# PRINCIPLES OF ELECTRON TUBES

GEWARTOWSKI
WATSON

# Principles of Electron Tubes

# Principles of Electron Tubes

## Including Grid-Controlled Tubes, Microwave Tubes, and Gas Tubes

*by*

### J. W. GEWARTOWSKI and H. A. WATSON

*Members of the Technical Staff*
*Bell Telephone Laboratories*

*To Our Parents*

# PREFACE

This book is a revision and extension of notes prepared by the authors for courses given to the Communications Development Training Program at Bell Telephone Laboratories. This study program is given to all new members of the technical staff who have completed their university training at the B.S. or M.S. level. Thus the book is primarily intended for use at the senior or first-year graduate level.

The book should also be useful to graduate engineers working on electron-tube development and manufacture and to engineers using electron tubes as circuit elements. Detailed descriptions are included of practical electron tubes as examples.

Throughout the text an effort has been made to present a coherent picture of the use of electron-field interactions to obtain useful device performance. The first 13 chapters relate primarily to vacuum tubes, and the last four chapters are concerned with gas-discharge devices. The text first considers the basic laws of electron motion in fields and electron emission. This is followed by a discussion of electron lenses and electron guns. Next, grid-controlled vacuum tubes are examined, and their equivalent circuits are derived. High-frequency limitations of grid-controlled tubes are explored through the concept of induced currents. This is followed by a detailed study of microwave tubes. A final chapter on vacuum tubes considers the noise performance of these devices. The last four chapters of the text consider first the Townsend discharge in a gas diode, followed by a discussion of cold-cathode and hot-cathode gas tubes, and finally a description of gas lasers.

Although considerable mathematical detail is included, an effort has been made to stress the physical principles of each device. Problems are included at the ends of most of the chapters to illustrate further concepts relative to the text material. References are cited for those who wish to pursue particular subjects in more detail. A notation has been adopted which is consistent with the symbols used in the literature, insofar as this is possible in a coherent presentation. No attempt has been made to include any historical comments concerning electron tubes. For the most part, tubes are discussed in configurations that are in practical use today rather than in those originally conceived.

The authors wish to thank their many associates at Bell Telephone Laboratories who provided information, drawings, and criticism of the manuscript. Dr. T. B. Ramachandran of Lehigh University carefully reviewed the earlier chapters and made a number of helpful suggestions concerning later chapters while these were in preparation. The authors also wish to express their appreciation to four secretaries who typed the manuscript and its many revisions: Miss J. Mishko, Miss D. J. Delong, Mrs. S. T. Otto, and Mrs. J. K. Ziegler.

J.W.G. AND H.A.W.

# CONTENTS

**APPENDICES**

# LIST OF PRINCIPAL SYMBOLS

$A$ = area

$A_1$ = initial loss factor, see Equation (10.1-62)

$A_2$ = space charge loss factor, see Equation (10.1-68)

$a$ = radius of a helix-type slow-wave circuit

$B, \mathbf{B}$ = magnetic flux density; $B$ = amplifier bandwidth

$B_m$ = magnetic flux density within a permanent magnet material

$b$ = susceptance term in electron beam admittance; velocity parameter, see Equation (10.1-40)

$C$ = capacitance; small-signal gain parameter, see Equation (10.1-36)

$c$ = velocity of light in free space

$D, \mathbf{D}$ = electric flux density; $D$ = diffusion coefficient

$d$ = molecular diameter; electrode spacing in a planar device; circuit loss parameter, see Equation (10.1-41)

$E, \mathbf{E}$ = electric field intensity

$\mathbf{E}_1(x,y,z)$ = a vector function of position having the dimensions of (meters)$^{-1}$, see Equation (6.1-5)

$-e$ = electric charge of the electron

$F$ = noise figure; $F_1, F_2$ = focal points of an electron lens

$f$ = frequency of a sinusoidally varying quantity; $f_1, f_2$ = focal lengths of an electron lens

$G$ = conductance; average number of metastables striking the cathode for each electron leaving the cathode (Chapter 14)

$\mathcal{G}$ = available power gain

$g$ = conductance term in electron beam admittance

$g_m$ = transconductance

$g_o = I_o/V_o$ = ratio of dc beam current to dc beam voltage

$H, \mathbf{H}$ = magnetic field intensity; $H$ = average number of metastables produced by the release of a single electron from the cathode (Chapter 14)

$H_m$ = magnetic field intensity in a permanent magnet material

$h$ = Planck's constant; hub thickness in a magnetron

$I$ = electric current; average number of ions generated by the release of a single electron from the cathode (Chapter 14)

$I_o$ = dc beam current; $I_{ST}$ = dc starting current in a backward-wave oscillator

$i = ReI\epsilon^{j\omega t}$ = instantaneous current, where $I$ is a phasor quantity having magnitude, phase, and the dimensions of an electric current

$i$ = electron convection current phasor (Chapters 10,11)

$J$ = current density; $J_o$ = dc beam current density; cathode emission current density; $J_T$ = total current density (convection + displacement)

$j = \sqrt{-1}$

$K$ = beam-coupling impedance

$k$ = Boltzmann's constant; wave number, $k = \omega\sqrt{\mu\mu_o\varepsilon\varepsilon_o}$

$L$ = inductance; periodic length; mean free path

$l$ = length

$M$ = mutual inductance; beam-coupling coefficient

$M_{1(n)}$ = gap factor for the $n$th space harmonic; see Equation (10.2-20)

$M_{2(n)}$ = impedance reduction factor for the $n$th space harmonic, see Equation (10.2-21)

$m$ = mass of the electron

$N$ = number of wavelengths, see Equation (10.1-70); number of molecules in one gram molecular weight

$n$ = index of refraction; number of atoms or molecules per unit volume

$P$ = average power; probability; pressure in mks units

$p$ = pressure in mm of Hg

$Q$ = a measure of quality of a resonant circuit or resonant cavity, see Equation (6.4-2); $Q_e$ = external $Q$; $Q_l$ = loaded $Q$; $Q_u$ = unloaded $Q$

$QC$ = space-charge parameter, see Equation (10.1-37)

$q$ = electric charge

$R$ = resistance; universal constant for one mole of gas, $R = Nk = PV/T$

$r$ = radius; $\mathbf{r}$ = a radial vector of length $r$

$r_a$ = dynamic anode resistance, or "plate" resistance

$\dot{r} = dr/dt$; $\ddot{r} = d^2r/dt^2$; $r' = dr/dz$; $r'' = d^2r/dz^2$

$S$ = surface area; beam cross-sectional area

$\mathbf{S}$ = Poynting vector

$T$ = temperature in °K; $T_e$ = electron temperature; $T_o$ = standard reference temperature of 290°K

$t$ = time

$u, \mathbf{u}$ = velocity; $u_o$ = dc beam velocity

$V$ = electric potential, voltage; volume

$V_o$ = dc beam voltage

$v = ReVe^{j\omega t}$ = instantaneous voltage, where $V$ is a phasor quantity having magnitude, phase, and the dimensions of voltage

$v_g$ = group velocity; $v_p$ = phase velocity

$W$ = energy, average stored energy; $W_T$ = electron-volt equivalent of $kT$; $W_t$ = emission energy parallel to the cathode surface; $W_n$ = emission energy normal to the cathode surface; $W_L$ = average stored energy in a unit cell of length $L$; $W_l$ = average stored energy per unit length

$X$ = bunching parameter, see Equation (9.2-9).

$x_1$ = rate of growth of the growing wave (Chapter 10)

$Y$ = admittance

$Z$ = impedance; $Z_o$ = characteristic impedance

$\alpha$ = ionization coefficient per centimeter; circuit attenuation per unit length

$\beta$ = phase shift per unit length along the axis of a slow-wave circuit; coefficient expressing the average number of metastables produced by a single electron in advancing one centimeter through a gas under the influence of an applied field (Chapter 14)

$\beta_e = \omega/u_o$; $\beta_n$ = propagation constant for the $n$th order space harmonic; $\beta_q = \omega_q/u_o$

$\Gamma$ = complex propagation coefficient

$\Gamma^2$ = space-charge smoothing factor for shot noise

$\gamma$ = average number of electrons emitted from a cold cathode per incident ion

$\gamma_n = \sqrt{\beta_n{}^2 - k^2}$ = transverse decay parameter for the fields in a slow-wave structure

$\delta$ = ratio of secondary electrons emitted from a surface to primary electrons; skin depth; dimensionless complex growth constant of a wave (Chapters 10 and 11); gap length in a periodic structure (Chapters 10 and 11)

$\varepsilon$ = relative dielectric constant; $\varepsilon_o$ = permittivity of free space

$\epsilon$ = base for natural logarithms

$\eta$ = efficiency; ionization coefficient per volt; $\eta = e/m$ = ratio of charge to mass of an electron; $\eta_e$ = electronic efficiency; $\eta_c$ = circuit efficiency

$\theta$ = angular coordinate in a cylindrical or spherical coordinate system; dc transit angle in a klystron amplifier

$\lambda$ = wavelength; $\lambda_o$ = wavelength in free space; $\lambda_t$ = guide wavelength; $\lambda_c$ = free-space wavelength at the cutoff frequency

$\mu$ = relative permeability; amplification factor; a dimensionless coordinate which varies from zero on the beam axis to unity at the beam edge (Chapter 4); ion mobility; $\mu_o$ = permeability of free space; $\mu_{es}$ = electrostatic amplification factor

$\nu$ = frequency of light radiation

$\rho$ = charge per unit volume

$\sigma$ = charge per unit area; conductivity of a medium; distance from the trajectory of a nonthermal electron to the trajectory of a thermal electron emitted from the same point on the cathode with transverse velocity $\sqrt{kT/m}$ (Chapter 4)

$\tau$ = charge per unit length

$\phi$ = work function in electron volts; magnetic flux

$\psi$ = magnetic potential; pitch angle of a helix

$\omega = 2\pi f$ = radian frequency; $\omega_p$ = plasma frequency in radians/sec, see Equation (9.3-23); $\omega_q$ = reduced plasma frequency in radians/sec, see Equation (9.3-30); $\omega_c = eB/m$ = radian cyclotron frequency

# Chapter 1

## ELECTRONS AND FIELDS

Electron-field interactions play an essential role in the operation of all electron tubes. Fields determine the motion of the electrons in the inter-electrode space of a tube, and the electron motion in the interelectrode space determines the currents that flow in the external circuit connected between the electrodes.[1] It is appropriate therefore that we begin this text with a review of the laws that govern the motion of electrons in electric and magnetic fields, as well as some properties of the fields themselves. The discussion of fields in the present chapter will be limited to static electric and magnetic fields. Time-varying fields will be considered in later chapters.

In describing fields and electron-field interactions, we must rely on certain experimental laws of physics. Several such laws from which much of our discussion of the present chapter will develop are:

1. A particle with charge $q$ is acted on by an electric field $\mathbf{E}$ with a force proportional to $q\mathbf{E}$, the force being in the direction of the field if $q$ is positive, and in the opposite direction if $q$ is negative.

2. When a particle with charge $q$ moves with velocity $\mathbf{u}$ in a magnetic field $\mathbf{B}$, it experiences a force proportional to the vector product $q\mathbf{u} \times \mathbf{B}$. The force is in the direction of $\mathbf{u} \times \mathbf{B}$ if the charge is positive, and in the opposite direction if the charge is negative.

3. The electric flux crossing a closed surface surrounding a quantity of charge is proportional to the amount of charge enclosed by the surface and is independent of the shape of the surface. This is known as *Gauss's Law*. A point charge therefore acts as a point source of electric flux, and with each unit of charge there is associated a certain total amount of electric flux.

4. In a static magnetic field the line integral of the magnetic field intensity $\mathbf{H}$ around any closed path surrounding a flow of current $I$ is propor-

---

[1]See Chapter 6.

tional to the flow of current through a surface enclosed by the path. This relationship is known as *Ampere's Circuital Law*. Lines of magnetic flux close on themselves, and there are no point sources of magnetic flux.

The constants of proportionality used in expressing the foregoing experimental laws, together with the units used to measure mass, length, time, and charge, serve to determine the units in which the electric and magnetic field quantities are measured. Several systems of units are in use at present, each with its own particular advantages. However, the meter-kilogram-second system is perhaps the most widely accepted in electron-tube work, and we shall adhere to it throughout this text. Appendix I lists the mks units in which electric and magnetic field quantities are measured, together with their dimensions. Appendix II lists values of a number of physical constants, and Appendix III presents a summary of relationships governing static electric and magnetic field quantities.

Relativistic effects will be neglected throughout this text; that is, electron velocities will be considered small compared with the velocity of light.

### 1.1   Electron Motion in an Electric Field

#### (a)   *Change of Kinetic Energy and the Concept of Electric Potential*

A charge of $q$ coulombs in an electric field $\mathbf{E}$ volts/meter is acted on by a force $q\mathbf{E}$ newtons. The force is in the direction of the field if $q$ is a positive charge, and in the opposite direction for a negative charge. Thus, when an electron moves in an electric field $\mathbf{E}$, it experiences a force $-e\mathbf{E}$ newtons, where $-e$ is the charge on the electron, $e$ being equal to $1.602 \times 10^{-19}$ coulomb. The resulting motion of the electron is described in rectangular coordinates by the three equations,

$$m\frac{d^2x}{dt^2} = -eE_x, \qquad m\frac{d^2y}{dt^2} = -eE_y, \qquad m\frac{d^2z}{dt^2} = -eE_z \qquad (1.1\text{-}1)$$

where $m$ is the mass of the electron, and $E_x$, $E_y$, and $E_z$ are the components of $\mathbf{E}$ in the coordinate directions. If the first of these equations is multiplied by $dx$ on both sides, we obtain

$$m\left[\frac{d(dx/dt)}{dt}\right]dx = m\frac{dx}{dt}d\left[\frac{dx}{dt}\right] = d\left[\tfrac{1}{2}m\left[\frac{dx}{dt}\right]^2\right] = -eE_x\,dx \quad (1.1\text{-}2)$$

The right-hand part of this equation states that the portion of the electron's kinetic energy associated with its motion in the $x$ direction is changed by an amount $-eE_x\,dx$ when the electron moves a distance $dx$ in the $x$ direction under the influence of the field. Similar expressions hold for motion in the $y$ and $z$ directions. It follows, therefore, that if the electron moves a distance

$d\mathbf{l}$ under the influence of the electric field, its net change in kinetic energy is equal to the vector product $-e\mathbf{E}\cdot d\mathbf{l}$. This quantity may be positive or negative depending on the angle between $\mathbf{E}$ and $d\mathbf{l}$.

If the electron travels from point $A$ to point $B$ under the influence of the electric field, its total change in kinetic energy is given by

$$\text{change in k.e.} = -e\int_{A}^{B} \mathbf{E}\cdot d\mathbf{l} \tag{1.1-3}$$

where the integral is taken over the path followed by the electron from $A$ to $B$. This expression is of much importance in determining the behavior of charged particles in electric fields. It holds for time varying fields as well as for static fields.

If the field is constant with time and if the work done by the field on the electron serves only to change the kinetic energy of the electron, the field is said to be conservative. For such a field the integral in Equation (1.1-3) is independent of the path taken from $A$ to $B$, and we can write

$$-e\oint_{\substack{\text{closed}\\\text{path}}} \mathbf{E}\cdot d\mathbf{l} = 0 \tag{1.1-4}$$

where the integral is taken around a closed path. In this case we can ascribe to each point in space a scalar potential such that the difference in potential between two points is equal to the line integral of $\mathbf{E}$ along any path between them. A potential difference of 1 volt exists between points $A$ and $B$ if the line integral of $\mathbf{E}$ along any path between them is equal to 1 volt. (Potential difference is sometimes called electromotive force or emf.)

If $dl$ is an increment of distance in the direction of an electric field $E$, the change in potential $dV$ over the distance $dl$ can be expressed as $|dV| = Edl$, and we can write that

$$\mathbf{E} = -\nabla V \tag{1.1-5}$$

where $V$ is the scalar potential. The minus sign implies that the field is directed from regions of higher potential to ones of lower potential. Equation (1.1-5) is valid in regions in which there is space charge as well as regions that are free of charge. From the equation, it is evident that $E$ has the dimensions of volts per meter.

If an electron starts from rest and is accelerated through a potential rise of $V$ volts, it acquires an amount of kinetic energy given by

$$\tfrac{1}{2}mu^2 = -e\int \mathbf{E}\cdot d\mathbf{l} = eV \text{ joules} \tag{1.1-6}$$

Substituting the experimentally measured values for $e$ and $m$ in this, we find the velocity of the electron to be

$$u = 5.93 \times 10^5 \sqrt{V} \text{ meters/sec} \tag{1.1-7}$$

A unit of energy frequently used to measure energies gained or lost by an electron is the electron volt. It is equal to $1.602 \times 10^{-19}$ joule and is the kinetic energy gained by an electron when it is accelerated through a potential rise of 1 volt. If the electron travels between points differing in potential by $V$ volts, its change in kinetic energy is $V$ electron volts.

### (b)   Electron Trajectories in an Electric Field

Figure 1.1-1(a) shows two electrodes, $A$ and $B$, of arbitrary shape. Electrode $A$ is grounded, and electrode $B$ is held at a positive potential with respect to ground. The path that might be followed by an electron which
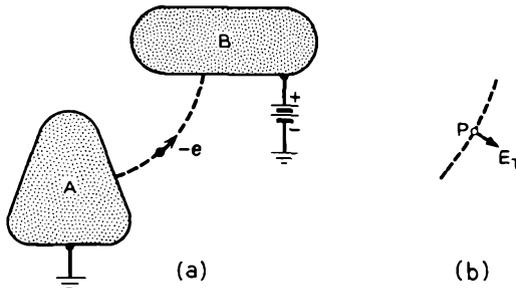


FIG 1.1-1 An electron trajectory between two conducting electrodes.

starts from rest at electrode $A$ and is accelerated in the direction of electrode $B$ is shown by a broken line. Figure 1.1-1(b) shows a curved portion of the path passing through point $P$. The electric field $\mathbf{E}$ acting on the electron at point $P$ can be resolved into two components, one parallel to the trajectory and one transverse to it. The transverse component, $E_T$, is responsible for bending the path of the electron and hence must lie in the plane of curvature of the trajectory. If $u$ is the velocity of the electron at point $P$ and $r$ is the radius of curvature of the trajectory at that point,

$$\frac{mu^2}{r} = eE_T \tag{1.1-8}$$

Since the electron started from rest at electrode $A$, its kinetic energy at point $P$ is given by

$$\tfrac{1}{2}mu^2 = eV \tag{1.1-9}$$

where $V$ is the potential at point $P$. Combining these two equations, we obtain

$$r = \frac{2V}{E_T} \tag{1.1-10}$$

Now $V$ and $E_T$ are directly proportional to the voltage applied to electrode $B$. Since $r$ is equal to twice the ratio of these quantities, it follows that $r$ is independent of the voltage applied to electrode $B$. Consequently, if the electron starts from rest, its trajectory is the same for all positive voltages applied to electrode $B$.

A second point, which may seem intuitively clear, follows from similar reasoning. When the linear dimensions in Figure 1.1-1(a) are scaled by a constant factor, the trajectory followed by the electron is scaled by the same factor. Let us suppose that all linear dimensions are multiplied by the factor $k$ and that the voltage applied to electrode $B$ remains unchanged. In this case the potential $V$ at corresponding points between the electrodes will be unchanged. The direction of the electric field intensity also will be unchanged, but its magnitude will be $1/k$ times as great. From Equation (1.1-10), it follows that $r$ becomes $k$ times its previous value, so that $r$ and the trajectory scale with the other linear dimensions.

A third conclusion we can draw from Equation (1.1-10) is that the trajectory is independent of the mass or charge of the particle, provided, of course, that the charge is finite and negative and the mass is not zero. Hence a negative ion would follow the same path as the electron, provided both started from rest at the same point on electrode $A$.

## 1.2 Motion in Combined Electric and Magnetic Fields

When a particle with charge $q$ coulombs moves with velocity $\mathbf{u}$ meters per second in a magnetic field $\mathbf{B}$ webers per square meter, it experiences a force $q\mathbf{u} \times \mathbf{B}$ newtons. Thus, an electron moving in a magnetic field $\mathbf{B}$ experiences a force $-e\mathbf{u} \times \mathbf{B}$ newtons, and the resulting acceleration of the electron is $-(e/m)\mathbf{u} \times \mathbf{B}$ meters per second$^2$.

The vector $\mathbf{u} \times \mathbf{B}$ has the components $B_z u_y - B_y u_z$ in the $x$ direction, $B_x u_z - B_z u_x$ in the $y$ direction, and $B_y u_x - B_x u_y$ in the $z$ direction, where $u_x$, $u_y$, and $u_z$ are the components of $\mathbf{u}$ in the coordinate directions, and $B_x$, $B_y$, and $B_z$ are the components of $\mathbf{B}$ in the coordinate directions. If both an electric field and a magnetic field act on an electron, the differential equations describing the motion of the electron are

$$\frac{d^2x}{dt^2} = -\frac{e}{m}\left(E_x + B_z\frac{dy}{dt} - B_y\frac{dz}{dt}\right) \tag{1.2-1}$$

$$\frac{d^2y}{dt^2} = -\frac{e}{m}\left(E_y + B_x\frac{dz}{dt} - B_z\frac{dx}{dt}\right) \tag{1.2-2}$$

and

$$\frac{d^2z}{dt^2} = -\frac{e}{m}\left(E_z + B_y\frac{dx}{dt} - B_x\frac{dy}{dt}\right) \tag{1.2-3}$$

where $E_x$, $E_y$, and $E_z$ are the components of the electric field in the co-ordinate directions. In cylindrical coordinates these equations become

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{e}{m}\left[E_r + B_z r\frac{d\theta}{dt} - B_\theta\frac{dz}{dt}\right] \qquad (1.2\text{-}4)$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = -\frac{e}{m}\left[E_\theta + B_r\frac{dz}{dt} - B_z\frac{dr}{dt}\right] \qquad (1.2\text{-}5)$$

$$\frac{d^2z}{dt^2} = -\frac{e}{m}\left[E_z + B_\theta\frac{dr}{dt} - B_r r\frac{d\theta}{dt}\right] \qquad (1.2\text{-}6)$$

We shall find a number of occasions to make use of these equations in later chapters.

Because the force resulting from the magnetic field is perpendicular to the motion of the electron, any component of force parallel to the trajectory must result from the electric field. However, it is the force parallel to the trajectory which changes the electron's kinetic energy, and consequently *only an electric field can change the kinetic energy of an electron.*

If the electric field is zero and if the velocity of the electron is perpendicular to the magnetic field, the electron moves in a circular path as illustrated in Figure 1.2-1. The radius $R$ of the path is determined by the relation
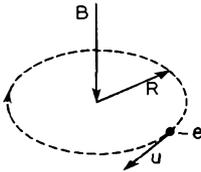


FIG. 1.2-1 The motion of an electron in a magnetic field when the velocity of the electron is perpendicular to the magnetic field.

$$\text{acceleration} = \frac{u^2}{R} = \frac{e}{m}uB \qquad (1.2\text{-}7)$$

or

$$R = \frac{mu}{eB} \qquad (1.2\text{-}8)$$

The angular frequency of the circular motion of the electron is given by

$$\omega = \frac{u}{R} = \frac{eB}{m} \qquad (1.2\text{-}9)$$

As a simple example of motion in combined electric and magnetic fields, let us consider the case illustrated in Figure 1.2-2. Here, an electric field $E$ lies parallel to the $-y$ direction of a rectangular coordinate system, and a magnetic field $B$ lies parallel to the $-z$ direction. We shall assume that an electron starts from the origin at time $t = 0$ with zero velocity. The elec-

tron is initially acted on only by the electric field, but as it advances in the $y$ direction and gains velocity, it is acted on by the magnetic field with a force proportional to the product of its velocity and the magnetic flux density.
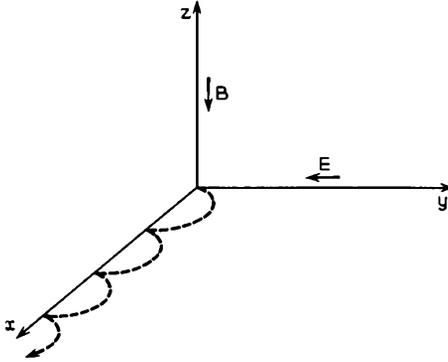


FIG. 1.2-2 The trajectory of an electron which starts from rest in crossed electric and magnetic fields.

As a result, the trajectory is bent back toward the $x$ axis. For this problem Equations (1.2-1), (1.2-2), and (1.2-3) reduce to the two simple equations,

$$\frac{d^2x}{dt^2} = \frac{e}{m}B\frac{dy}{dt}$$

$$\frac{d^2y}{dt^2} = \frac{e}{m}E - \frac{e}{m}B\frac{dx}{dt}$$

(1.2-10)

It is easily shown that these equations have the solutions

$$\frac{dx}{dt} = \frac{E}{B}(1 - \cos \omega t)$$

$$\frac{dy}{dt} = \frac{E}{B}\sin \omega t$$

(1.2-11)

and

$$x = \frac{E}{\omega B}(\omega t - \sin \omega t)$$

$$y = \frac{E}{\omega B}(1 - \cos \omega t)$$

(1.2-12)

where $\omega = eB/m$. Equations (1.2-12) are the equations of a cycloid, the electron trajectory being as illustrated in Figure (1.2-2). Each $2\pi/\omega$ seconds the electron returns to the $x$ axis and then repeats the curved part of the trajectory.

Next let us consider the scaling of electron trajectories in a region in which there is both an electric field and a magnetic field. It is convenient to rewrite Equation (1.1-8) to express the radius of curvature of the trajectory as

$$r = \frac{mu^2}{\text{transverse force}} \qquad (1.2\text{-}13)$$

where the transverse force in this case may result from both an electric field and a magnetic field. The transverse force, of course, lies in the plane of curvature of the trajectory. Clearly, if we change the electric field intensity and the magnetic flux density in such a manner that the right-hand side of this equation is unchanged for all points on the trajectory, the shape of the trajectory will not be changed. Suppose the electric field intensity at all points is increased by the factor $a^2$ and the magnetic flux density is increased by the factor $a$. Then an electron which starts from rest at the beginning of the trajectory and travels to point $P$ on the trajectory will have $a^2$ times as much energy at point $P$, and its velocity will be $a$ times as great. The part of the transverse force resulting from the electric field will also be $a^2$ times as great; and since the part of the transverse force that results from the magnetic field is proportional to the product of $u$ and $B$, this also will be increased by the factor $a^2$. Hence both the numerator and denominator of the right-hand side of Equation (1.2-13) will be increased by the factor $a^2$, and the radius $r$ will be unchanged. Thus, if we increase the electric field intensity at all points in space by the factor $a^2$ and the magnetic flux density by the factor $a$, the trajectory of an electron which leaves a given point in space with zero initial velocity will remain unchanged, but the electron will travel $a$ times as fast. (The reader will readily verify this to be the case for the trajectories given by Equations (1.2-12).)

By similar reasoning it is easily shown that, if the linear dimensions of the electrodes are increased by the factor $b$, and if all the voltages applied to the electrodes are increased by the factor $b^2$, and if the magnetic flux density *at corresponding points* between the electrodes is unchanged, the electron trajectory will also scale with the other linear dimensions of the system. In this case the electron velocity at corresponding points of the trajectory will be increased by the factor $b$.

As a final point, we should note that the motion of an electron in an electric or magnetic field is governed entirely by the forces acting on it. The only way we can change the kinetic energy of an electron is to cause the electron to be acted on by an electric field. *Changing the potential in the region does not in itself change the kinetic energy of the electron.*

## 1.3   Conservation of Energy and Charge

One of the most important laws governing the behavior of physical processes is the principle of conservation of energy. It states that energy

can never be created or destroyed. As applied to electron tubes, it tells us that whenever an electron gains kinetic energy, we can in principle account for the source of kinetic energy and show that the source lost an equal amount of energy. Similarly, when an electron loses kinetic energy, we can in principle find an amount of energy which has appeared elsewhere in the system equal to the lost kinetic energy.

Another significant law we learn from experimental physics is the principle of conservation of charge. This principle states that the total charge of a system, both positive and negative, can be changed only by adding charge to the system or removing charge from the system. In later chapters we shall frequently have occasion to consider volume charge densities or "space-charge densities" arising from a large number of electrons in a region of space. If $\rho(x,y,z)$ is the volume charge density, the total charge in an element of volume $\Delta v$ is $\rho(x,y,z)\,\Delta v$. The principle of conservation of charge tells us that, if this quantity is changing with time, charge is flowing across the surface of the volume element, such that the total amount of charge both inside and outside is constant. Expressed mathematically, the principle states that

$$\int_{\substack{\text{closed}\\\text{surface}}} \mathbf{J} \cdot \mathbf{n}\,dS = -\frac{\partial}{\partial t} \int_{\text{volume}} \rho(x,y,z)\,dv \qquad (1.3\text{-}1)$$

where $\mathbf{J}(x,y,z)$ is the current density associated with the flow of charge, and $\mathbf{n}$ is a unit vector normal to the surface element $dS$ and pointing outward. Dividing both sides by $\Delta v$ and taking the limit as $\Delta v \to 0$, the left-hand side becomes the divergence of $\mathbf{J}$, and we obtain

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \qquad (1.3\text{-}2)$$

This is known as the *equation of continuity*. We shall find a number of occasions to make use of it in later chapters.

### 1.4 Static Electric Fields — Gauss's Law, Poisson's and Laplace's Equations

*(a) Gauss's Law*

In mks units the electric flux density $\mathbf{D}$ is related to the electric field intensity $\mathbf{E}$ by $\mathbf{D} = \varepsilon\varepsilon_o\mathbf{E}$, where $\varepsilon$ is the relative dielectric constant of the medium, and $\varepsilon_o$ is the permittivity of free space. The relative dielectric constant $\varepsilon$ is a dimensionless constant, which in free space has the value 1.

The constant $\varepsilon_o$ is approximately equal to[2] $8.854 \times 10^{-12}$ and has the dimensions of farads per meter or coulombs per volt per meter. Since $\mathbf{E}$ has the dimensions of volts per meter, the vector $\mathbf{D}$ has the dimensions of coulombs per square meter. (The vector $\mathbf{D}$ is sometimes called the displacement vector.)

If we surround a quantity of charge by a closed surface, a certain total amount of electric flux crosses the surface because of the charge inside. Gauss's Law states that no matter what surface we choose to surround the charge, the total flux crossing the surface is the same. Furthermore, the amount of flux crossing the surface is proportional to the charge enclosed. *Hence with each unit of charge there is associated a certain total amount of flux.* In mks units the flux crossing the surface is numerically equal to the charge in coulombs enclosed by the surface. Gauss's Law therefore can be expressed as

$$\int_{\substack{\text{closed} \\ \text{surface}}} \mathbf{D} \cdot \mathbf{n} dS = \int_{\substack{\text{closed} \\ \text{surface}}} \varepsilon \varepsilon_o \mathbf{E} \cdot \mathbf{n} dS = q \qquad (1.4\text{-}1)$$

where $\mathbf{n}$ is a unit vector normal to the surface element $dS$, and $q$ is the charge enclosed by the surface. If there is a distribution of charge within the region, the theorem can be written in the form

$$\int_{\substack{\text{closed} \\ \text{surface}}} \mathbf{D} \cdot \mathbf{n} dS = \int_{\text{volume}} \rho(x,y,z) dv \qquad (1.4\text{-}2)$$

where $\rho(x,y,z)$ is the volume charge density, and the integral on the right is taken over the volume enclosed by the surface. Equations (1.4-1) and (1.4-2) are valid even if the surface over which the integrals are taken passes through a conductor or other solid matter, or if it passes through a region of space charge. (However, if the surface element $dS$ lies in a conductor, $\mathbf{E} = 0$, and the flux crossing $dS$ is zero.)

If the volume enclosed by the surface in Equation (1.4-2) is $\Delta v$, and if both sides of the equation are divided by $\Delta v$, and the limit is taken as $\Delta v \to 0$, we obtain

$$\nabla \cdot \mathbf{D} = \rho \qquad (1.4\text{-}3)$$

This provides another useful expression of Gauss's Law.

---

[2]In mks units the magnetic permeability of free space $\mu_o$ is defined to be equal to $4\pi \times 10^{-7}$, and the constants $\mu_o$ and $\varepsilon_o$ are related by $\mu_o \varepsilon_o = 1/c^2$, where $c$ is the velocity of light. Hence $\varepsilon_o$ can be determined by experimental measurement of the velocity of light. It is found that $c \approx 2.996 \times 10^8$ meters/sec, so that $\varepsilon_o \approx 8.854 \times 10^{-12}$, or approximately $1/(36\pi \times 10^9)$.

If $q$ in Equation (1.4-1) is positive, the net electric flux crossing the surface is directed outward, and if $q$ is negative, the net electric flux is directed inward. If the charge enclosed by the surface consists of two equal but opposite charges, the net electric flux crossing the surface is zero.

Two results that follow directly from Gauss's Law and symmetry arguments are:

1. The electric field in free space at a distance $r$ from a point charge $q$ is given by

$$E = \frac{q}{4\pi\epsilon_o r^2}\text{volts/meter} \tag{1.4-4}$$

2. The electric field in free space outside a cylindrical charge distribution of uniform axial charge density is given by

$$E = \frac{\tau}{2\pi\epsilon_o r}\text{volts/meter} \tag{1.4-5}$$

where $\tau$ is the axial linear charge density in coulombs per meter, and $r$ is the radius at which $E$ is determined.

The concept of lines of electric flux, or field lines, is useful in presenting a picture of an electric field distribution. In the case of two equal but opposite point charges, the electric field lines terminate on the two charges and extend from one charge to the other, the lines being directed from the positive charge to the negative charge. The total number of lines is proportional to the amount of charge at the ends of the field lines. The field lines are parallel to the direction of the electric field, and the number of lines crossing unit area normal to the direction of the field is proportional to the average electric flux density over the unit of area.

Static electric fields are always associated with coulomb charges — either point charges, surface charges, volume charges, or perhaps a combination of the three. In electron-tube work a density of electrons in the interelectrode space of a tube can often be considered to be a volume charge density, or "space-charge density," even though it is really a cloud of individual point charges.

If a point charge is brought close to a conductor, currents flow in the conductor until a charge distribution is built up on its surface which exactly cancels the electric field that would otherwise be present within the conductor. The surface charge is said to be an induced charge. Thus, when electrons are present in the interelectrode space of a vacuum tube, an amount of positive charge equal to the total charge on the electrons is induced on the electrodes or other nearby surfaces, and one can imagine electric field lines extending from the induced surface charges to the electrons in the interelectrode space.

Charges on conductors are always surface charges. A net volume charge density within a conductor would lead to electric fields within the conductor with the result that currents would flow causing neutralization of the charge. Similarly, a static electric field at the surface of a conductor is always normal to the surface of the conductor, since otherwise it would have a component parallel to the surface, and charge would flow along the surface.

By a further application of Gauss's Law, it is easily shown that the electric field intensity $E$ in free space at the surface of the charged conductor is given by

$$E = \frac{\sigma}{\varepsilon_o} \tag{1.4-6}$$

where $\sigma$ is the surface charge density.

Equation (1.4-6) can be used to obtain an expression for the capacitance of a parallel-plate capacitor. When the capacitor is charged, electric field lines extend from the surface charge on one plate to the surface charge on the other, the charge on the plate at higher potential being positive, and that on the plate at lower potential being negative. If the spacing between plates is small compared with their linear dimensions so that edge effects are negligible, the potential difference from one plate to the other can be expressed as $V = Ed = \sigma d/\varepsilon_o = qd/\varepsilon_o A$, where $d$ is the spacing between the plates, $A$ is the area of a single plate, $\sigma$ is the surface charge density, and $q$ is the total charge on a single plate. The capacitance of the device is defined as the ratio of $q$ to $V$, or

$$C = \frac{q}{V} = \frac{\varepsilon_o A}{d} \tag{1.4-7}$$

In mks units, $C$ is measured in farads. If the space between the plates were filled with a material of relative dielectric constant $\varepsilon$, it is easily shown that $E = \sigma/\varepsilon\varepsilon_o$, and $C = \varepsilon\varepsilon_o A/d$.

### (b)  Poisson's and Laplace's Equations

Equation (1.4-3) can be written in the form

$$\nabla \cdot \mathbf{D} = \nabla \cdot (\varepsilon\varepsilon_o \mathbf{E}) = \rho \tag{1.4-8}$$

Now $\mathbf{E} = -\nabla V$, and in free space $\varepsilon = 1$. It follows that in a region of free space in which there is a distributed charge density $\rho(x,y,z)$, the potential $V$ is described by the equation

$$\nabla \cdot (\nabla V) = \nabla^2 V = -\frac{\rho}{\varepsilon_o} \tag{1.4-9}$$

This relationship is known as Poisson's Equation.

If there is no space charge in the region, $\rho = 0$, and the potential satisfies Laplace's Equation,

$$\nabla^2 V = 0 \tag{1.4-10}$$

As an example of a problem that can be solved with the aid of Poisson's Equation, let us consider the potential within a long conducting cylindrical tube filled with a uniform charge density $\rho_o$. (We can imagine that an electron beam of uniform charge density is directed down inside the tube and that the beam just fills the tube.) Using cylindrical coordinates, Poisson's Equation for this problem becomes

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dV}{dr}\right) = -\frac{\rho_o}{\varepsilon_o} \tag{1.4-11}$$

since there is no variation of $V$ in the $\theta$ or $z$ directions. The reader will readily verify that $V = -(\rho_o/4\varepsilon_o)r^2 + c_1 \ln r + c_2$ is a solution of this equation, where $c_1$ and $c_2$ are constants. Evidently $c_1 = 0$, since $V$ is finite at $r = 0$. If the inside radius of the conducting tube is $R$ meters, and if the tube is at zero potential, the potential at radius $r$ is given by $V = (\rho_o/4\varepsilon_o)$ $(R^2 - r^2)$ for $r \leq R$. Positive space charge raises the potential within the cylinder, and negative space charge lowers it.

A problem that can be solved with the aid of Laplace's Equation is that of finding the potential in the region between two long concentric conducting cylinders which are held at different potentials. Since $V$ does not change in the $\theta$ or $z$ directions, Laplace's Equation for this problem becomes

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dV}{dr}\right) = 0 \tag{1.4-12}$$

This equation has the solution $V = c_1 \ln r + c_2$, where $c_1$ and $c_2$ are constants. If the inner cylinder is held at potential $V_o$ and the outer cylinder is at zero potential, and if their radii are $a$ meters and $b$ meters, respectively, it is easily shown that $V = (V_o \ln r/b)/(\ln a/b)$. A solution of Laplace's Equation which satisfies a particular set of boundary conditions is always unique, and the first and second derivatives of such a solution are continuous at all points between the bounding surfaces.

Potential distributions can also be obtained by integrating known electric field distributions along the direction of the field. In this case use is made of the relation $\mathbf{E} = -\nabla V$. Thus, if the axial charge density on the inner cylinder in the above problem were specified, we could integrate Equation (1.4-5) with respect to $r$ to obtain the potential as a function of $r$. In a similar manner, Equation (1.4-4) can be integrated with respect to $r$ to obtain the potential due to an isolated point charge. Thus

$$V = \frac{q}{4\pi\varepsilon_o r} + c_1 \tag{1.4-13}$$

where $c_1$ is a constant, and $r$ is the distance from the charge $q$ to the point at which $V$ is determined. If $V$ is assumed to be zero at large distances from the point charge, then $c_1 = 0$.

### (c) Superposition

Because Laplace's Equation is linear, the sum of the potentials arising from two or more point charges also satisfies it. If a region of space contains a number of point charges as well as surface charges and volume charges, the potential at point $P$ can be expressed as

$$V_p = \sum \frac{dq}{4\pi\varepsilon_o r} \qquad (1.4\text{-}14)$$

where $dq$ is a point charge or element of surface charge or volume charge, and $r$ is the distance from the point charge or element of charge to point $P$.

A problem that can be solved with the aid of Equation (1.4-14) is that of finding the potential at point $P$ outside a conducting sphere with uniform charge density $\sigma$. We shall assume that there are no other point charges,
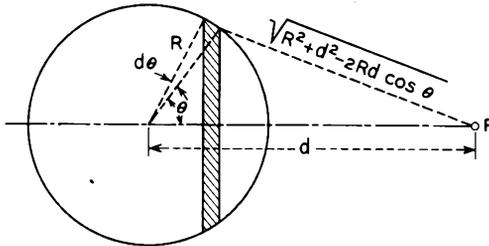


FIG. 1.4-1 A construction which may be used in determining the potential at a point $d$ meters from the center of a uniformly charged conducting sphere.

volume charges, or solid bodies nearby. With the aid of Figure 1.4-1 we can show that

$$V_p = \int_0^\pi \frac{\sigma 2\pi R^2 \sin\theta \, d\theta}{4\pi\varepsilon_o\sqrt{R^2 + d^2 - 2Rd\cos\theta}} = \frac{R^2\sigma}{\varepsilon_o d} = \frac{q}{4\pi\varepsilon_o d} \qquad (1.4\text{-}15)$$

where $R$ is the radius of the sphere, $\sigma$ is the surface charge density, $d$ is the distance from point $P$ to the center of the sphere, and $q$ is the total charge on the sphere.

Finally, let us note that, since the electric field at a given point is related to the potential gradient at the point by $\mathbf{E} = -\nabla V$ and since the gradient operator is linear, the total electric field is a vector sum of contributions arising from each of the separate point charges, and elements of surface

charge and volume charge in the region. Hence superposition applies to fields as well as potentials.

## 1.5   Static Magnetic Fields — Ampere's Circuital Law, Permanent Magnets

Static magnetic fields always result from charge in motion — sometimes an electron current in a conducting medium, or a beam of charged particles, or, in the case of permanent magnets, a preferred orientation of the electron spins or orbits in the solid matter of which the magnets are made. As in the case of an electric field, it is often convenient to picture a magnetic field in terms of magnetic flux or magnetic field lines. The lines lie parallel to the direction of the magnetic flux density **B**, and the number of lines crossing unit area normal to the direction of the field is proportional to $|\mathbf{B}|$.

When current flows in a long cylindrical conductor and the direction of flow is parallel to the axis of the conductor, the magnetic field lines are circles concentric with the conductor and lying in a plane perpendicular to

⊗ CURRENT FLOW IN

⊙ CURRENT FLOW OUT

(a)                                                  (b)
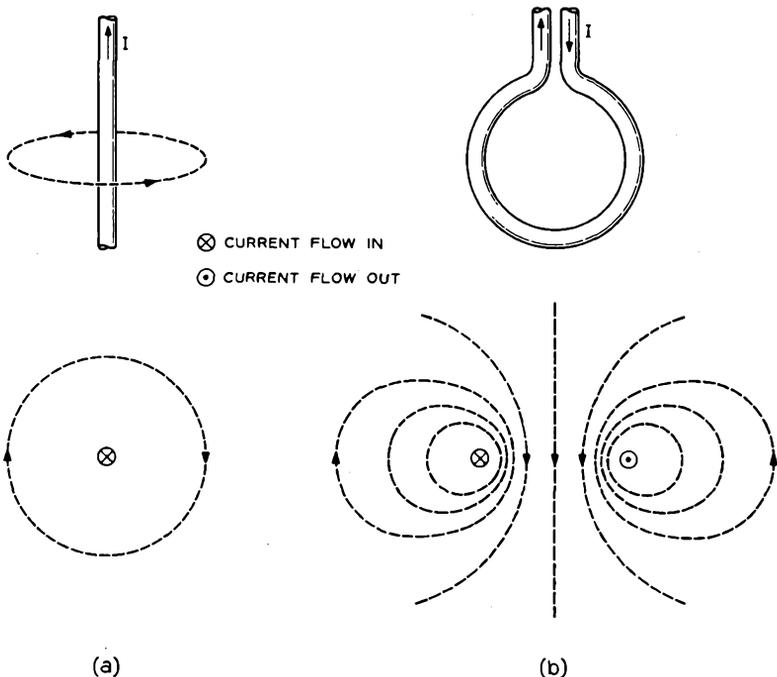
FIG. 1.5-1 Magnetic field lines associated with current flow in a wire and a loop.

the axis. The magnetic flux density is found to be greatest at the surface of the conductor and falls off inversely with distance from the axis at larger distances from the axis. Figure 1.5-1 illustrates the direction of the magnetic field in relation to the direction of current flow. If the conductor is bent in the form of a loop, the magnetic field lines still surround the flow of current, and each field line threads through the loop. In all cases the field lines close on themselves, and there are no point sources of magnetic field. Magnetic field lines never start or stop at a point or surface as do electric field lines.

Since the magnetic field lines close on themselves, the total magnetic flux crossing a closed surface must be zero. The magnetic flux crossing an element of area $dS$ can be expressed as $\mathbf{B} \cdot \mathbf{n} dS$, where $\mathbf{n}$ is a unit vector normal to the element of area. Hence

$$\int_{\substack{\text{closed} \\ \text{surface}}} \mathbf{B} \cdot \mathbf{n} dS = 0 \qquad (1.5\text{-}1)$$

If the volume enclosed by the surface is very small and can be represented by $\Delta v$ and if we take the limit as $\Delta v \to 0$, we obtain

$$\int_{\substack{\text{closed} \\ \text{surface} \\ \Delta v \to 0}} \frac{\mathbf{B} \cdot \mathbf{n} dS}{\Delta v} = \nabla \cdot \mathbf{B} = 0 \qquad (1.5\text{-}2)$$

In the mks system the unit of magnetic flux is the weber, and magnetic flux density $\mathbf{B}$ is measured in webers per square meter.

For some purposes it is convenient to define a vector $\mathbf{H}$, known as the magnetic field intensity vector, such that

$$\mathbf{B} = \mu\mu_o \mathbf{H} \qquad (1.5\text{-}3)$$

where $\mu$ is the relative permeability of the medium, and $\mu_o$ is the permeability of free space. The relative permeability $\mu$ is a dimensionless constant, which in free space is equal to 1. In mks units the constant $\mu_o$ is defined to be equal to $4\pi \times 10^{-7}$ and has the dimensions of henries per meter or webers per ampere-meter. Since $\mathbf{B}$ has the dimensions of webers per square meter, $\mathbf{H}$ has the dimensions of amperes per meter.

## (a)  Ampere's Circuital Law

Ampere's Circuital Law states that the line integral of $\mathbf{H}$ around any closed path which surrounds a flow of current $I$ is equal to the flow of current across the area enclosed by the path, or

$$\oint_{\text{closed path}} \mathbf{H} \cdot d\mathbf{l} = I \qquad (1.5\text{-}4)$$

If the closed path in this equation lies in a plane normal to a current density $J$ and if the area surrounded by the closed path is very small and can be represented by $\Delta A$, we can divide both sides of the equation by $\Delta A$ and take the limit as $\Delta A \rightarrow 0$ to obtain

$$\oint_{\substack{\text{closed path} \\ \Delta A \rightarrow 0}} \frac{\mathbf{H} \cdot dl}{\Delta A} = J \qquad (1.5\text{-}5)$$

or, since the left-hand side is the definition of the curl of $\mathbf{H}$,

$$| \nabla \times \mathbf{H} | = J$$

and

$$\nabla \times \mathbf{H} = \mathbf{J}$$

where $\mathbf{J}$ is a vector parallel to the flow of current and of magnitude equal to $J$. Ampere's Circuital Law applies when the closed path lies within solid bodies, conductors, or magnetic materials, as well as in regions of free space.

Equation (1.5-4) can be used to obtain the magnetic field intensity at a distance $a$ from the axis of a long cylindrical conductor in free space which conducts a current $I$ amperes parallel to its axis. If the closed path in the equation is a circle of radius $a$ and if the circle is normal to the axis of the conductor with center on the axis, so that $H$ is parallel to the path at all points, we obtain

$$H 2\pi a = I \qquad (1.5\text{-}6)$$

Hence the magnetic flux density $B$ at a distance $a$ from the axis of a long cylindrical conductor, which carries a current $I$ and which is surrounded only by free space, is given by

$$B = \frac{\mu_o I}{2\pi a} \qquad (1.5\text{-}7)$$

Actually the magnetic field generated by a long straight conductor is a vector sum of contributions resulting from each element of length of the conductor. Ampere deduced that when a current $I$ amperes flows in an element of length $dl$ of a conductor, the magnetic flux density $d\mathbf{B}$ at a point $r$ meters from the length $dl$ is given by

$$d\mathbf{B} = \frac{\mu_o I (d\mathbf{l} \times \mathbf{r})}{4\pi r^3} \qquad (1.5\text{-}8)$$

where $d\mathbf{l}$ is a vector of length $dl$ and direction parallel to the current flow. The vector $\mathbf{r}$ is of length $r$ and directed away from the element $dl$ along a line joining $dl$ to the point at which $d\mathbf{B}$ is determined. This result is known as *Ampere's Rule*. It applies only when there is no magnetic material in the

region. With the aid of Figure 1.5-2 it is easily shown that the sum of the contributions to the net magnetic flux density at a point $a$ meters from the axis of a long cylindrical conductor which carries a current $I$ amperes is given by

$$B = \int_{-\pi/2}^{\pi/2} \frac{\mu_o I \, \cos\phi \, d\phi}{4\pi a} = \frac{\mu_o I}{2\pi a} \qquad (1.5\text{-}9)$$

in agreement with Equation (1.5-7). Ampere's Rule is really a special form of the Circuital Law.
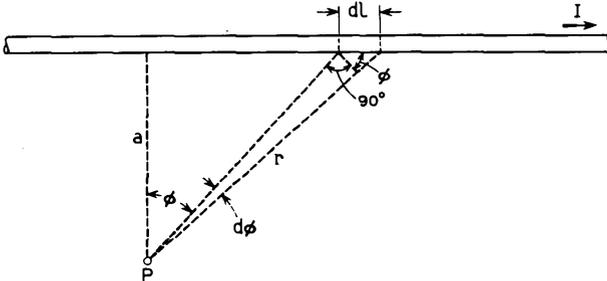


FIG. 1.5-2 A long cylindrical conductor carrying a current $I$ amperes.

Equation (1.5-8) can in principle be used to determine the magnetic flux density at any point in space resulting from a coil of any shape, if sufficient ingenuity is used in carrying out the vector addition of the contributions $d\mathbf{B}$ from each element of current flow.

Perhaps the simplest application of Equation (1.5-8) is the problem of determining the magnetic flux density at the center of a circular loop of wire
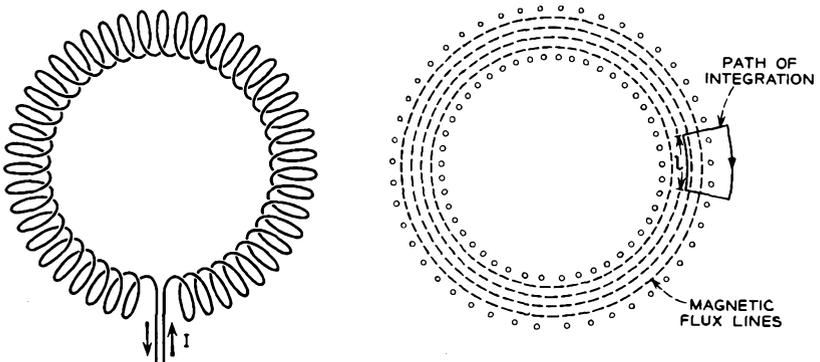


FIG. 1.5-3 Magnetic field lines associated with a toroidal coil which conducts a current $I$ amperes.

which carries a current $I$ and is of radius $a$. In this case the vectors $d\mathbf{B}$ at the center of the loop resulting from each element $dl$ of the loop are all parallel. The total magnetic flux density at the center is easily shown to be

$$B = \frac{\mu_o I}{2a} \qquad (1.5\text{-}10)$$

and is parallel to the axis of the loop.

Figure 1.5-3 shows qualitatively the shape of the magnetic flux lines associated with a toroidal coil. If the turns are close together and regularly spaced, it is evident from symmetry considerations that the magnetic field lines must all lie within the toroid and that $B$ outside the coil is essentially zero. If there are $n$ turns per unit length around the periphery of the coil, application of Ampere's Circuital Law to the path of integration shown in the figure gives

$$Hl = nlI$$

or

$$H = nI \qquad (1.5\text{-}11)$$

where $l$ is the length of the curved part of the path within the toroid. (The only non-zero contribution to the line integral comes from the curved part of the path within the toroid.) The magnetic flux density within the coil is therefore given by $B = \mu_o nI$. This is also the magnetic flux density at the center of a long straight coil of $n$ turns per meter.

The inductance of a coil is equal to the number of "flux linkages" per ampere of current passed through the coil, where the number of flux linkages is equal to the product of the number of webers linking each turn of the coil and the number of turns in the coil. In the case of the toroidal coil shown in Figure (1.5-3), the flux linking each turn of the coil is $\pi r^2 B = \pi r^2 \mu_o nI$, where $r$ is the radius of the turns. If the total number of turns in the coil is $N$, the number of flux linkages per ampere is $\pi r^2 \mu_o nN$, or

$$L = \pi r^2 \mu_o nN \qquad (1.5\text{-}12)$$

where $L$ is the inductance of the coil. In the mks system inductance is measured in henries. If the coil were filled with a medium of relative permeability $\mu$, the inductance would be $L = \pi r^2 \mu \mu_o nN$.

## (b)  Permanent Magnets

A number of metals including the elements iron, nickel, and cobalt, and certain alloys, as well as a group of ceramics called ferrites, exhibit a property known as ferromagnetism. When a long cylindrical rod of one of these materials is placed along the axis of a coil and a current is passed through the coil, the magnetic flux density $B$ within the rod is often hun-

dreds or thousands of times that which would be obtained along the axis
of the coil in the absence of the ferromagnetic material. The ratio of the
magnetic flux density within the sample to that which would be obtained in
free space with the same value of $H$ is known as the relative permeability
of the material and is designated by $\mu$. The magnetic flux density $B$ with-
in the material can therefore be expressed as $B = \mu\mu_o H$, as in Equation
(1.5-3).

Figure 1.5-4(a) shows a coil wound around a toroidal sample of ferro-
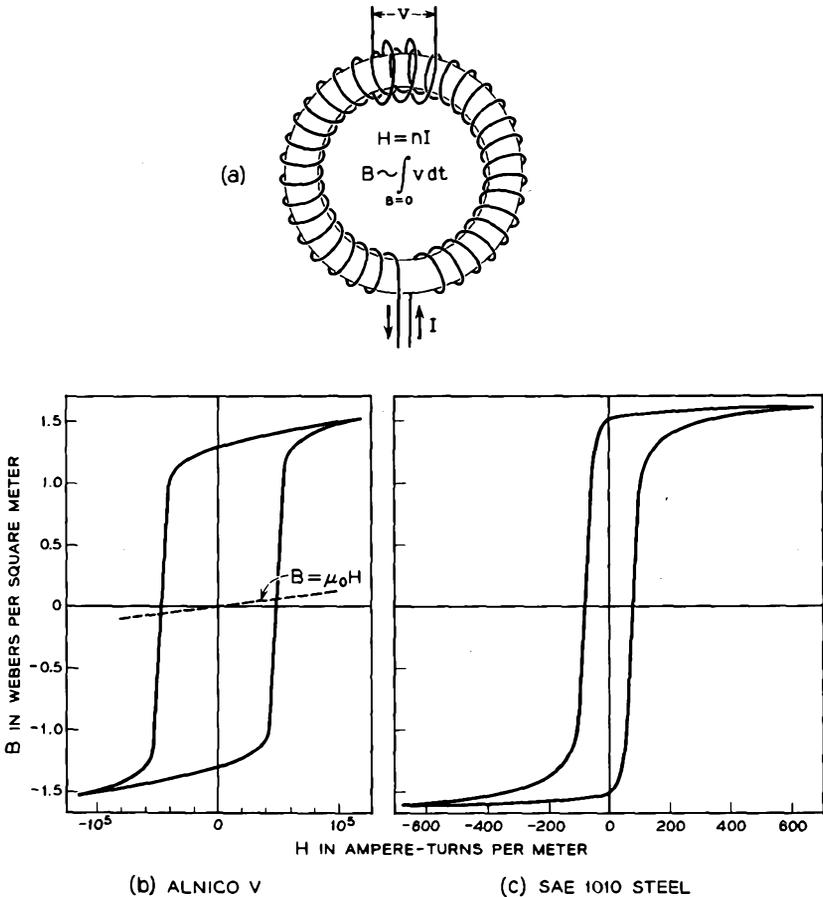


Fig. 1.5-4 A coil surrounding a toroidal sample of ferromagnetic material and
hysteresis loops for two ferrous alloys. Alnico V is frequently used as a permanent
magnet material, and SAE 1010 steel is often used for pole pieces.

(a)

(b)

(c)

(d)

(e)
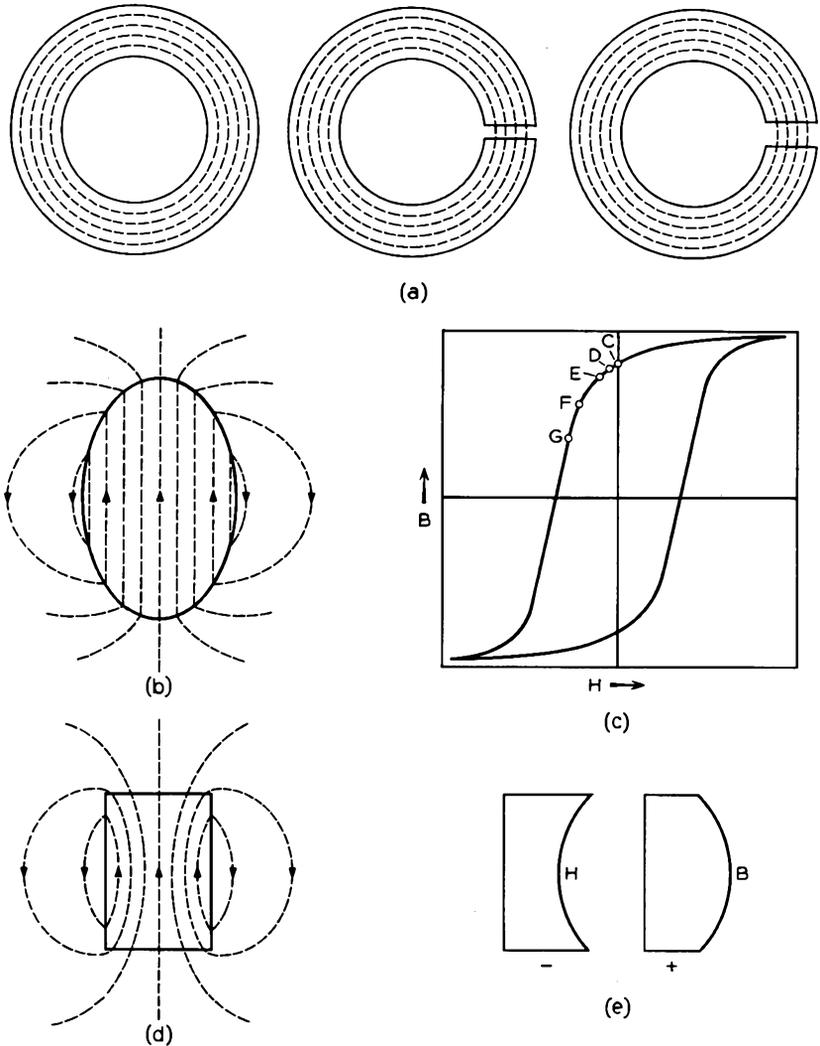
FIG. 1.5-5  The magnetic flux lines associated with several shapes of permanent magnets.  (a) Three toroidal magnets, two with air gaps.  A small amount of flux leakage which would take place from the sides of the two magnets with the air gaps is not shown.  (b) An ellipsoidal magnet.  (c) A hysteresis loop.  (d) A cylindrical magnet.  (e) Plots of $B$ and $H$ along the axis of the cylindrical magnet.

magnetic material.  By passing a current $I$ through the coil, a magnetic field intensity $H = nI$ is established within the sample, where $n$ is the

number of turns per unit length around the periphery of the toroid. If a low-frequency alternating current is passed through the coil, the magnetic flux density[3] $B$ within the material is found to lag the applied $H$. The familiar "hysteresis loop" is a plot of $B$ vs. $H$ obtained in this manner. Two examples of hysteresis loops are shown in Figure 1.5-4. The shape of the hysteresis loop is characteristic of the particular ferromagnetic material. (Notice the difference in the horizontal scale for the two hysteresis loops shown in the figure.) Materials having hysteresis loops with large enclosed areas make the best permanent magnet materials.

Figure 1.5-5(a) shows three toroidal rings of ferromagnetic material. In one the ferromagnetic material forms a closed ring, in one there is a small air gap, and in one there is a larger air gap. We shall assume that each has been "magnetized" by winding a toroidal coil around it and momentarily passing a large current through the coil. When the magnetizing current is removed, the line integral of $H$ around any closed path in the region must be zero, since there is no flow of current in or around the sample. From symmetry arguments we can easily deduce that within the closed ring, $H = 0$, and that $B$ has the value indicated by point $C$ on the hysteresis loop. The flux lines take the form of circles concentric with the axis of the toroid, and all are within the sample. There is no magnetic flux outside the sample.

In the case of the sample with the small air gap, nearly all the lines of flux cross the gap, so that $B$ in the gap is approximately equal to $B$ in the solid. However, since $H$ is parallel to the direction of $B$ in the gap and since the line integral of $H$ along a path following the flux lines must be zero, $H$ must be in the opposite direction to $B$ in the magnetic material. It will be convenient to denote the values of $B$ and $H$ in the air gap with the subscript $g$ and the values of $B$ and $H$ in the magnetic material with the subscript $m$. Then $B_g \approx B_m$. If $H$ is integrated along a path followed by a flux line which crosses the center of the gap, we obtain

$$\oint \mathbf{H} \cdot d\mathbf{l} = lH_g + LH_m = 0 \qquad (1.5\text{-}13)$$

where $l$ is the length of the air gap, and $L$ is the length of the path in the magnetic material. Evidently $H_m$ is small and negative and the values of $B_m$ and $H_m$ might be those corresponding to point $D$ on the hysteresis loop. Since $B$ is positive, it follows from Equation (1.5-3) that $\mu$ for the magnetized toroid is negative. In the case of the sample with the larger air gap, the values of $B_m$ and $H_m$ corresponding to point $E$ might apply. In both samples with the air gap there will actually be "flux leakage" outside the

---

[3]Changes in the magnetic flux density $B$ within the sample are linearly proportional to the time integral of the voltage generated in an auxiliary coil surrounding the sample and can be measured in this manner.

gap since $H$ is not zero within the sample, and $\oint \mathbf{H} \cdot d\mathbf{l}$ must equal zero for all closed paths.

Figure 1.5-5(b) shows qualitatively the shape of the magnetic flux lines associated with an ellipsoidal sample of ferromagnetic material when the sample is magnetized parallel to the long axis of the ellipsoid. It can be shown that, when an ellipsoidal sample is magnetized parallel to one of its axes, the $B$ lines within the sample are all parallel to each other and to the axis. The values of $B_m$ and $H_m$ in this case might correspond to point $F$ on the hysteresis loop.

Figure 1.5-5(d) shows qualitatively the shape of the field lines associated with a cylindrical bar magnet[4]. Some of the flux lines leave the sample through the sides in this case, with the result that $B$ is less at the ends than at the center. Consequently, although the values of $B_m$ and $H_m$ at the center of the magnet might correspond to point $F$ on the hysteresis loop, the values of $B_m$ and $H_m$ at the ends might correspond to point $G$. Figure 1.5-5(e) shows qualitatively the variation of $H$ and $B$ along the axis of the bar magnet.

From the foregoing discussion it is apparent that the operating point on the hysteresis loop is determined by the geometry of the permanent magnet. To illustrate this point further, let us return to the two toroidal magnets with air gaps illustrated in Figure 1.5-5(a). If it is assumed that all the lines of $B$ cross the gap and that there is no flux leakage from the sides of the magnet, then

$$B_m = B_g = \mu_o H_g \qquad (1.5\text{-}14)$$

Combining this with Equation (1.5-13), we obtain

$$\frac{B_m}{H_m} = -\frac{\mu_o L}{l} \qquad (1.5\text{-}15)$$

This defines the slope of a line through the origin of the coordinate system for the hysteresis loop, and the intersection of this line with the hysteresis loop defines the operating point for $B_m$ and $H_m$.

Since $\oint \mathbf{H} \cdot d\mathbf{l} = 0$ for all closed paths in the neighborhood of a permanent magnet, it is possible to define a magnetic potential $\psi$ such that the potential difference between points $A$ and $B$ is given by $\psi_{AB} = -\int_A^B \mathbf{H} \cdot d\mathbf{l}$. (The magnetic potential difference between two points is often called the magnetomotive force, or mmf, in analogy to the electromotive force, or emf, in electrostatics.) The magnetic field intensity is related to the magnetic potential $\psi$ by $H = -\nabla \psi$. Since $\mathbf{B} = \mu_o \mathbf{H}$ in the region *outside* a per-

---

[4]After M. Abraham, R. Becker, *Classical Theory of Electricity and Magnetism*, p. 137, Blackie and Son, 1932.

manent magnet, and since $\nabla \cdot \mathbf{B} = 0$ and $\nabla \cdot (\nabla) = \nabla^2$, the magnetic potential in the space surrounding a permanent magnet satisfies Laplace's Equation, $\nabla^2 \psi = 0$.

Magnetic fields are used to focus, or confine, the electron beams of a number of microwave tubes including traveling-wave tubes, klystron amplifiers, and backward-wave oscillators. Magnetic fields also play an
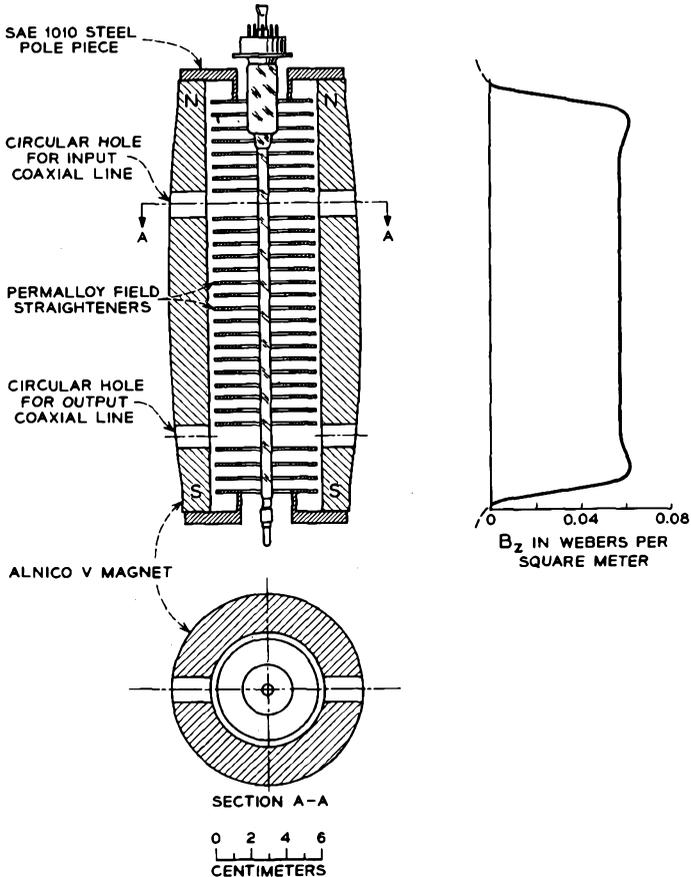


FIG. 1.5-6 A permanent magnet circuit used to focus the electron beam of a traveling-wave tube. The outline of the tube is shown in the figure. A plot of the axial magnetic field $B_z$ is shown at the right. The slight peaking of the axial magnetic field near the ends of the circuit results from the "re-entrancies" in the pole pieces. Within the pole pieces the axial magnetic field changes direction, and beyond the pole pieces the axial magnetic field has the opposite direction to that which it has in the center of the magnet.

essential role in the operation of magnetron oscillators. By using permanent magnets rather than electromagnets to provide the magnetic field, the total power consumption of the tubes can be reduced.

Figure 1.5-6 shows a permanent magnet circuit for a traveling-wave tube. The circuit produces a magnetic flux density[5] of nearly 0.06 weber/meter$^2$ along the axis of the tube in the region between the pole pieces. The magnetic flux density $B$ in the pole pieces is well below that needed to saturate the pole piece material, so that $H$ within the pole pieces is extremely small (see hysteresis loop for SAE 1010 steel in Figure 1.5-4). The pole pieces, therefore, serve as equipotential bodies, the mmf being nearly constant throughout their volume. In a similar manner, the permalloy "field straighteners" are flat discs of high-permeability steel which serve as equipotential planes and assure that the lines of $B$ are parallel to the axis of the traveling-wave tube. Since $B = \mu\mu_oH$, and $\mu$ is very large for the field-straightener material, $H$ within the field straighteners is correspondingly small. The permanent magnet is larger at its center than at its ends to account for flux leakage from its sides.
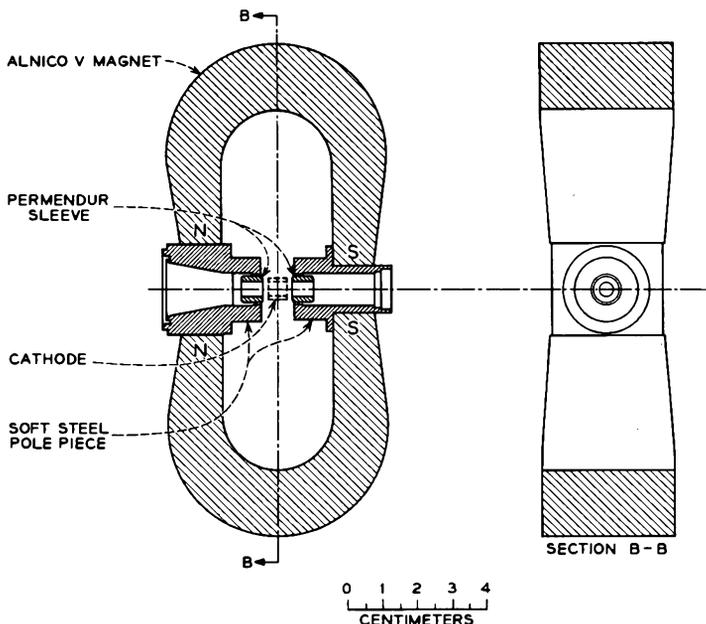


FIG. 1.5-7 A permanent magnet circuit for a magnetron.

---

[5]One weber per square meter = 10$^4$ gauss.

Figure 1.5-7 shows a permanent magnet circuit for a magnetron. The circuit produces a magnetic flux density of about 0.5 weber/meter² in the neighborhood of the magnetron's cathode. The permendur sleeves inside the pole pieces serve to shape the magnetic field in the region between the cathode and anode so as to obtain electron trajectories which give optimum interaction between the electrons and the rf field.

*PROBLEMS*
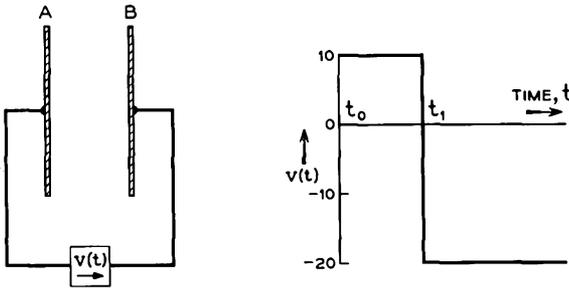


Problem 1.1

1.1 At time $t_o$ a single electron is emitted from electrode $A$ with zero velocity, and at this time a voltage $V = +10$ volts is applied between the electrodes in such a direction that it accelerates the electron toward electrode $B$. It is assumed that the electric field intensity is uniform at all points between the electrodes. At time $t_1$ the electron is halfway to electrode $B$, and the voltage $V$ changes discontinuously to $-20$ volts and remains at that value. Which electrode does the electron strike, and what is its kinetic energy in electron volts when it strikes the electrode?
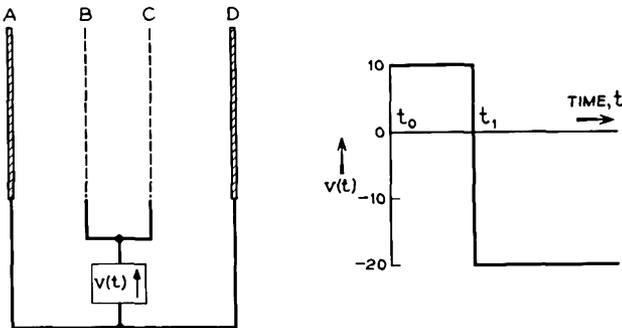


Problem 1.2

1.2 Grids $B$ and $C$ are assumed to be ideal grids having the properties that they do not intercept electrons and that field lines do not penetrate through the grids.

A single electron leaves electrode $A$ with zero velocity at time $t_o$. At this time the voltage $V$ is $+10$ volts and is in such a direction that the electron is accelerated toward grid $B$. At time $t_1$ the electron is midway between grids $B$ and $C$, and the voltage $V$ changes to $-20$ volts. Which electrode (either $D$ or $A$) does the electron strike, and what is its kinetic energy in electron volts when it strikes the electrode?
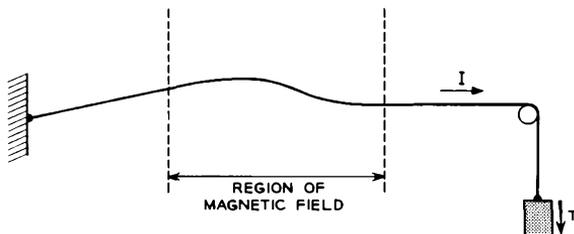


Problem 1.3

1.3 A very fine wire is held stationary at one end, while the other end passes over a pulley and is fixed to a weight which maintains a tension $T$ newtons in the wire. Over a limited region between the fixed end of the wire and the pulley there is a magnetic field that varies across the region both in magnitude and direction. If a current $I$ amperes is passed through the wire, the magnetic field causes a force to act on the wire which tends to deflect it. The force is equal to $BI$ newtons per meter length of the wire and acts in the direction normal to both the current flow and the magnetic field. The resulting shape of the wire might be that shown in the figure. Suppose that the wire were removed and that an electron were directed toward the magnetic field along the path previously followed by the wire. Show that, if the electron momentum $mu$ satisfies the relation $mu/e = T/I$, the electron trajectory through the region of the field will coincide with the path followed by the wire. Assume that the stiffness of the wire can be neglected and that its mass is negligible.
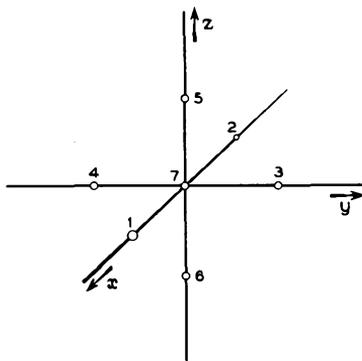


Problem 1.4

1.4 Points 1, 7, and 2 lie on the $x$ axis of a rectangular coordinate system. Points 3, 7, and 4 lie on the $y$ axis, and points 5, 7, and 6 lie on the $z$ axis. The distance from

point 7 to each of its neighboring points is $d$ meters. The region is filled with a uniform charge density $\rho_o$ coulombs/meter³. Show that if the distance $d$ is very small, the potential at point 7 is approximately given by

$$V_7 = \frac{V_1 + V_2 + V_3 + V_4 + V_5 + V_6}{6} + \frac{d^2\rho_o}{6\varepsilon_o}$$

where $V_1$ is the potential at point 1, and so on. What effect does the presence of space charge have on the potential at point 7?

1.5 Use Equation (1.4-14) to show that the potential at the center of an isolated spherical cloud of charge of radius $R$ and uniform charge density $\rho_o$ is given by

$$V_{\text{center}} = \frac{\rho_o R^2}{2\varepsilon_o} = \frac{3q}{8\pi\varepsilon_o R}$$

where $q$ is the total amount of charge in the cloud.



(a)                                                        (b)
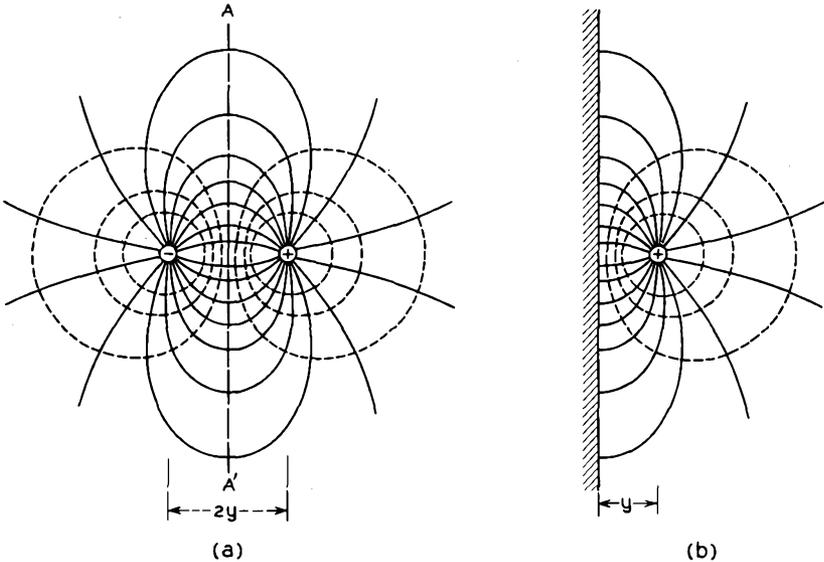
Problem 1.6

1.6 Part (a) of the figure shows qualitatively the field lines associated with two isolated point charges $+q$ and $-q$. The plane $A$-$A'$ lies midway between the two point charges. Since all points on the plane are equidistant from the two point charges, the potential on the plane is zero. Both charges contribute to the electric field intensity at the plane $A$-$A'$. Show that the total electric field intensity at the plane can be expressed as

$$E = \frac{qy}{2\pi\varepsilon_o(r^2 + y^2)^{3/2}}$$

where $y$ is the distance from the point charges to the plane $A$-$A'$, and $r$ measures the distance along the surface of the plane from the line joining the point charges to the point at which $E$ is determined. The electric field intensity at the plane $A$-$A'$ is, of course, normal to the plane.

Since all points on the plane $A$-$A'$ are at zero potential, a thin planar conductor could be inserted along the plane without disturbing the potential and field distribution in the region. Suppose such a planar conductor were inserted and the left-hand charge were then removed. Evidently the right-hand half of the field pattern would remain unchanged. Hence the field distribution shown in part (b) of the figure must be that which applies when a point charge $+q$ is $y$ meters from planar conductor. Field lines originating on the charge $+q$ terminate on negative induced charges on the surface of the conductor. Use the above expression for $E$ to obtain an expression for the surface charge density induced on the planar conductor by the charge $+q$. Show that the total induced charge is equal to $-q$.

Show that the force tending to draw the charge $+q$ toward the planar conductor in part (b) of the figure is $q^2/[4\pi\varepsilon_o(2y)^2]$ newtons and that the work required to remove the charge $+q$ to infinite distance from the planar conductor is $q^2/[4\pi\varepsilon_o(4y)]$ joules.

1.7 A dc current $I$ amperes flows within a long cylindrical conductor of radius $R$. The current density is assumed to be uniform across the wire and directed parallel to the axis. Sketch qualitatively how the magnetic flux density $B$ varies with radius $r$ from the axis of the wire out to several times $R$. Make a similar sketch for the radial electric field intensity associated with a cylindrical beam of electrons. Assume uniform space charge density across the beam cross section.

## REFERENCES

Several general references covering electron motion in fields and properties of static electric and magnetic fields are listed below:

1a. W. G. Dow, *Fundamentals of Engineering Electronics*, 2nd Ed., Chapters 1 and 2, John Wiley and Sons, Inc., New York, 1952
1b. J. Millman and S. Seely, *Electronics*, 2nd Ed., Chapters 2 and 3, McGraw-Hill Book Co., Inc., New York, 1951.
1c. W. W. Harman, *Fundamentals of Electronic Motion*, Chapters 1, 2, and 4, McGraw-Hill Book Co., Inc., New York, 1953.
1d. J. R. Pierce, *Theory and Design of Electron Beams*, 2nd Ed., Chapters 1, 2, and 3, D. Van Nostrand Co., Inc., Princeton, N. J., 1954.
1e. K. R. Spangenberg, *Vacuum Tubes*, Chapters 5 and 6, McGraw-Hill Book Co., Inc., New York, 1948.
1f. J. D. Kraus, *Electromagnetics*, Chapters 1 and 2, McGraw-Hill Book Co., Inc., 1953.

# Chapter 2

## ELECTRON EMISSION

The great majority of electron tubes depend upon thermionic emission as their source of free electrons. In this emission process, electrons within a conductor or semiconductor receive sufficient energy by thermal excitation to overcome the forces tending to keep them within the solid.

Our interest in thermionic emission at this point lies in the fact that some inherent properties of thermionic emission seriously affect the design and performance of electron tubes. It is found, for instance, that the emitted electrons have small, but finite, velocities upon emission, so that in considering the shapes of the electron trajectories in the interelectrode space of a tube we must take into account the distribution of emission velocities. When we try to design an electron gun that will produce a thin beam of electrons with high current density, we find that we are seriously limited in doing so by the finite electron emission velocities.

It is also found that both the current of emitted electrons and the velocity distribution of the emitted electrons fluctuate with time. These fluctuations constitute two principal sources of "noise" in electron tubes and lead to serious limitations in the performance of many amplifier tubes.

We are further limited in designing an electron tube by the fact that each thermionic cathode material is characterized by a maximum emission current density consistent with long life of the emitter. Since the performance requirements of a tube are generally such that a certain total current must be drawn from the cathode, the maximum emission current density serves to determine the minimum area of the cathode emitting surface. In grid-controlled tubes this minimum cathode area usually determines the area of the remaining electrodes. The high-current-density beams used in klystrons and traveling-wave tubes are frequently obtained by making use of a relatively large cathode emitting surface in order to draw the required total emission current and by using electrostatic fields

30

to focus the electrons to a beam of smaller diameter. In later chapters we shall find that a number of the ultimate performance limitations of vacuum tubes result from the need for using a large cathode emitting surface.

Finally, some properties of the oxide-coated cathode are of concern to the tube designer and tube user. It is found, for instance, that the resistance of the oxide coating is often of the order of a few ohms across a square centimeter of coating. This resistance is effectively inserted between the cathode lead and the emitting surface. In the case of grid-controlled tubes, it serves as a negative feedback mechanism which has the effect of reducing the transconductance of the tube. Furthermore, during the processing of a tube and during the life of the tube, compounds form at the interface between the oxide coating and the base metal. These compounds place an additional impedance between the cathode lead and the emitting surface. The interface resistance is often ten or more times that of the oxide coating, and it increases with the life of the tube. End of life for many grid-controlled tubes occurs when the growth of interface resistance reduces the transconductance of the tube below a minimum useful value.

These limitations are of sufficient importance to merit further consideration of the thermionic emission process, and accordingly a major part of the present chapter will be concerned with this subject. (Discussion of noise in electron beams and its excitation by current and velocity fluctuations at the cathode will be deferred until Chapter 13, however.)

Electrons can be emitted from a solid by processes other than thermionic emission. Whenever electrons near the surface of a solid are given sufficient energy to overcome the forces tending to keep them within the solid, some of the electrons escape. Excitation of the electrons near the surface can be caused by incident electrons, photons, positive ions, or excited atoms, and each of these means of excitation can lead to electron emission. Electron emission also can be caused by the application at the surface of sufficiently high electric fields that the surface forces are reduced to the point where electrons escape. Electron emission resulting from electrons striking a surface is known as secondary emission; emission resulting from photons striking a surface is known as photoelectric emission; and emission resulting from the application of very high electric fields is known as field emission, or Schottky effect. Some characteristics of secondary emission and photoelectric emission will be described in Sections 2.5 and 2.6, respectively. Emission caused by positive ions and excited atoms striking a surface is of interest in connection with gas-discharge devices, and we shall defer discussion of it until Chapter 14. Field emission has found only limited application in electron-tube work[1] and will not be described further.

---

[1]Some special-purpose cathode-ray tubes and x-ray tubes have field-emission cathodes. Certain arc-discharge processes are also known to depend upon field emission.

We shall begin by describing briefly the behavior of electrons in the interior of a metal and the forces that act on electrons at the surface of a metal.

## 2.1   Electrons in a Conductor, Work Function, and Contact Potential

### (a)   Electrons in a Conductor

It is well known that the structure of metals is crystalline. Three crystal structures frequently formed by the metallic elements are illustrated in Figure 2.1-1. These are the body-centered cubic structure, the face-



|      | $\underline{d}$ |
|------|------|
| Na   | 3.67Å |
| K    | 4.62Å |
| W    | 2.74Å |

VIEWED FROM ABOVE

(a)

|      | $\underline{d}$ |
|------|------|
| Al   | 2.86Å |
| Ni   | 2.49Å |
| Cu   | 2.55Å |

(b)

|      | $\underline{d}$ | $\underline{d'}$ |
|------|------|------|
| Be   | 2.28Å | 2.23Å |
| Mg   | 3.20Å | 3.19Å |
| Zn   | 2.66Å | 2.91Å |

(c)
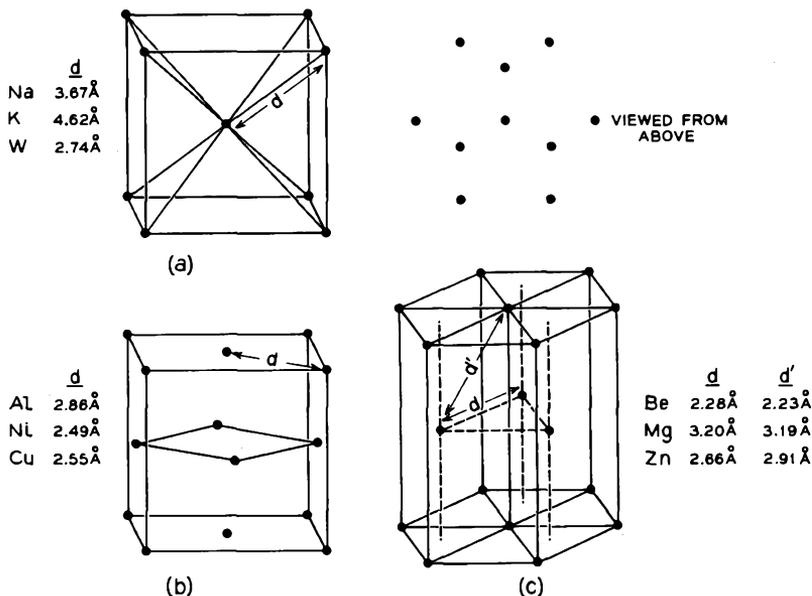
FIG. 2.1-1 Three common metallic crystal structures: (a) body-centered cubic structure; (b) face-centered cubic structure; (c) hexagonal structure.

centered cubic structure, and the hexagonal structure. The atomic spacings of several metallic elements that form these structures are shown in Figure 2.1-1. Practically all nearest-neighbor spacings of the atoms in metallic crystals lie between 2 and 5 angstroms (1 angstrom = $10^{-8}$ cm).

Let us now look at the electronic structure of the isolated atoms. An atom is always characterized by certain discrete total energies which its electrons can have. Each electron in the atom has one of these energies and therefore is said to be in an energy state of the atom. With each energy state there is associated a certain characteristic motion of the electron

about the nucleus. For many purposes, an electron in a given energy state can be thought of as causing a cloud of charge about the nucleus. The probability of finding the electron in a given volume element about the nucleus is proportional to the charge density of the cloud at the volume element. It is found that electrons in certain energy states tend to contribute their maximum charge density at approximately the same distance from the nucleus, and consequently there are said to be shells of electrons about the nucleus. Thus, copper has two electrons in its innermost shell, eight electrons in the next shell, eighteen in the next, and one in the outer shell. Metallic atoms have one to four electrons in the outer shell. Generally, the radius at which the outer-shell electrons in the isolated atoms contribute their maximum charge density is nearly half the nearest-neighbor spacings of atoms in the metallic crystal and at least several times the radius at which the electrons in the next smaller shell contribute their maximum charge density.[2]
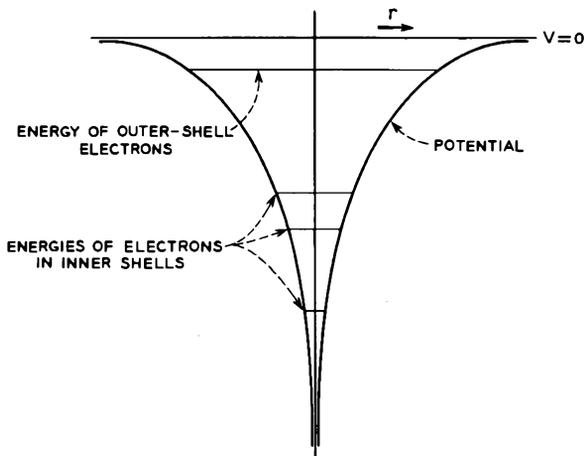


FIG. 2.1-2 The potential experienced by an atomic electron as a function of radius $r$ from the nucleus of the atom.

Figure 2.1-2 shows qualitatively how the potential that acts on an electron in an isolated atom varies with distance $r$ from the nucleus. As the electron travels to large distances from the nucleus, so that it is outside the charge clouds of the other electrons, it leaves behind a net charge of $+e$ on the atom and experiences a potential $-e/4\pi\varepsilon_0 r$ volts. At smaller distances, such that the electron is within the charge clouds of the other electrons, its

[2]Reference 2.1, p. 349, lists the radii at which electrons in the various energy states of the lighter atoms contribute their maximum charge density.

potential is complicated by interactions with other electrons. When the
electron is inside the charge clouds of all the other electrons, its potential
approaches $-(Ze/4\pi\varepsilon_o r) + C$, where $Ze$ is the positive charge on the
nucleus, and $C$ is a constant.
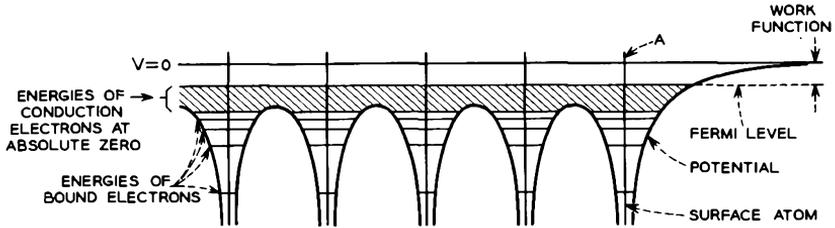


FIG. 2.1-3 The potential experienced by an electron in a metallic crystal. The
potential is plotted along a line running through the centers of several of the atoms.
The vertical lines in the figure indicate the locations of the atomic centers. The
conduction band might extend from the bottom of the shaded region to well above
$V = 0$.

Figure 2.1-3 shows qualitatively how we may expect the potential to
vary along a line of atoms in a crystal. Clearly, the potential will be
periodic with extreme local variations near the nuclei. The crystal may be
thought of as being made of a lattice of positive charge centers with the
space between the charge centers filled with a cloud of negative charge.
The positive charge centers are the metal atoms minus their outer-shell
electrons, and the cloud of negative charge arises from the outer-shell
electrons. At the interatomic spacings that atoms assume in a crystal, the
charge clouds of the outer-shell electrons overlap appreciably, whereas
very little overlapping takes place for the charge clouds of the electrons in
the inner shells.

The overlapping of the outer-shell charge clouds causes a broadening of
the possible energies that the outer-shell electrons can have into a band of
energy states, known as the conduction band. Each energy state in the
conduction band can be occupied by a maximum of two electrons. In a
metal not all the energy states in the conduction band are filled. At ab-
solute zero only the lower part of the band is filled, while the higher energy
states are unoccupied. If there are $N$ outer-shell electrons in a piece of
metal which is at absolute zero, the outer-shell electrons occupy the $N/2$
lowest energy states in the conduction band, there being two electrons in
each state. However, when the metal is at room temperature or higher,
a few of the electrons are excited to higher states by thermal excitation.
(In Figure 2.1-3 the conduction band would extend from the lower part
of the shaded region to well above $V = 0$.)

Excitation to the higher states also takes place when an electric potential gradient is established within the metal, and it can be shown that the existence of unfilled states just above the occupied part of the band is an essential condition for electric conduction. It is the outer-shell electrons that are responsible for electrical conductivity, for they find themselves relatively free to drift through the crystal under the influence of an applied electric field, while those in the inner shells remain bound to their atoms.

## (b)  Work Function

Next let us consider the forces acting on an electron at the edge of a metallic crystal. The letter $A$ in Figure 2.1-3 marks the position of a surface atom. An electron moving to the right from $A$ would at first experience a potential similar to that in an isolated atom. However, at somewhat larger distances from $A$, the main effect results from a force called the image force. A well-known problem in electrostatics[3] shows that a point charge $+q$ located $y$ meters from the surface of a conducting plane is acted on by a force directed toward the plane and equal in magnitude to that which the charge would experience from an equal and opposite charge $2y$ meters away. The actual force arises from an induced surface charge $-q$ on the conducting plane. In the case of an electron $y$ meters from a planar conductor, the potential arising from the image force is $-e/16\pi\varepsilon_o y$ volts. If $y$ is measured in angstroms, this becomes $-3.6/y$ volts.

As the distance $y$ decreases and approaches the interatomic spacing $d$, the concept that the surface is a planar conductor becomes no longer valid, and the potential merges with that arising from the surface atoms. The location of the region over which the merging takes place depends upon the number of atoms per unit area of the crystal surface and upon the shape of the outer-shell electron clouds. Consequently, we would expect that the height to which the potential curve rises above the filled part of the conduction band will differ for crystals of different metals.

In Figure 2.1-3 we have shown the outer-shell electrons, or conduction electrons, to have energies distributed over a band having a definite maximum and minimum. Such is the case at absolute zero, and it is almost the case at room temperature. However, as already noted, when the metal is at room temperature or higher, a number of the electrons are excited to states just above the part of the conduction band that is filled at absolute zero. If the temperature is sufficiently high, a few electrons gain sufficient energy to overcome the image potential and leave the surface of the metal. This is the basis of thermionic emission.

------

[3]See Problem 1.6.

The potential corresponding to the top of the part of the conduction band that is filled at absolute zero is frequently called the Fermi level. The work function of a metal is the energy that must be given to an electron at the Fermi level to enable it to escape from the metal with zero velocity. In the discussion that follows, we shall designate the work function by $\phi$ and assume that it is measured in electron volts. (*The work function in electron volts is numerically equal to the potential rise in volts from the Fermi level to the potential $V = 0$ shown in Figure* 2.1-3.)

The thermionic emission current density from a surface is closely related to the work function of the surface. With a large work function, the electrons must be excited to higher energy states in the conduction band in order to be able to escape, and the emission current density for a given cathode temperature will be lower.

Table 2.1-1 lists the work functions of several metals.[4] The work function

TABLE 2.1-1. MEAN WORK FUNCTIONS

| Metal | $\phi$ (Electron Volts) | Metal | $\phi$ (Electron Volts) |
|-------|------------------------|-------|------------------------|
| Ag | 4.28 | Mo | 4.27 |
| Al | 3.74 | Na | 2.27 |
| Au | 4.58 | Ni | 4.84 |
| Ba | 2.29 | Pd | 4.82 |
| C | 4.39 | Sb | 4.08 |
| Ca | 2.76 | Sr | 2.35 |
| Cb | 3.99 | Ta | 4.12 |
| Cs | 1.89 | Th | 3.41 |
| Cu | 4.47 | Ti | 4.09 |
| Fe | 4.36 | W | 4.50 |
| K | 2.15 | | |

of a clean metal surface is always of the order of a few electron volts, and the energy separation between the bottom of the conduction band and the Fermi level is also of about this magnitude.

Actually, $\phi$ is different for different faces of a metal crystal. The variation arises in part from the fact that the density of surface atoms changes from face to face, and the distance from the surface atoms at which the image potential merges into that of the surface atoms shows a corresponding variation. In the case of tungsten, the measured values for the different faces[5] range from 4.30 electron volts to nearly 6 electron volts. The values listed in Table 2.1-1 are for polycrystalline surfaces.

---

[4]Reference 2.2.
[5]Reference 2.3.

(c)   *Contact Potential*

It can be shown that if two dissimilar metals are brought together at absolute zero, the potentials within the metals immediately become adjusted so that the Fermi levels in the two systems coincide. This situation is illustrated in Figure 2.1-4 for two metals at absolute zero. The figure would
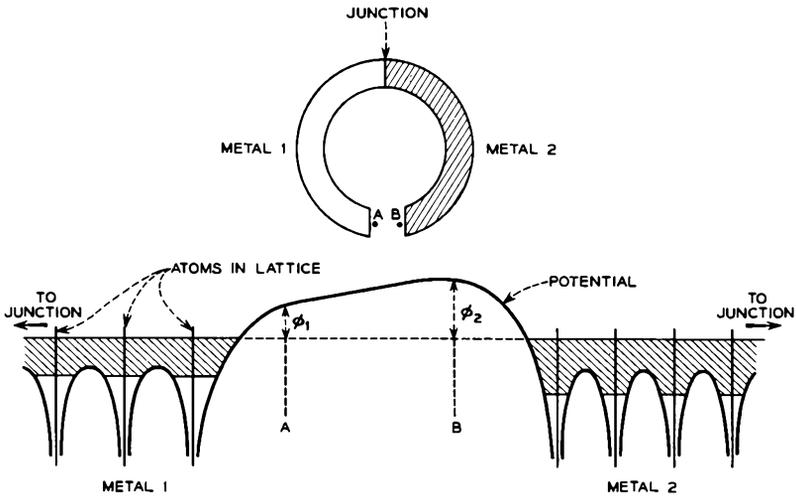


FIG. 2.1-4 Two metals joined at absolute zero. The symbols $\phi_1$ and $\phi_2$ indicate potential differences which are numerically equal to the work functions of the two metals.

be the same at higher temperatures except that the tops of the filled parts of the conduction bands would not be well defined, since some of the electrons are excited to higher energies. If the two metals have different work functions, the potential rises to different heights outside the two metals, and a potential difference must exist between a point just outside one metal and a point just outside the other. From the definition of the work function, we can see that this contact potential difference, as it is called, is numerically equal to the difference between the work functions of the two metals.

Since there is a potential difference between points $A$ and $B$ in Figure 2.1-4, this can result only from the presence of induced surface charges on the conductors. The potential in the region between the two metals is therefore a superposition of the potential caused by the image force and that arising from the induced surface charge. The surface charges result from a flow of charge between the metals that takes place when the metals

are first joined. The flow of charge is such that the material of lower work function is positively charged.

If several wires made of different metals are connected in series, the potential difference in volts between a point just outside the wire on one end and a point just outside the wire on the other end is numerically equal to the difference between the work functions in electron volts of the two end metals. Let us suppose that two parallel metal plates made of different metals form a parallel-plate capacitor and that the plates are joined by a wire. If $\phi_1$ and $\phi_2$ are the work functions of the plates in electron volts, the potential difference in volts between a point just outside one plate and a point just outside the other will be numerically equal to $\phi_1 - \phi_2$. We shall designate it $V_{12}$. If $d$ is the spacing between the plates, the electric field intensity in the region between the plates will be $V_{12}/d$, neglecting edge effects, and the surface charge density $\sigma$ on the two surfaces that face each other will be given by

$$\sigma = \varepsilon_o \frac{V_{12}}{d} \qquad (2.1\text{-}1)$$

Evidently as the distance between the plates is varied, charge must flow along the wire joining them.

When a voltage difference is applied between two electrodes of a tube, the electric field intensity in the interelectrode space effectively results from the sum of the applied potential difference and the contact potential difference. Since the contact potential may amount to two or three volts, there are many cases where it cannot be overlooked.

If the cathode of a tube is oxide-coated, it will likely have the lowest work function of any of the electrodes. Thus, if all the electrodes in such a tube are directly connected together, the emitted electrons experience a retarding field that returns almost all of them to the cathode. Part of the "aging" or drift in the electrical characteristics of some grid-controlled tubes with life has been attributed to a change in the contact potential difference between the control grid and cathode as a result of a gradual contamination or decontamination of the grid surface.

## 2.2   The Richardson-Dushman Equation

A derivation of the law governing the thermal emission of electrons from a metal[6] involves results that are closely related to the physics of solids and hence represents a departure from most of the work that will be emphasized in later chapters. In brief summary, an expression can be derived for the

[6]See Reference 2.4, p. 137.

number of electrons per unit volume of the metal which at temperature $T$ are excited to states characterized by sufficient energy to enable the electrons to overcome the surface forces. (In Figure 2.1-3, these states would lie above the horizontal line corresponding to $V = 0$.) The current of electrons in these states striking unit area of surface from within the metal is assumed to be the emission current density and is found to be

$$J_o = \frac{4\pi mek^2}{h^3}T^2\epsilon^{-\phi/W_T} \text{ amps/meter}^2 \qquad (2.2\text{-}1)$$

where $h$ is Planck's constant, $k$ is Boltzmann's constant, $T$ is the absolute temperature of the emitting surface, and $\phi$ is the work function of the metal in electron volts. $W_T$ is the "electron-volt equivalent" of the energy $kT$ and is given by

$$W_T = \frac{kT}{|e|} = \frac{T}{11,600} \text{ electron volts} \qquad (2.2\text{-}2)$$

where $|e|$ is a positive dimensionless constant numerically equal to the charge of the electron. Equation (2.2-1) is known as the Richardson-Dushman Equation for the emission current density. The equation can be expressed more conveniently in the form

$$J_o = AT^2\epsilon^{-\phi/W_T} \text{ amps/meter}^2 \qquad (2.2\text{-}3)$$

where

$$A = \frac{4\pi mek^2}{h^3} = 120 \times 10^4 \text{ amps/meter}^2(°K)^2$$
$$= 120 \text{ amps/cm}^2(°K)^2$$

To a first approximation, $\phi$ is independent of temperature. However, experimental evidence indicates that it has a small temperature coefficient, and hence that it can be expressed as $\phi = \phi_o + \alpha T$, where $\phi_o$ is the work function at absolute zero. (Measurements of the coefficient $\alpha$ for tungsten[7] indicate that it is of the order of a few times $10^{-5}$ electron volt/$°K$.) Substituting $\phi = \phi_o + \alpha T$ in Equation (2.2-3), we obtain

$$J_o = AT^2\epsilon^{-(\phi_o+\alpha T)/W_T} = A\epsilon^{-\alpha|e|/k}T^2\epsilon^{-11,600\phi_o/T}$$
$$= A'T^2\epsilon^{-11,600\phi_o/T} \qquad (2.2\text{-}4)$$

where use has been made of the relations $|e|W_T = kT$ joules and $W_T = T/11,600$ electron volts. A non-zero coefficient $\alpha$, therefore, has the effect of modifying the constant $A$ in the Richardson-Dushman Equation.

Experimental values[8] of $A' = A\epsilon^{-\alpha|e|/k}$ for clean surfaces of several metals

[7]Reference 2.5.
[8]Reference 2.6.

TABLE 2.2-1

| Metal | $A'$<br>$(Amp/cm^2(°K)^2)$ | $\phi$<br>(Electron Volts) |
|---|---|---|
| Mo.................55 | | 4.27 |
| Ni.................30 | | 4.84 |
| Pt.................32 | | 5.29 |
| Ta.................37 | | 4.12 |
| W..................70 | | 4.50 |

are given in Table 2.2-1, along with the corresponding work functions from Table 2.1-1. (It should be noted that accurate measurement of $A'$ is extremely difficult, since a small error in the absolute temperature $T$ can cause a large error in $A'$.)

Equations (2.2-3) and (2.2-4) indicate a critical dependence of the emission current density upon both the work function $\phi$ and the temperature $T$ of the emitting surface. Decreasing the work function by one electron volt increases the emission current density by $\epsilon^{11.600/T}$ for the same temperature $T$. Table 2.2-2 lists values of the emission current density in

TABLE 2.2-2. EMISSION CURRENT DENSITY
AMPS/Cm²

| $T - °K$ | $\phi = 1.0$ ev | $\phi = 2.0$ ev | $\phi = 3.0$ ev | $\phi = 4.0$ ev | $\phi = 5.0$ ev |
|---|---|---|---|---|---|
| 1000 | 360 | $3.3 \times 10^{-3}$ | $3 \times 10^{-8}$ | — | — |
| 1500 | — | 17 | $7.6 \times 10^{-3}$ | $3 \times 10^{-6}$ | $1.4 \times 10^{-9}$ |
| 2000 | — | — | 4.4 | $1.3 \times 10^{-2}$ | $4 \times 10^{-5}$ |
| 2500 | — | — | 230 | 2.1 | $2.1 \times 10^{-2}$ |

amperes per square centimeter for several values of $\phi$ and $T$ assuming that $A' = 40$ amps/cm²(°K)². Evidently cathodes with a high work function must be operated at a high temperature in order to obtain an appreciable emission current density.

The emission current density given by Equations (2.2-3) and (2.2-4) is often referred to as the saturation emission current density. Under most operating conditions less current is actually drawn from the cathode. If more electrons are emitted from the cathode than are drawn to the other electrodes, negative space charge accumulates just outside the cathode. This causes a small retarding field at the cathode surface which returns some of the emitted electrons to the cathode. The current drawn from the cathode is then said to be space-charge-limited. If full saturation emission current flows to the other electrodes, the current drawn from the cathode is said to be temperature limited, since the cathode temperature then determines the magnitude of the current flowing to the other electrodes.

## 2.3   Cathode Materials

The choice of cathode materials to be used in a particular tube is determined by such factors as the environment in which the cathode is to be operated, the required emission current density, the tube life requirements, the temperature at which the cathode must be operated to obtain the emission density, and the power which must be supplied to heat the cathode. Generally, the choice of materials which will satisfy a particular application is quite limited.

As a first consideration, all cathode materials are characterized by a maximum ‘operating temperature consistent with long life of the emitting surface.  This means that they are characterized by a maximum emission current density consistent with long cathode life.  In the case of pure metal cathodes, such às tungsten filaments, the limiting temperature is that at which evaporation of the metal starts to be appreciable.  In the various other forms of cathodes to be described in this section, additional chemical and physical processes tend to limit the life of the cathode when too high temperatures are used.  Since for many applications it is desirable to have available a high emission current density and since the emission current density increases rapidly with temperature, the cathode operating temperature is often determined as a compromise between requirements for high emission density and requirements for long cathode life.  Of the relatively large number of possible cathode materials that have been investigated,[9] only a very few are capable of simultaneously giving appreciable emission current density and good life performance.

A second important consideration is the desirability of operating the cathode at as low a temperature as possible.  A low cathode operating temperature means low heater or filament power and greater power efficiency for the tube.  If less heat is dissipated in the tube, less heat is radiated and conducted from the tube, and there is less heating of the surrounding apparatus.  Furthermore, as the cathode warms up, thermal expansion of the cathode and its supports frequently causes a change in the electrode spacings and hence a change in the electrical characteristics of the tube. With a low cathode operating temperature, it is much easier to minimize these changes in spacing, and greater reproducibility of the electrical characteristics from device to device can be obtained.  Finally, the noise which appears in the tube output, and which results from fluctuations in the emission current and velocities, is less when the required emission current density is obtained at a lower operating temperature.

Since a cathode with a low work function can provide a given emission current density at a lower operating temperature than one with a high work

---

[9]Reference 2.6.
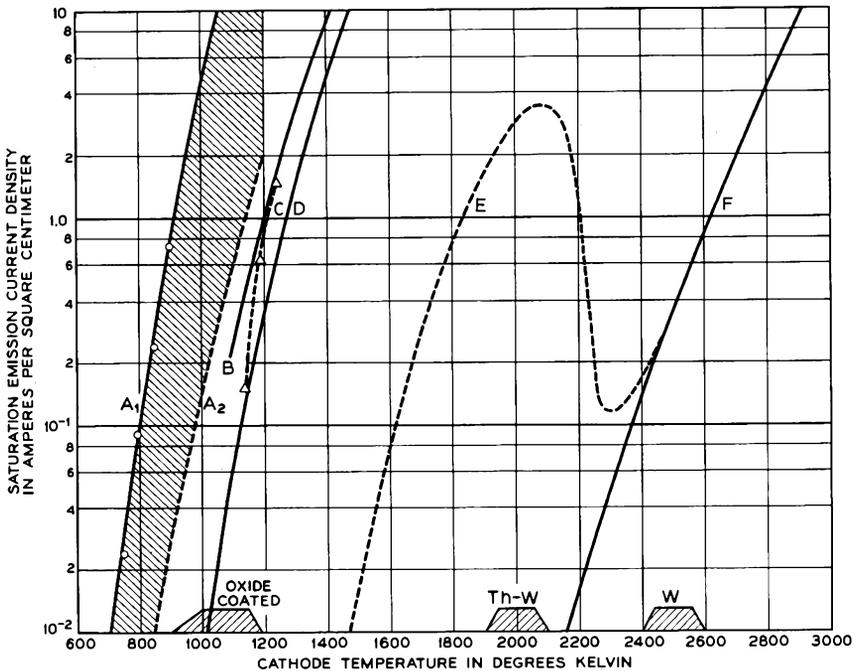
FIG. 2.3-1 Emission current density vs. cathode temperature for several types of thermionic emitters. The shaded blocks at the bottom of the figure show the normal operating range for three of the cathodes. (a) *The oxide-coated cathode.* Curve $A_1$ gives the saturation emission current density under pulsed conditions. (Circular points, Reference 2.11; solid curve, Reference 2c, Volume I.) Curve $A_2$ gives the dc saturation emission density. The position of this curve may vary substantially with environmental conditions. DC current densities much in excess of 0.5 amp/cm$^2$ lead to relatively short cathode life. (b) *The pressed nickel cathode.* Curve B shows the dc saturation emission current density obtained from a pressed nickel cathode (Reference 2.15). (c) *The impregnated nickel cathode.* Curve C shows the saturation emission current obtained from the impregnated nickel cathode. The measurements were taken with 40 microsec pulses and a repetition rate of 60 pulses/sec (Reference 2.12) (d) *Pressed and impregnated tungsten cathodes.* Curve D shows the saturation emission density obtained from pressed and impregnated tungsten cathodes based on $A' = 2.5$ amps/cm$^2$($^\circ$K)$^2$ and $\phi = 1.67$ electron volts. (These constants are given in Reference 2.13. However, it is the experience of the writers that under practical operating conditions somewhat higher cathode temperatures are needed to yield a given emission current density than those indicated by curve D). *The thoriated tungsten cathode.* Curve E shows the measured saturation emission current density of an uncarbonized thoriated tungsten filament (Reference 2.9). (f) *Tungsten filaments.* Curve F shows the saturation emission current density of a tungsten filament based on $A' = 70$ amps/cm$^2$($^\circ$K)$^2$ and $\phi = 4.5$ electron volts.

function, much effort has been devoted to obtaining cathode materials with low work functions. The oxide-coated cathode combines good emission properties and appreciable life with a particularly low work function, and consequently it is the most widely used type of cathode emitter.

Some discussion of the more frequently used cathode materials is given in Sections (a) to (d) below. Methods for heating the cathode are described in Section (e).

## (a)  Pure Tungsten

Filaments of pure tungsten[10] are operated at about 2500°K. At this temperature they yield a saturation emission density of 300 to 400 ma/cm² and radiate about 70 watts/cm². This high radiation per unit area means that a relatively large amount of power is required to heat the filament. If the ratio of emission current density to power radiated per unit area is taken as a measure of efficiency in obtaining emitted electrons for a given amount of heater power, tungsten cathodes are the least efficient of the commonly used cathode materials.

The vapor pressure of tungsten at 2500°K is $1.3 \times 10^{-8}$mm of Hg, which is sufficiently small that the filaments are able to survive several thousand hours of operation before failure occurs. Nearly all other common pure metals have vapor pressures much too high at temperatures at which appreciable thermionic emission takes place. (One exception is tantalum, which has found some application as a cathode material.) The melting point of tungsten is 3640°K.

Curve $F$ in Figure 2.3-1 shows a plot of emission current density vs. temperature for a tungsten emitting surface.

Tungsten filaments find their chief application in tubes that operate with anode voltages greater than about 20 thousand volts. Other cathode materials suffer severe damage from bombardment by positive ions if used in tubes that operate at very high voltages. The positive ions are formed by the emitted electrons striking molecules of residual gas in the tube and are accelerated toward the cathode by the same field that accelerates the electrons away from the cathode. If they strike an oxide-coated cathode with sufficient energy, they may chip away part of the emitting surface. However, pure tungsten filaments show less damage as a result of such bombardment. A number of x-ray tubes, high-voltage diode rectifier tubes, and some high-voltage transmitting tubes use tungsten filaments.

---

[10]The fabrication of tungsten filaments is described in Reference 2.7, Chapter 8.

*(b)   Thoriated Tungsten*

Thoriated tungsten filaments are made by adding 1 or 2 per cent of thorium oxide, $ThO_2$, to the tungsten before it is sintered and drawn into filaments. After drawing, the filament is heated in an atmosphere of hydrocarbon vapor, causing the surface of the wire to be converted to tungsten carbide, $W_2C$, to a depth of about one tenth of the wire radius, a process known as carbonization. The normal operating temperature of the filament is about 2000°K, and at this temperature the tungsten carbide slowly reduces the thorium oxide. Free atoms of thorium thus produced diffuse through the metal and eventually reach the surface, where some are adsorbed and others evaporate onto surrounding electrodes. Under normal operating conditions, there is probably somewhat less than a mono-layer of thorium atoms adsorbed on the surface of the filament. At the same rate that free thorium atoms diffuse to the surface, other thorium atoms that were adsorbed on the surface are lost as a result of evaporation, reaction with residual gases in the tube, and positive ion bombardment.

It is found that a partial layer of thorium atoms adsorbed on a tungsten carbide surface evaporates at a much slower rate at a given temperature than it would from solid thorium. As a result, the filament can operate at a much higher temperature than would be possible for solid thorium. How-ever, if several layers of thorium are adsorbed on the surface, the evapora-tion rate of the outer layers is much the same as from solid thorium, so that there is a tendency for additional layers to be lost, leaving only a single layer, or perhaps a little less than a layer.

The adsorbed atoms form a dipole layer at the surface with positive charge on the outside. This modifies the potential acting on a conduction electron at the surface of the filament causing the work function to be lower. The resulting work function is not that of thorium, but is one characteristic of thorium atoms adsorbed on a tungsten carbide surface. At 2000°K the work function is between 2.6 and 2.7 electron volts and $A'$ is about 4 amps/cm$^2$(°K)$^2$. The reason that $A'$ is lower than the values measured for clean metal surfaces is not well understood.

Early thoriated tungsten filaments were not carbonized; thermal re-duction of the thorium oxide was relied upon to release free thorium. Generally, the filaments were first "activated" by heating them well above the normal operating temperature for a few minutes to effect appreciable reduction of the thorium oxide. This was followed by operation for about a half hour at a somewhat lower temperature, still above the normal operat-ing temperature, to permit diffusion of the free thorium to the surface. However, the rate of thermal reduction of the thorium oxide at the normal operating temperature was insufficient to keep up with the loss of thorium from the surface, and the filaments had to be "reactivated" from time to

time. Carbonization both increases the rate of reduction of the thorium oxide and reduces the rate of evaporation of thorium atoms from the surface of the filament.[11]

As the filament temperature is increased above 2000°K, the rate of evaporation of thorium atoms from the surface increases faster than the rate of diffusion of thorium atoms from the interior of the filament, with the result that the fraction of the surface covered with thorium atoms decreases. This causes $\phi$ to increase toward the value for a clean tungsten-carbide surface, and one might expect that the value of $A'$ would likewise approach the corresponding value for a clean tungsten-carbide surface.

Curve $E$ in Figure 2.3-1 shows the variation of the saturation emission current density with temperature for an uncarbonized thoriated tungsten filament according to measurements by Langmuir[12]. Between 2100°K and 2300°K, the emission density actually falls because of the rapid increase of $\phi$ with temperature. Langmuir estimated that at the maximum of the curve, the surface was covered with about 0.8 of a layer of thorium atoms, whereas at the minimum to the right, he estimated there was only about 0.15 layer of thorium on the surface.

A plot similar to curve $E$ for a carbonized thoriated tungsten filament does not appear to be available at the time of writing. However, at 2000°K carbonized filaments give an emission current density of about 3 amps/cm², which is comparable to that indicated by curve $E$ for uncarbonized filaments. The radiation from a carbonized filament at 2000°K is about 28 watts/cm², so that the thoriated tungsten filament offers considerable advantage over pure tungsten filaments with respect to the ratio of emission current density to power radiated per unit area.

Thoriated tungsten filaments are used in a number of moderate-voltage transmitting tubes and a class of hot-cathode gas tubes, known as tungar rectifiers. Cathodes of thoriated tungsten are also used in high-power beam-type microwave tubes, where the high emission capabilities of the thoriated tungsten cathode are an important advantage. The cathodes in this case are heated by electron bombardment.

(c)  *Oxide-Coated Cathodes*

Although the physical processes involved in thermionic emission from an oxide-coated cathode are not well understood at present, the description of the physical processes and the interpretation of the tube processing given in the following paragraphs appear to be the prevailing thought in a somewhat simplified form.

---

[11]Reference 2.8.
[12]Reference 2.9.

*General.* The oxide-coated cathode is used in nearly all general-purpose tubes, many low-voltage transmitting tubes, and most hot-cathode gas tubes. It consists of a thin coating of a mixture of metallic oxides applied to a base of nickel or some nickel alloy.[13] The oxides most frequently used are those of the alkaline earths, barium, strontium, and calcium, (i.e., $BaO$, $SrO$, and $CaO$). Since these oxides are unstable in the presence of atmospheric moisture, the coating is applied to the base in the form of the corresponding carbonates ($BaCO_3$, etc.), and during the processing of the tube the cathode is heated to decompose the carbonates and release $CO_2$.

To prepare the carbonates for application to the cathode, they are first ground into a fine powder and mixed with an organic binder and a suitable solvent. Often nitrocellulose serves as the binder and amyl acetate as the solvent. The resulting mixture is then applied to the cathode, often by spraying, although filamentary cathodes are frequently dipped in the mixture or drawn through it. The coated cathode is then dried in room air.

While the tube is being pumped, the envelope and electrodes are heated, generally by applying rf induction to the metal parts. Sometimes an oven is also placed over the tube for a period before the rf induction is applied. This heating of the envelope and electrodes drives off appreciable amounts of gases that are adsorbed on the inside surfaces of the tube. It also causes the organic binder in the cathode coating to decompose into volatile gases, which are pumped away. When the envelope and electrodes are suitably "outgassed," the cathode temperature is raised to approximately $1000°K$ for about a minute. This causes the carbonates in the cathode coating to decompose into carbon dioxide and the metal oxides, a process known as "breakdown." The carbon dioxide is evolved as a gas and is pumped away.

A coating of the pure oxides is an insulator, and as such is capable of supporting very little sustained emission. To become suitable for thermionic emission, the coating must first be "activated." In this process, barium oxide is partially reduced, given rise to free barium atoms within the coating, which in turn aid in making the coating a semiconductor and increase its emission capabilities. Cathode activation is accomplished by heating the cathode to a temperature above the normal operating temperature. Sometimes current is drawn from the cathode while at this elevated temperature. (Often the normal operating temperature lies between $1000°K$ and $1150°K$.) The heating of the cathode above the normal operating temperature causes impurity atoms in the base nickel to diffuse through the nickel, with the result that some reach the interface between the coating and the nickel, where they reduce the oxides in the coating. It

---

[13]One exception is fluorescent lamps in which tungsten is used for the base.

is found that the impurity atoms principally react with the barium oxide, so that the reactions at the interface lead to the release of free barium atoms. The passage of current through the coating probably also causes some electrolytic dissociation of the coating.

Most cathodes are activated partly while the tube is being pumped and partly after the tube is removed from the pump. Usually the cathodes of general-purpose tubes are raised to a temperature of 1200° to 1450°K for about 30 seconds during pumping and then "aged" for a length of time after removing the tube from the pump. Often a series of aging steps is used in which both the cathode temperature and the anode current are varied from step to step. The aging time may range from 10 minutes to several days, depending largely on the quality of the tube being made. Tubes intended for applications needing high reliability are generally aged for a longer period of time at a lower cathode temperature.

The activated coating is white in appearance, its thickness is of the order of $0.5 \times 10^{-2}$ cm (0.002 inch), and it is highly porous, having a density of about one quarter that of the solid oxides. Its electrical properties are those of an $n$-type semiconductor.[14]

In some tubes a mixture of barium, strontium, and calcium oxides is used, the molecular proportions consisting of 10 to 14 per cent calcium oxide and about equal percentages of barium and strontium oxides. However, there has been a trend in the tube industry toward the use of "double-carbonate" coatings containing barium and strontium oxide in about equal molecular proportions. The "triple-carbonate" coating (barium, strontium, and calcium oxides) has been found to give faster activation and consequently is attractive from a manufacturing standpoint. However, the adherence of the coating to the base nickel is somewhat poorer, and failure due to peeling of the coating is more likely. In addition, some recent studies of the life capabilities of grid-controlled tubes made with double- and triple-carbonate coatings indicate that greater life can be achieved with the double-carbonate coating. The use of barium oxide alone has been found to give particularly unreproducible results,[15] whereas an active coating of strontium oxide has a higher work function than that of a mixture of barium and strontium oxides.

---

[14]At room temperature and higher, electrons are excited to the conduction band of the coating by donors which are distributed throughout the coating. At the cathode operating temperature, a few of the conduction electrons gain sufficient energy to overcome the work function of the coating and escape from the surface. The current of escaping electrons is the thermionic emission current. At present there is some difference of opinion as to what constitutes the principal donor in the coating; possibly it is the free barium atoms.

[15]It is thought that the barium oxide dissolves into the barium carbonate forming a solution which melts at about 1175°K. (Reference 2c, Vol. I, p. 62.)

At the cathode operating temperature, barium oxide has an appreciably higher vapor pressure than either strontium oxide or calcium oxide. Consequently, the barium oxide in the outermost part of the coating gradually becomes depleted, and evaporated barium oxide deposits on the surfaces surrounding the cathode throughout the operational life of the tube.

Equation (2.2-3), which gives the emission current density from a metal, does not apply to emission from a semiconductor and hence does not apply to emission from an oxide-coated cathode. However, it is found that the variation of emission density with temperature for an oxide-coated cathode is governed principally by the factor $\epsilon^{-|e|\phi/kT}$ as in the case of emission from metal surfaces. Experimental values of $\phi$ for commonly used coatings lie between 1.0 and 1.3 electron volts.

The cathodes of many general-purpose tubes are operated at temperatures in the range from 1000° to 1150°K. At 1050°K, the heat radiation from an oxide coating[16] is about 3 watts/cm², and mean cathode current densities of a few hundred milliamperes per square centimeter are found to be consistent with reasonably long life of the cathode, perhaps 10 to 30 thousand hours. However, if the cathode temperature is raised in order to increase the emission current density beyond 500 ma/cm², the life of present types of oxide-coated cathodes is found to decrease rapidly with increasing temperature and cathode current density. When particularly long life is desired from a tube, lower operating temperatures are used, and the emission density must be correspondingly less. The oxide-coated cathodes used in repeater tubes for recently developed underwater telephone cables operate at 940°K, and are expected to have an average operational life in excess of 40 years. The mean current density drawn from the cathodes in this case is only 10 ma/cm².

*The Base Nickel.* During operation of the cathode, free barium atoms in the coating diffuse through the coating, eventually reaching the outer surface, whereupon many evaporate onto surrounding electrodes and the walls of the tube. Since an excess of barium atoms in the coating is necessary for the coating to be an active emitter, impurity atoms in the base nickel must continually reduce the barium oxide and release free barium atoms. For this reason, much attention has been given to the impurity content of nickels used as the base material. A nickel too rich in reducing agents will cause short cathode life, whereas a very pure nickel will lead to insufficient emission. In addition, most reducing agents form compounds at the interface between the coating and the nickel, and these compounds often have the effect of placing an electrical impedance between the cathode lead and

---

[16]Black-body radiation at this temperature is nearly 7 watts/cm².

the emitting surface. In grid-controlled tubes, such an impedance is likely to cause adverse effects on the electrical performance of the tube.

The principal reducing agents found in cathode nickels are listed in Table 2.3-1. Those elements near the top of the table are found to be more active reducing agents than those near the bottom of the table[17]. Zirconium

TABLE 2.3-1. CONCENTRATION OF REDUCING AGENTS IN PER CENT
BY WEIGHT IN AN ACTIVE NICKEL AND A PASSIVE NICKEL

| Reducing Agent | Atomic Number | Type A (Active) | Type B (Passive) |
|---|---|---|---|
| C | 6 | 0.06 | < 0.01 |
| Mg | 12 | 0.04 | < 0.01 |
| Al | 13 | 0.006 | < 0.005 |
| Si | 14 | 0.03 | < 0.01 |
| Ti | 22 | 0.02 | < 0.005 |
| Zr | 40 | None | None |
| W | 74 | None | None |

and tungsten are found in cathode nickel only if they have been intentionally added to the nickel. A nickel relatively rich in reducing agents is said to be an "active" nickel and has the property that shorter times are required to activate and age the coating. The concentration of reducing agents in such a nickel might be as indicated for Type A in the table. A nickel such as Type A is frequently used in the manufacture of commercial grid-controlled tubes. However, because of the relatively rapid rate of reduction of the barium oxide and because the formation of interface compounds (see below) may be appreciable, the life of the coating may be less than might be obtained with smaller amounts of reducing agents.

A nickel that is nearly free of reducing agents, such as Type B in the table, would be considered to be a "passive" nickel. Such a nickel would be incapable of supporting substantial emission from an oxide cathode over an appreciable length of time because the rate of diffusion of impurity atoms through the nickel to the oxide coating would be too slow. In the manufacture of tubes requiring particularly long life there has been a trend in the industry toward the use of nickels that are more passive than Type A. Recently, studies have been carried out on the performance of cathodes in which the base consists of a passive nickel to which one or two reducing agents are added in controlled amounts. One "single-additive" nickel which appears to perform satisfactorily contains about 0.1 per cent zirconium. The zirconium both increases the mechanical strength of the nickel and

[17]Part of this undoubtedly results from the fact that the elements near the top of the table are lighter and hence diffuse faster.

acts as the reducing agent to provide long-term activity of the cathode. A "double-additive" nickel which has also given good results contains 0.01 to 0.03 per cent magnesium and 2 to 4 per cent tungsten. The magnesium makes possible the initial activity of the cathode, but it is soon lost from the nickel by reaction with the coating and evaporation from the surface. The tungsten makes possible the long-term activity and considerably increases the mechanical strength of the nickel.

It is believed that the nickel itself does not reduce the oxides in the coating. Other impurities which are sometimes present in the nickel and which probably have little effect upon the coating activity are[18] cobalt, iron, copper, and molybdenum. One nickel-cobalt-iron alloy which is sometimes used in oxide-coated filamentary cathodes contains about 19 per cent cobalt and 2 per cent iron. The addition of the cobalt and iron gives greater mechanical strength to the filament and increases its resistivity over that which can be obtained with pure nickel.

*Interface.* Next let us consider the events that take place at the interface between the base metal and the coating. When carbon atoms reach the interface, they react with the oxide to form CO and free barium. The CO is evolved as a gas, some of which contributes to the residual pressure in the tube, some is absorbed by the getter material, and some becomes adsorbed on the inside surfaces of the tube. Other reducing agents in the base form solid compounds upon reacting with the barium oxide. Various workers have reported finding one or more of $MgO$, $BaAl_2O_4$, $Ba_2SiO_4$, $BaTiO_4$, and $Ba_3WO_6$ at the interface. Often the layer of compounds is referred to as "interface."

The presence of such a layer between the coating and the base has the effect of placing an electrical impedance between the cathode lead and the emitting surface. An approximate equivalent network[19] for the impedance consists of a resistance shunted by a capacitance as shown in Figure 2.3-2. Values of the resistance often range from a few tenths of an ohm to a few tens of



FIG. 2.3-2 Approximate equivalent network for the interface impedance. A more accurate network would include an additional resistance and capacitance in series which would shunt the network shown in the figure.

---

[18]Manganese and sulfur are known to have adverse effects on cathode activity.
[19]A more accurate network would include an additional resistance and capacitance in series which would shunt the network shown in Figure 2.3-2.

ohms for one square centimeter of emitting surface.[20] The capacitance shunting the layer gives the parallel $RC$ combination a time constant of about a microsecond. In the case of base nickels containing greater than 0.05 per cent silicon, by far the greater part of the interface resistance results from $Ba_2SiO_4$, whereas base nickels having less than 0.01 per cent silicon lead to particularly small interface resistances. Often the interface resistance increases when the cathode is held at operating temperature for extended periods with no current being drawn. In cases where the silicon content of the base nickel is high, interface resistances as high as 1000 ohms and more have been observed after such operation.

The coating also exhibits a resistance, often of the order of a few ohms across the thickness of a square centimeter of coating. This resistance and the interface resistance add in series and are effectively inserted between the cathode lead and the cathode emitting surface. Such a resistance $R$ in series with the cathode lead of a grid-controlled tube reduces the low-frequency transconductance of the tube. In a triode, the transconductance is reduced by the factor $1/(1 + g_m R)$, where $g_m$ is the transconductance in the absence of the resistance $R$. For $R = 100$ ohms and $g_m = 10,000$ micromhos, the transconductance is reduced by 50 per cent. At higher frequencies, of the order of a megacycle or more, the interface resistance is bypassed by the capacity shunting it, with the result that the transconductance approaches the value it would have in the absence of interface resistance.

During the life of a tube, the interface resistance increases because of the formation of additional quantities of interface compounds. This causes a further reduction in the transconductance, partly because of the factor $1/(1 + g_m R)$, and partly because an increase in $R$ causes the cathode current to decrease, and this in turn reduces the transconductance. Often the two effects contribute comparable amounts to the decrease in transconductance, and together will account for the failure of a tube.[21]

Since much smaller interface resistances are obtained with base nickels having low silicon content, there has been increasing use of such nickels in grid-controlled tubes and other tubes where minimum interface resistance is desirable.

*Pulsed and DC Emission Current Densities.* It is found that the saturation current density drawn from an oxide-coated cathode under pulsed conditions with pulse lengths of the order of a few microseconds and a low-duty cycle is often of the order of 10 times that which can be drawn under dc conditions. However, as the pulse length is increased to a few milliseconds,

---

[20]Sometimes the coating tends to blister or peel, and this adds to the apparent interface resistance.

[21]Frequently end of life of a grid-controlled tube is assumed to take place when the transconductance falls below about 65 per cent of its initial value.

or a second, the amplitude of the current pulse decays toward the end of the pulse. Figure 2.3-3 shows the decay of current density drawn from a particular diode when a step-function voltage is applied to the anode. Two
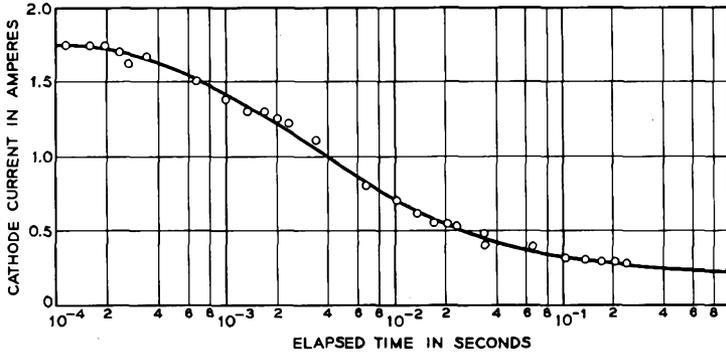


FIG. 2.3-3 The decay of the saturation emission current density drawn from a particular diode when a step-function voltage is applied to the anode. The abscissa indicates time elapsed after application of voltage to the anode. (From L. S. Nergaard, *RCA Rev.* **13**, 464, December 1952)

effects are thought to be principally responsible for the decay of the pulsed emission current when the pulse length is increased to times of the order of a millisecond or a second:

1. Gas released from the anode and other electrodes struck by the electrons tends to destroy cathode activity by oxidizing the impurity centers in the coating. The gas is released both as a direct result of electron bombardment and as a result of heating caused by the kinetic energy of the incident electrons. It is known that $O_2$, $CO_2$, $SO_2$, $H_2O$, and $CO$ are all effective in destroying the impurity centers.

2. Under the influence of the potential gradient established in the coating when current is conducted through the coating, the impurity centers tend to migrate toward the nickel base leaving a layer of oxide near the surface that is partially depleted of impurity centers.[22]

In Figure 2.3-1, curve $A_1$ shows the saturation emission current density that can be drawn from an oxide-coated cathode in a very clean environment under pulsed conditions. Curve $A_2$ shows the dc saturation emission current that might be obtained under normal operating conditions. The position of curve $A_2$ depends much on the environment in which the cathode is operated.

---

[22]Reference 2.10.

*Causes of Failure.* Finally, let us list the principal causes of failure of oxide-coated cathodes. These are:

1. Formation of excessive interface.
2. Peeling or blistering of the coating.
3. Destruction of the coating by ion bombardment.
4. Evaporation of the coating.
5. "Poisoning" of the cathode by residual gases or foreign matter within the tube.
6. Depletion of the activating agents in the base nickel.

Of course, in filamentary cathodes mechanical failure of the filament itself is still another cause of failure.

*(d)   The L-Cathode; Pressed and Impregnated Cathodes of Tungsten and Nickel*
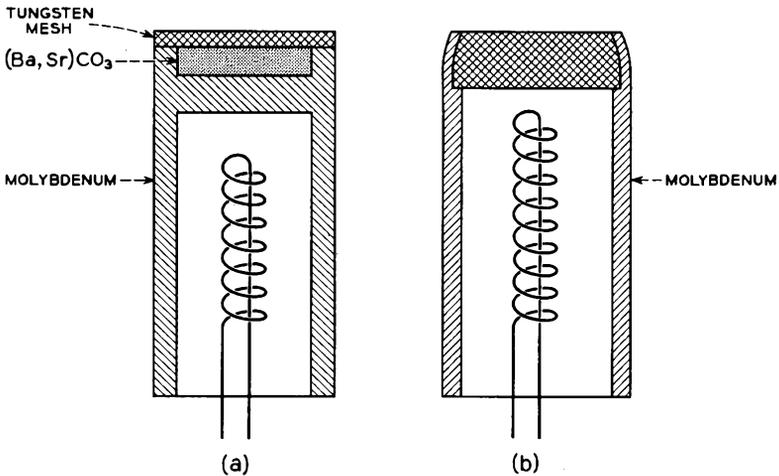


FIG. 2.3-4 The construction of the Philips cathodes: (a) the L-Cathode; (b) the pressed or impregnated cathode. The heaters are shown schematically. (From A. Venema *et al.*, *Philips Tech. Rev.* **19**, 177, 1957)

Several forms of cathode in which emission takes place from the surface of a porous tungsten body, which is covered with adsorbed barium and oxygen atoms, have been developed by the Philips Laboratories.[23] One of these,

[23]Philips Laboratories, Eindhoven, The Netherlands, and Irvington-on-Hudson, N.Y., U.S.A. The fabrication of the Philips cathodes is described in Reference 2.13.

known as the L-Cathode, is illustrated in Figure 2.3-4(a). Here the emitting surface consists of a porous tungsten body of approximately 27 per cent porosity which is welded to a nonporous molybdenum support. A chamber between the tungsten and molybdenum initially contains (Ba, Sr)CO₃ in solid solution. However, breakdown of the cathode causes the carbonates to be reduced to a solid solution of (Ba, Sr)O. During operation of the cathode, evaporated BaO diffuses and migrates through the tungsten pores, and some of it is reduced by the tungsten. As a result, the outer surface of the tungsten body becomes covered with adsorbed barium and oxygen atoms, perhaps nearly a monolayer of oxygen covered with a monolayer of barium.[24]  The emission constants for such a surface as measured by Philips Laboratories are approximately $\phi = 1.67$ electron volts and $A' = 2.5$ amps/cm²(°K)². Curve $D$ in Figure 2.3-1 shows a plot of emission current density vs cathode temperature based on these constants.

In two other cathodes developed by the Philips Laboratories, the oxides are contained within the pores of the metal body rather than in a reservoir beneath it. In one of these, known as the pressed cathode, a powdered mixture of CaCO₃, BaCO₃, and Al₂O₃ is mixed with a powdered tungsten-molybdenum alloy containing 75 per cent molybdenum, and the resulting mixture is pressed together in a die. The compact is then removed from the die and sintered at 1850°C. The resulting metal body has a porosity of about 40 per cent. It is mounted in a molybdenum support as shown in Figure 2.3-4(b). Emission properties of the cathode are similar to the L-Cathode, but the rate of evaporation of BaO from the pressed cathode is somewhat higher. The use of Al₂O₃ serves to increase the stability of the cathode in the presence of atmospheric moisture.

In still another Philips cathode, known as the impregnated cathode, a porous tungsten body of about 17 per cent porosity is first machined into the desired geometry. Then a powdered mixture of CaCO₃, BaCO₃, and Al₂O₃ is brought in contact with the tungsten body and heated until the powder melts and is drawn into the pores by capillary action. The resulting cathode shows a lower rate of evaporation of BaO than the pressed cathode and has generally found favor over both the pressed cathode and the L-Cathode. (The L-Cathode requires a relatively long activation cycle, and difficulties are experienced in welding the porous tungsten body to the molybdenum holder so that there are no cracks through which the BaO can evaporate.) The emission properties of all three Philips cathodes are about the same.

---

[24]Reference 2.14.

Several advantages of the Philips cathodes are:

1. High emission capability consistent with reasonably long life. Emission of a few amperes per square centimeter can be obtained with a cathode life of the order of several thousand hours when the cathode is operated in a relatively clean environment.

2. The emitting surface is physically strong and can be shaped very accurately.

3. The surface is capable of withstanding strong ionic bombardment.

4. There is negligible potential drop across the cathode.

The cathodes have the disadvantage of requiring a higher operating temperature than either the oxide-coated cathode or the pressed or the impregnated nickel cathodes discussed below.

Philips cathodes have found application in a number of beam-type tubes, including cathode-ray tubes and certain microwave tubes. They have also been used in some magnetron applications.

Pressed and impregnated cathodes can also be made with the porous metal body made of nickel. Such cathodes combine most of the advantages of the Philips cathodes with a somewhat lower operating temperature for a given current density. The pressed cathode is made by pressing together a mixture of powdered $(Ba, Sr)CO_3$ and powdered nickel in a die and sintering the resulting compact at $1000°C$. The nickel body then has a porosity of about 50 per cent. Curve $B$ in Figure 2.3-1 shows measurements[25] of the dc saturation emission current density vs. cathode temperature for the pressed nickel cathode. Pressed nickel cathodes are used in a number of beam-type microwave tubes.

Most magnetron oscillators use either pressed or impregnated nickel cathodes. The magnetron cathode is cylindrical in shape with the emitting surface on the outer side of the cylinder. During operation of the tube, many of the emitted electrons are returned to the cathode with appreciable velocity after traveling some distance through the tube. This "back bombardment" tends to destroy an oxide-coated cathode, but pressed or impregnated cathodes show little effect from the returning electrons. Generally, a molybdenum sleeve provides a base for the cathode, molybdenum having greater strength at the cathode operating temperature than nickel. However, because the coefficient of expansion of the outer nickel body is somewhat different from that of molybdenum, one or more intervening layers of porous nickel or nickel-molybdenum alloy are used to help absorb the difference in expansion. The structure of a typical impregnated

[25]Reference 2.15.

cathode is illustrated in Figure 2.3-5. Next to the molybdenum, there is a
layer of nickel-molybdenum alloy followed by a coarse porous nickel layer
onto which is sintered the finer outer layer. The porosity of the outer layer
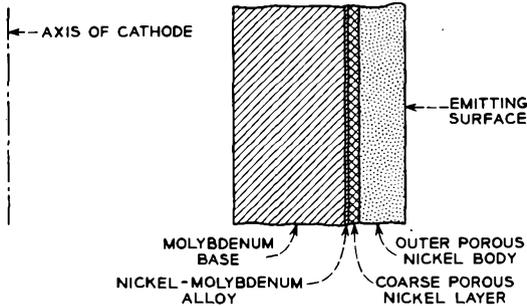


FIG. 2.3-5 A cross-sectional view of a portion of an impregnated magnetron cathode

is about 50 per cent. The outer layer is generally impregnated by bringing
it in contact with a colloidal suspension of $(Ba, Sr)CO_3$. This is drawn into
the pores by capillary action. Curve $C$ in Figure 2.3-1 shows measurements
of the saturation emission current density from the impregnated nickel
cathode as a function of cathode temperature.

### (e)   Heating the Cathode

Most oxide-coated cathodes are indirectly heated. Often the cathode
emitting surface in grid-controlled tubes has the shape of a circular or
elliptical cylinder, and the heater consists of a coiled or folded tungsten wire
which is inserted inside the cathode. To prevent the heater from making
electrical contact with the cathode or short-circuiting to itself, a coating of
aluminum oxide is applied to the heater after forming the wire into its final
shape. The coated heater is then fired in an oven at about $2000°K$ to sinter
the aluminum oxide. The resulting coating is hard and insulating and can
withstand a moderate amount of abrasion during assembly of the tube.
The normal operating temperature of the heater is about $400°C$ above that
of the cathode.

Filamentary cathodes have the advantage over indirectly heated cath-
odes of being quicker to reach the operating temperature and of requiring
somewhat less power. However, they have a number of important limita-
tions which preclude their use in many vacuum-tube applications. Principal
among these are:

1. The cathode emitting surface has the shape of a long slender wire.
2. The voltage drop across the filament is often comparable to the inter-
electrode voltages.

3. Mechanical motion of the filament as it warms up generally precludes the use of filamentary cathodes in tubes where close electrode spacings are needed.

4. The filament is mechanically weak at the operating temperature and hence must be supported at one or more points over its length in order to prevent it from contacting the other electrodes. If the supports are conductors, they must be insulated from the other electrodes.

5. Direct current must be used to heat the filaments of high-gain amplifier tubes, since an ac filament current would introduce hum in the output of the tube.

Oxide-coated filamentary cathodes have found application in a number of diode rectifiers, where the voltage drop across the filament and variations in the electrode spacings do not have an important effect on the tube performance. They are also used in portable radio applications, where low power consumption is desirable, and in proximity fuses and other military applications where extremely fast warmup times are needed.

Cathodes that operate at temperatures above about 1400° or 1500°K are generally directly heated, either as filaments or by electron bombardment. Indirect heating would require sufficiently high heater temperatures that the aluminum-oxide coating on the heater would melt or evaporate.

A number of high-power klystron amplifiers which require an ampere or more of beam current use thoriated tungsten cathodes to take advantage of the high emission capabilities of thoriated tungsten. The cathode emitting surface is in the form of a concave disc and is heated by electron bombardment from the reverse side. A bombarding current of several hundred milliamperes is obtained from an auxiliary oxide-coated cathode which is held a kilovolt or more negative with respect to the thoriated-tungsten cathode. The bombarding electrons generate sufficient heat in the thoriated tungsten cathode to raise it to an operating temperature of about 2000°K.

## 2.4 Thermionic Emission Energies

The equations given in Section 2.2 for the emission current density can be used to predict the distribution of emission energies in the direction normal to the emitting surface. If the work function $\phi$ of a metal is increased by $W_n$ electron volts, it follows from Equation (2.2-1) that the emission current density at temperature $T$ is changed by the factor $\epsilon^{-W_n/W_T}$. Consequently, this fraction of the emitted electrons is able to overcome a work function $W_n$ electron volts greater than that actually present. It follows that the electrons comprising this fraction have more than $W_n$ electron volts of kinetic energy associated with their motion normal to the cathode at the time of

emission. Similarly, the fraction of electrons that leave the cathode with greater than $W_n + dW_n$ electron volts of kinetic energy in the direction normal to the cathode is $\epsilon^{-(W_n+dW_n)/W_T}$, and the fraction of electrons that leave the cathode with normally directed energy in the range $W_n$ to $W_n + dW_n$ electron volts is

$$dP(W_n) = \epsilon^{-W_n/W_T} - \epsilon^{-(W_n+dW_n)/W_T} = \epsilon^{-W_n/W_T}[1 - \epsilon^{-dW_n/W_T}] \qquad (2.4\text{-}1)$$

If $dW_n$ is small compared with $W_T$, Equation (2.4-1) can be written as

$$dP(W_n) = \epsilon^{-W_n/W_T}\frac{dW_n}{W_T} \qquad (2.4\text{-}2)$$

This expression gives the probability that the part of the emission energy associated with motion normal to the cathode lies in the range $W_n$ to $W_n + dW_n$. It can be used as a weighting factor to calculate the average emission energy in the direction normal to the cathode. Thus,

$$\text{average normally directed energy} = \int_0^\infty W_n dP(W_n) = W_T \text{ electron volts} \qquad (2.4\text{-}3)$$

If $u_n$ is the electron velocity corresponding to a kinetic energy of $W_n$ electron volts, then $W_n = mu_n^2/(2\,|\,e\,|)$, where $|\,e\,|$ is a positive dimensionless constant numerically equal to the charge on the electron. Substituting for $W_n$ in Equation (2.4-2), we find that the probability that the emission velocity of an electron has a component normal to the cathode surface in the range $u_n$ to $u_n + du_n$ is given by

$$dP(u_n) = \frac{mu_n}{kT}\epsilon^{-mu_n^2/2kT}du_n \qquad (2.4\text{-}4)$$

It can further be shown[26] that the emitted electrons have an average kinetic energy associated with their motion parallel to the cathode surface of $W_T$ electron volts, and the probability that the part of the emission energy associated with motion parallel to the cathode surface lies in the range $W_t$ to $W_t + dW_t$ electron volts is

$$dP(W_t) = \frac{1}{W_T}\epsilon^{-W_t/W_T}dW_t \qquad (2.4\text{-}5)$$

where the subscript $t$ refers to motion in the "transverse" direction. If $u_t$ is the electron velocity corresponding to an energy of $W_t$ electron volts, the probability that the emission velocity has a component parallel to the cathode surface in the range $u_t$ to $u_t + du_t$ is

$$dP(u_t) = \frac{mu_t}{kT}\epsilon^{-mu_t^2/2kT}du_t \qquad (2.4\text{-}6)$$

---

[26]Reference 2.4, p. 141.

It will be helpful to combine Equations (2.4-4) and (2.4-6) to determine the distribution of total emission velocities and the angular distribution of the emission velocities. Multiplying the left-hand sides and right-hand sides of the two equations, we find that the probability that the emission velocity simultaneously has a component normal to the cathode surface in the range $u_n$ to $u_n + du_n$ and a component parallel to the cathode surface in the range $u_t$ to $u_t + du_t$ is given by

$$dP(u_n,u_t) = dP(u_n)dP(u_t) = \left(\frac{m}{kT}\right)^2 u_n u_t \epsilon^{-mu^2/2kT} du_n du_t \qquad (2.4-7)$$

where $u^2 = u_n{}^2 + u_t{}^2$. If the emission velocity $u$ makes an angle $\theta$ with the normal to the surface, then $u_n = u \cos\theta$, and $u_t = u \sin\theta$. (See Figure 2.4-1.)
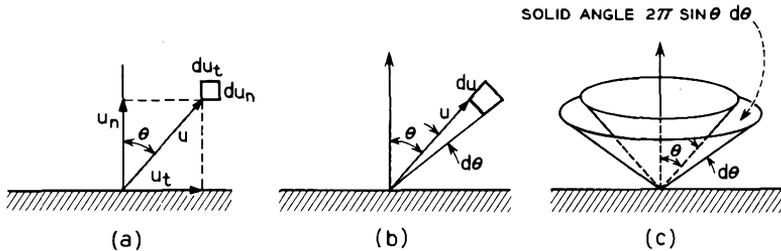


FIG. 2.4-1 The relationship between $u$, $u_n$, $u_t$, and $\theta$.

The probability that the emission velocity lies in the range $u$ to $u + du$ and makes an angle with the normal in the range $\theta$ to $\theta + d\theta$ can therefore be expressed as

$$dP(u,\theta) = \sin\theta \cos\theta \left(\frac{m}{kT}\right)^2 u^3 \epsilon^{-mu^2/2kT} du d\theta \qquad (2.4-8)$$

Integrating this expression with respect to $\theta$ from 0 to $\pi/2$, we find that the probability that the emission velocity lies in the range $u$ to $u + du$ is given by

$$dP(u) = \frac{mu^2}{2kT}\epsilon^{-mu^2/2kT} d\frac{mu^2}{2kT} \qquad (2.4-9)$$

Substituting $W = mu^2/2 \, |\, e \, |$ in this, we obtain the probability that the total emission energy lies in the range $W$ to $W + dW$, or

$$dP(W) = \frac{W}{W_T}\epsilon^{-W/W_T} d\frac{W}{W_T} \qquad (2.4-10)$$

Figure 2.4-2 shows a plot of this probability function vs. $W$ for a cathode temperature of 1000°K. The average total emission energy is $2W_T$.

From Equation (2.4-8) we can also obtain the angular distribution of the emission velocities. By integrating the equation with respect to $u$ from 0 to
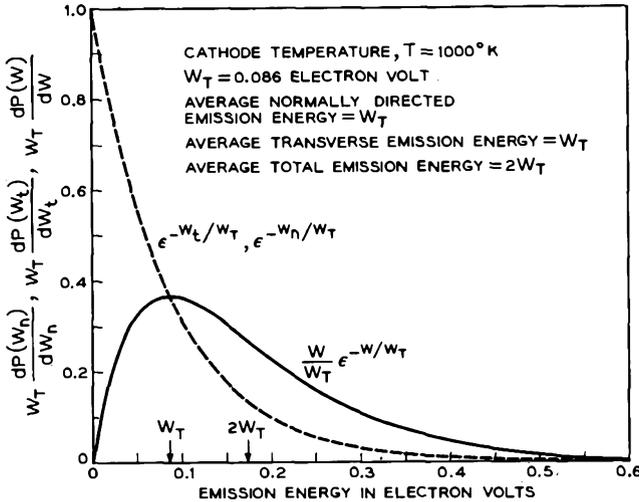


FIG. 2.4-2 The probability functions $W_T dP(W_n)/dW_n$, $W_T dP(W_t)/dW_t$, and $W_T dP(W)/dW$ are plotted vs. $W_n$, $W_t$, and $W$, respectively.

$\infty$, the probability that the direction of the emission velocity lies in the range of angle $\theta$ to $\theta + d\theta$ with respect to the normal is found to be

$$dP(\theta) = 2 \sin\theta \cos\theta \, d\theta \qquad (2.4\text{-}11)$$

Since the range of angle $\theta$ to $\theta + d\theta$ with respect to the normal defines a solid angle $2\pi \sin\theta \, d\theta$, which is subtended at the surface of the emitter, the current density emitted per unit solid angle at an angle $\theta$ with respect to the normal is

$$J_o \frac{dP(\theta)}{2\pi \sin\theta \, d\theta} = J_o \frac{\cos\theta}{\pi} \qquad (2.4\text{-}12)$$

where $J_o$ is the emission current density.

In light optics a source of brightness $B$ is said to emit according to Lambert's Law if the radiation per unit area per unit solid angle in the direction $\theta$ with respect to the normal is $B \cos\theta$. The total radiation from unit area of the surface is then

$$\int_0^{\pi/2} B \cos\theta \, 2\pi \sin\theta \, d\theta = \pi B \qquad (2.4\text{-}13)$$

A thermionic cathode is therefore said to emit according to Lambert's Law, and the quantity $J_o/\pi$ is analogous to the "brightness" of the emitter, where $J_o$ is the emission current density.

Appendix IV gives a summary of the important relations presented in this section. We shall make use of these relations in Section 4.4, where the effects of the electron emission velocities in electron guns are discussed.

It is interesting to note that the total kinetic energy of an emitted electron plus the energy required to overcome the work function of the metal must come from heat energy supplied by the cathode heater. Thus, in addition to making up for heat that is radiated and conducted away from the cathode, the heater must supply an amount of power given by $I_o V_{\phi+2W_T}$ to the emitted electrons, where $I_o$ is the net current drawn from the cathode,
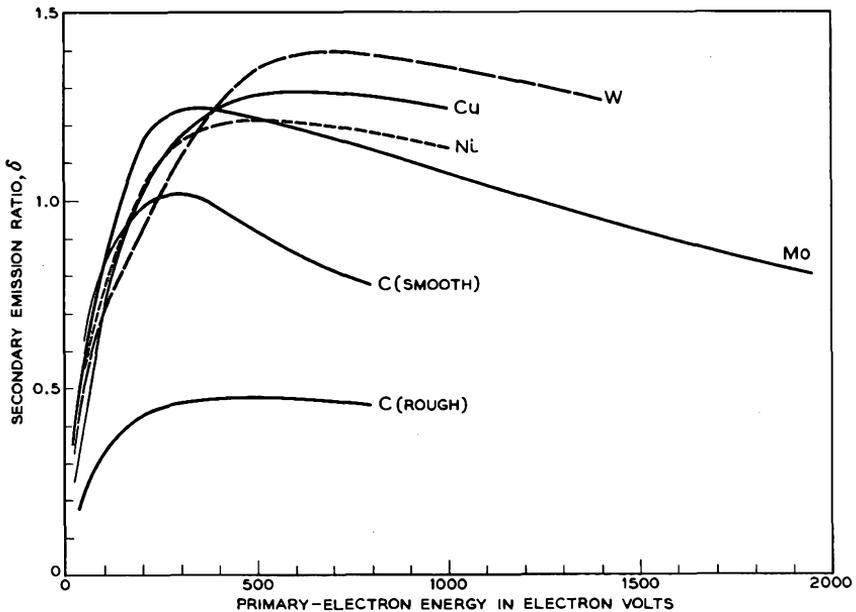


Fig. 2.5-1 Secondary emission ratio vs. primary-electron energy for clean surfaces of several metals. (Cu — Reference 2.16; C, Ni — Reference 2.17; Mo, W — Reference 2.18.)

and $V_{\phi+2W_T}$ is a voltage numerically equal to $\phi + 2W_T$ electron volts. If several amperes are drawn from the cathode, this power amounts to several watts.

## 2.5   Secondary Emission

All solid surfaces, both conducting and insulating, are capable of second-ary electron emission. The secondary-electron yield for a given bombarding energy is found to be directly proportional to the current of primary elec-trons incident upon the surface, and for metals the yield is nearly independ-ent of the temperature of the emitting surface.

Two important characteristics of the secondary emission from a surface are the total yield of secondary electrons as a function of the incident elec-tron energy, and the distribution of secondary-electron energies for a given primary-electron energy. Figure 2.5-1 illustrates the results of measure-ments of the first of these quantities for electrons incident upon clean surfaces of several metals. Plotted in the figure is the ratio of secondary electrons to primary electrons (often designated as $\delta$) as a function of the primary-electron energy. The same general shape of curve is found to apply to a large number of pure metals. In all cases, the curve rises to a maximum at a particular primary-electron energy and then falls off slowly with in-creasing primary energy.
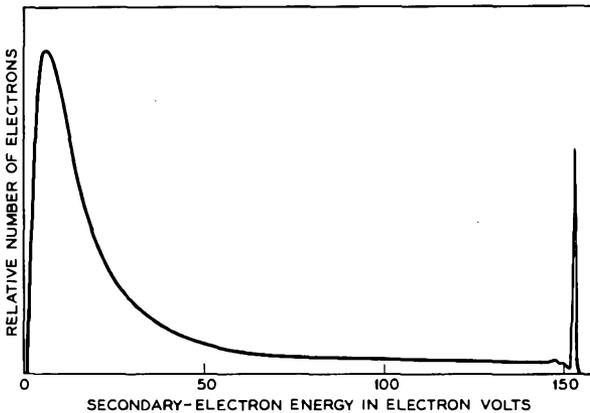


Fig. 2.5-2  Distribution of energies of secondary electrons from a gold target when 155-volt primary electrons are incident upon the surface. (From E. Rudberg, *Phys. Rev.* **50**, 138, 1936)

Figure 2.5-2 illustrates the distribution of energies of secondary electrons from a gold target when 155-volt primary electrons are incident upon the surface. The majority of emitted electrons have energies less than 30 elec-tron volts. However, a few have energies ranging all the way up to that of the incident electrons. It is probable that the true secondary electrons are emitted with kinetic energies less than 50 electron volts, and that most of

the electrons emitted with energies between 50 electron volts and the primary energy are actually inelastically reflected primary electrons. The spike at the right of Figure 2.5-2 results from elastically reflected primaries. For the majority of metals, the most probable emission energy of the true secondary electrons lies between 1.3 and 6 electron volts.

Next let us say a few words about the interactions that occur within a clean metal target when primary electrons with energies of a few hundred electron volts are incident upon the surface. Generally, the primary electrons travel through many atomic layers of solid before their kinetic energy is reduced to that of the conduction electrons in the metal. It is believed that the primary electrons lose their kinetic energy through excitation of both individual electrons and through electron plasma excitations.[27] With each excitation the primary electron abruptly loses an amount of energy equal to that imparted to the excited electron or electron plasma.

The excitation energy of an electron plasma is a discrete quantized value which is characteristic of the particular metal target. In most cases this energy lies between 5 and 30 electron volts. After a very short time interval the electron plasma excitation energy is in turn imparted to one or more electrons in the form of kinetic energy.

It is the electrons that are excited nearest the surface of the metal that have the greatest chance of escaping from the surface and being observed as secondary electrons. Measurements of the depths from which secondary electrons are emitted from a platinum target[28] indicate that, for 500-volt primary electrons, some of the emitted electrons receive their excitation energy as far as 15 atomic layers beneath the surface, and for 1000-volt primary electrons, some of the emitted electrons receive their excitation energy as far as 30 to 35 atomic layers beneath the surface.[29]

Of course, many of the excited electrons never reach the surface, and others lose sufficient energy before reaching the surface that they are unable to overcome the work function and escape. To evaluate the effect of change in work function on the secondary emission yield, McKay[30] evaporated somewhat less than a monolayer of sodium onto a tungsten surface, thereby reducing the work function of the surface to about half that of clean tungsten. He found that the secondary-electron yield at the primary energy giving maximum δ increased by about 60 per cent. Since the amount of secondary emission from the sodium was probably very small, he assumed

---

[27]Reference 2.20.

[28]Reference 2.21.

[29]Elastic reflection of primaries probably takes place from the first one or two atomic layers. However, some inelastic reflection undoubtedly takes place from appreciably greater depths.

[30]Reference 2.22.

that this increase resulted almost entirely from the change in work function of the surface.

The interactions between the primary electrons and the electrons of the solid, while best described by means of quantum mechanics, effectively result from interactions between the coulomb fields of the electrons. The excited electron receives an impulse that can be expressed as an integral of force $\times$ time. If the primary electron travels faster, the time of interaction is shorter, so that the impulse is smaller, and the probability of excitation is correspondingly less. This picture can be used to explain the shape of the curve giving secondary-electron yield vs. primary-electron energy. At low primary energies, the yield of secondary electrons increases with increasing primary energy, because the primary electrons expend more energy in slowing down and hence cause a greater number of excitations. At appreciably higher primary energies, the primary electrons are less effective in causing excitations in atomic layers close to the surface (where the escape probability is greatest), since they travel faster, and the time of interaction is shorter. Consequently, the secondary-electron yield at higher primary energies falls off with increasing energy.

Table 2.5-1 lists values of $\delta_{max}$, the maximum secondary-electron yield, and the corresponding primary-electron energy for a number of metals.[31]

<div align="center">TABLE 2.5-1*</div>

| Metal | $\delta_{max}$ | Corresponding Primary Electron Energy (Electron Volts) |
|-------|----------------|--------------------------------------------------------|
| Al | 1.0 | 300 |
| Au | 1.46 | 750 |
| C | 1.0 | 300 |
| Cu | 1.3 | 600 |
| Mo | 1.25 | 375 |
| Ni | 1.3 | 550 |
| Pt | 1.8 | 800 |
| Ti | 0.9 | 280 |
| W | 1.4 | 700 |
| Zr | 1.1 | 350 |

*From H. Bruining, *Physics and Applications of Secondary Electron Emission*, Pergamon Press, London, 1954.

The values given in the table apply only to clean surfaces of the metals. However, since the electrode surfaces in an electron tube are generally contaminated with adsorbed gases, oxides, and material evaporated from the

[31]Reference 2f, p. 39.

cathode, the secondary-electron yield from the electrodes is likely to differ substantially from that for a clean metal surface. An appreciable amount of oxide on a metal surface often increases the secondary-electron yield by a factor of 2 or more.
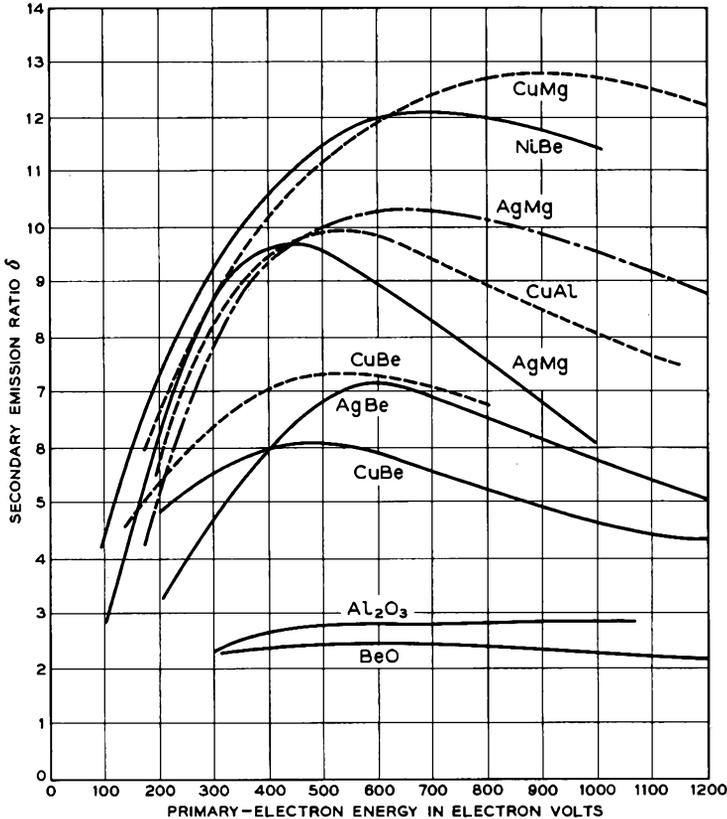


FIG. 2.5-3 Secondary emission yield vs. primary-electron energy for several compound surfaces. (After R. Kollath, *Handbook of Physics*, Vol. 21, p. 232, Springer Verlag, Berlin, 1956)

Certain compound surfaces have been found to give particularly high secondary-electron yields. Values of $\delta_{max}$ ranging as high as 10 or 12 have been reported in some cases. Several surfaces which have found application in electron tubes as good secondary-electron emitters are: a film of $Cs_3Sb$ deposited on a metal electrode, magnesium oxide on the surface of a silver-

magnesium alloy, beryllium oxide on the surface of a copper-beryllium alloy
or a nickel-beryllium alloy, and cesium oxide partially reduced on a base of
silver. Table 2.5-2 lists values[32] of $\delta_{max}$ and the corresponding primary-

<div align="center">TABLE 2.5-2</div>

| Metal | $\delta_{max}$ | Corresponding Primary Electron Energy (Electron Volts) |
|---|---|---|
| Cs₃Sb | 8.0 | 500 |
| AgMg | 9.8 | 500 |
| CuBe | 3.5–5.5 | 500–700 |
| NiBe | 12.3 | 700 |
| Ag-CsO₂-Cs | 5.8–9.5 | 500–1000 |

electron energy for these surfaces. Figure 2.5-3 shows a plot of the second-
ary-electron yield for several compound surfaces as a function of the pri-
mary-electron energy. The values of $\delta_{max}$ obtained for these surfaces depend
markedly on the manner of preparation of the surface, wide variations being
possible.

If a surface is very rough, the escape probability of the emitted electrons
may be substantially reduced, since electrons emitted from the bottom of a
hole or valley may strike other projecting parts of the surface and be re-
captured. In cases where it is desirable to reduce the secondary emission
from an electrode, the electrode is often coated with fine carbon granules.
Bruining[33] found that optimum reduction in the secondary-electron yield
occurs when the carbon granules are about 30 angstroms in diameter and
form a fine labyrinth. A plot of Bruining's measured secondary-electron
yield for carbonized nickel is shown in Figure 2.5-1 (curve "C(rough)").
The carbon can be deposited on an electrode either by spraying the elec-
trode with a suspension of lamp black or by passing it through a flame that
is generating carbon.

Curves giving the secondary-electron yield vs. primary-electron energy
for an insulator are similar in shape to those given in Figure 2.5-1, but the
values of $\delta_{max}$ are often appreciably greater. Figure 2.5-4 shows a plot of
the secondary-electron yield from mica. In experiments to measure the
secondary emission from an insulator, care must be taken to avoid charging
the surface of the insulator, or the primary-electron energy will be in-
determinate. Several experimenters have devised pulsed techniques which
overcome this difficulty.

---

[32]Courtesy Allen B. Dumont Laboratories, Clifton, N.J.
[33]Reference 2.17.
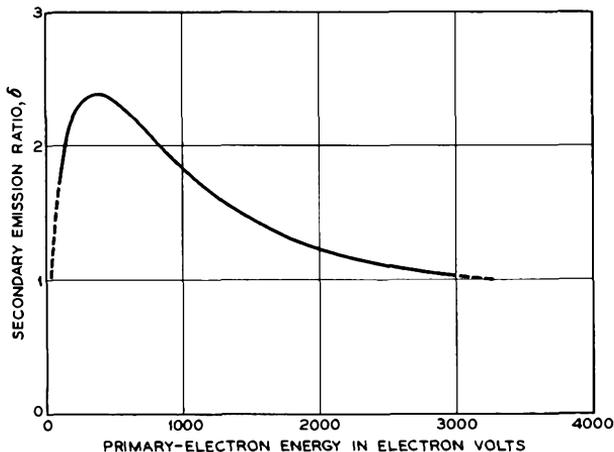
Fɪɢ. 2.5-4 Secondary emission yield vs. primary-electron energy for mica. (After H. Salow, *Z. Tech. Physik* **21**, 8, 1940)

Secondary emission is of interest in electron-tube work, sometimes as a useful effect and sometimes as an undesirable effect. In photomultiplier tubes a number of high-yield secondary emitting surfaces ("dynodes") are operated in cascade with the result that very high overall gains are obtained. In tubes using 10 dynodes, overall gains of the order of $10^5$ or $10^6$ are often obtained.

In a magnetron many of the emitted electrons are accelerated by the rf field and then driven back against the cathode with appreciable velocity. The resulting secondary-electron emission probably accounts for a majority of the total emission from the cathode, although thermionic emission is generally also needed to keep the device in operation.

On the other hand, secondary emission often adversely affects the characteristics of certain multielectrode tubes. One such tube is the tetrode vacuum tube, described in Chapter 5. If the screen grid of a tetrode is more positive than the anode, secondary electrons emitted from the anode are drawn to the screen grid; and over a range of the operating parameters, the anode current of some tubes decreases with increasing anode voltage.

Sometimes circuits employing grid-controlled tubes have been found to be bistable because of secondary emission from one of the grids. In one state a grid connected through a resistance to a fixed positive supply acts as an electron collector, $\delta$ being less than 1.0. In this case, the potential of the grid is biased to a value less than that of the fixed supply. In the other state the grid potential is above that of the fixed supply, with the result that the electrons arrive with more velocity causing $\delta$ to be greater

than 1.0, and the grid acts as an electron emitter. To obtain this second
state there must be another electrode nearby at a still higher potential in
order to collect the emitted electrons.

## 2.6   Photoelectric Emission

Photoelectric emission results from the interaction of photons incident
upon a solid and electrons within the solid. The interaction is such that an
individual photon imparts all its energy to a single electron within the solid.
If the photons are incident upon a conductor or semiconductor and if the
photon energy is greater than the work function of the surface, a fraction
of the photons will be effective in causing the emission of electrons from the
surface.

The energy of a photon is given by $h\nu = hc/\lambda$, where $h$ is Planck's con-
stant, $\nu$ is the frequency of the radiation, $c$ is the velocity of light, and $\lambda$
is the wavelength of the radiation. Expressed in electron volts, the photon
energy is equal to $12.4 \times 10^3/\lambda$ electron volts, where $\lambda$ is measured in
angstrom units. (The visible spectrum extends from about 3800 to 7600
angstroms corresponding to photon energies ranging from 3.3 to 1.6 electron
volts, respectively.)

At room temperature very few electrons in a conductor are in energy
states above the Fermi level, and hence the maximum energy of the escap-
ing electrons is very nearly given by $h\nu - \phi$, where $\phi$ is the work function
of the surface. If the work function is greater than 3.3 electron volts, only
ultraviolet radiation will cause photoelectron emission from the surface.

Figure 2.6-1 shows plots of photoelectric emission from clean surfaces of
several metals, semimetals, and one semiconductor as a function of the
energy of the incident photons. The ordinate in the figure indicates the
fraction of the photons incident upon the surface that cause emission of a
photoelectron. This fraction is called the quantum efficiency of the surface.
Since visible light corresponds to photon energies in the range between 1.6
and 3.3 electron volts, we see that some surfaces respond only to ultraviolet
light. Thus a clean platinum surface requires photon energies in excess of
4.6 electron volts to cause electron emission. The quantum efficiencies of
clean metal surfaces are generally extremely low, of the order of $10^{-4}$
or $10^{-5}$, and consequently clean metal surfaces have found little application
in photoelectric devices.

Much better photoelectron yields are obtained from certain compound
surfaces.[34] Figure 2.6-2 shows the relative photoelectric emission of several

---

[34]The preparation of high-yield photoemissive surfaces is described in Reference 2k,
Chapter 1.

WAVELENGTH IN ANGSTROMS
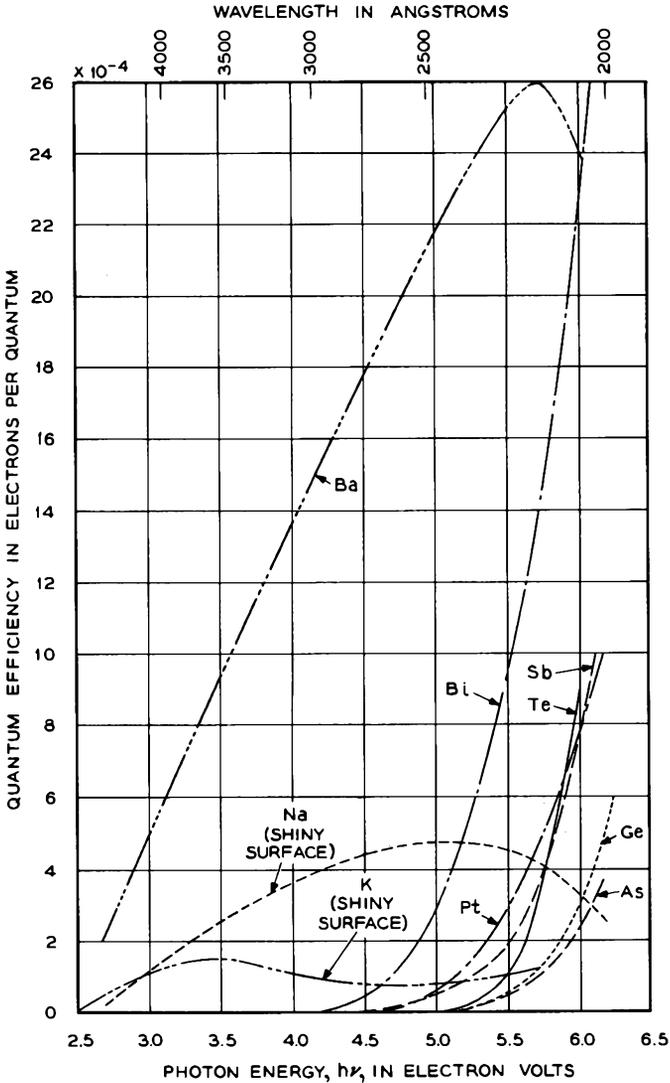


FIG. 2.6-1 Quantum efficiency of clean surfaces of several metals, semimetals, and one semiconductor as a function of the energy of the incident photons. (After *Phys. Rev.* **74**, 1462, 1948; **76**, 270, 1949; **84**, 508, 1951; **81**, 612, 1951)

frequently used compound surfaces as a function of the wavelength of the incident light. The S-11 surface is a cesium-antimony film so processed that

it is largely $Cs_3Sb$.  Often the film is deposited as a layer several hundred angstroms thick on a part of the glass envelope of the tube.[35]  Light is shone through the glass at the photoemissive surface, and the photoelectrons are drawn from the opposite side, or vacuum side, of the surface.  Electrical contact to the photoemissive surface is made through an evaporated film
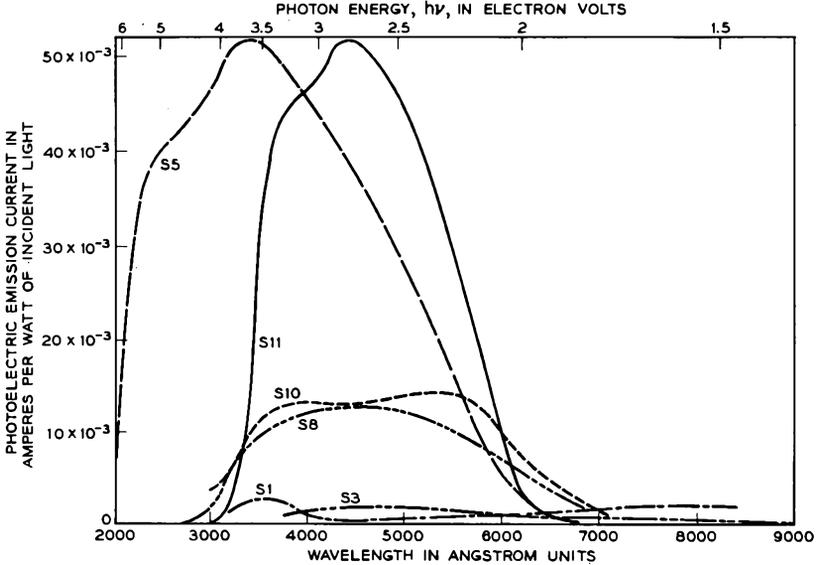


FIG. 2.6-2 Photoelectric emission current from several frequently used compound surfaces as a function of wavelength of the incident light.  The curves are for equal values of incident radiant flux at all wavelengths.  (Courtesy Allen B. DuMont Laboratories, Inc., Clifton, N.J.)

of metal which is deposited on the glass around the edge of the region through which the light is shone.  The S-5 surface is also a cesium-antimony surface, the processing being somewhat different from that for the S-11 surface.  The high response of the S-5 surface in the ultraviolet region is obtained by making the tube envelope of a glass which is transparent to ultraviolet light.

The S-8 and S-10 surfaces are compound surfaces prepared from silver, bismuth, and cesium.  Cathodes having the S-3 response have compound surfaces of silver, rubidium oxide, and rubidium, whereas those having the S-1 response have compound surfaces of silver, cesium oxide, and cesium.

---

[35]Sometimes a layer of manganese oxide is first deposited on the glass.

Both the S-3 and S-1 surfaces have relatively low photoelectron yields but are capable of emission at wavelengths extending well into the infrared region. The upper limit of sensitivity for the S-3 surface occurs at about
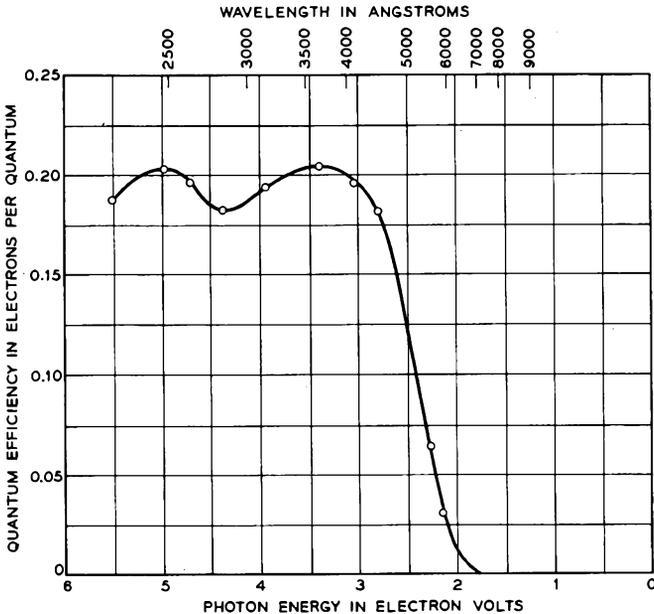


FIG. 2.6-3 Quantum efficiency of a thick film of $Cs_3Sb$ as a function of photon energy in electron volts. The data refer to electron emission on the same side of the film as that on which the light was incident.

9000 angstroms, and that for the S-1 surface occurs at about 12,000 angstroms.

The ordinate in Figure 2.6-2 is amperes per watt of incident light, which is numerically equal to coulombs per joule or electrons per electron volt. Hence multiplying the ordinate of the curve by the photon energy in electron volts (i.e., electron volts per photon) gives the quantum efficiency in electrons per photon. Thus the S-11 curve has a maximum of nearly $52 \times 10^{-3}$ amp/watt at a wavelength of 4400 angstroms. The corresponding photon energy is 2.8 electron volts, and the quantum efficiency of the surface for photons of this energy is $52 \times 10^{-3} \times 2.8 = 0.15$ electron/photon. Figure 2.6-3 shows results of measurements of the quantum efficiency of a thick film of $Cs_3Sb$ when light is incident upon the same side

of the film as that from which the emission takes place. Over a range of wavelengths from 2200 angstroms to 4000 angstroms, about one quantum in five is effective in causing the emission of a photoelectron.

*REFERENCES*

Several general references on thermionic emission are:

2a. C. Herring and M. H. Nichols, *Revs. Modern Phys.* **21**, 185, 1949.

2b. A. S. Eisenstein, *Advances in Electronics* **I**, 1, 1948.

2c. G. Hermann, S. Wagner, *The Oxide-Coated Cathode,* Vols. I and II, Chapman and and Hall, London, 1951.

2d. W. B. Nottingham, *Handbook of Physics*, Vol. 21, p. 1, Julius Springer, Berlin, 1956.

Several general references on secondary emission and photoelectric emission are:

2e. K. G. McKay, "Secondary Electron Emission," *Advances in Electronics* **I**, 65, 1948.

2f. H. Bruining, *Physics and Applications of Secondary Electron Emission*, McGraw-Hill Book Co., Inc., New York, 1954.

2g. O. Hachenberg and W. Brauer, "Secondary Electron Emission from Solids," *Advances in Electronics and Electron Physics* **XII**, 413, 1959.

2h. V. K. Zworykin and E. G. Ramberg, *Photoelectricity and Its Application*, John Wiley and Sons, Inc., New York, 1949.

2i. G. L. Weissler, "Photoionization in Gases and Photoelectric Emission from Solids," *Handbook of Physics*, Vol. 21, p. 304, Julius Springer, Berlin, 1956.

2j. P. Görlich, "Recent Advances in Photoemission," *Advances in Electronics and Electron Physics* **XII**, 1, 1959.

The preparation of high-yield secondary-electron and photo-electron emitting surfaces is described in Chapter 1 of the following reference:

2k. V. K. Zworykin and G. A. Morton, *Television*, 2nd Ed., John Wiley and Sons, Inc., New York, 1954.


Other references covering specific subjects discussed in Chapter 2 are:

2.1 J. C. Slater, *Introduction to Chemical Physics*, McGraw-Hill Book Co., Inc., New York, 1939.

2.2 H. B. Michaelson, *J. Appl. Phys.* **21**, 536, 1950.

2.3 E. W. Muller, *J. Appl. Phys.* **26**, 732, 1955.

2.4 J. Millman and S. Seely, *Electronics*, 2nd Ed., McGraw-Hill Book Co., Inc., New York, 1951.

2.5 A. R. Hutson, *Phys. Rev.* **98**, 889, 1955.

2.6 D. A. Wright, *Proc. Inst. Elec. Engrs. (London)*, Part III, Vol. 100, 125, 1953.

2.7 W. H. Kohl, *Materials Technology for Electron Tubes*, Reinhold Publishing Corp., New York, 1951.

2.8 M. R. Andrews, *Phys. Rev.* **33**, 454, 1929.

2.9 I. Langmuir, *Phys. Rev.* **22**, 357, 1923.

2.10 L. S. Nergaard, *RCA Rev.* **13**, 464, 1952.

2.11 F. A. Horak, *J. Appl. Phys.* **23**, 346, 1952.

2.12 H. J. Robinson, private communication

2.13 A. Venema, R. C. Hughes, P. P. Coppola, and R. Levi, *Philips Tech. Rev.* **19**, 177, 1957.

2.14  E. S. Rittner, R. H. Ahlert, and W. C. Rutledge, *J. Appl. Phys.* **28**, 156, 1957.
2.15  D. MacNair, private communication.
2.16  H. Bruining and J. H. DeBoer, *Physica* **5**, 17, 1938.
2.17  H. Bruining, *Philips Tech. Rev.* **3**, 80, 1938.
2.18  R. Kollath, *Physik. Z.* **38**, 202, 1937.
2.19  E. Rudberg, *Phys. Rev.* **50**, 138, 1936.
2.20  P. Nozières and D. Pines, *Phys. Rev.* **113**, 1254, 1959.  Also D. Pines, *Revs. Modern Phys.* **28**, 184, 1956, and D. Pines, *Solid State Phys.* **1**, 367, 1955.
2.21  P. L. Copeland, *Phys. Rev.* **58**, 604, 1940.
2.22  K. G. McKay, *Phys. Rev.* **61**, 708, 1942.
2.23  H. Salow, *Z. Tech. Physik* **21**, 8, 1940.
2.24  L. Apker, E. Taft, and J. Dickey, *Phys. Rev.* **74**, 1462, 1948; **76**, 270, 1949; **84**, 508, 1951.
2.25  J. Dickey, *Phys. Rev.* **81**, 612, 1951.
2.26  J. A. Burton, private communication.
2.27  R. Kollath, *Handbook of Physics*, Vol. 21, p. 232, Julius Springer, Berlin, 1956.

# Chapter 3

## BEAMS AND LENSES

An axially symmetric electric or magnetic field can be used to focus a beam of electrons much as a light lens focuses visible rays. Figure 3-1 illustrates the focusing action of a converging light lens. Rays that pass through the lens close to the axis and in directions nearly parallel to the
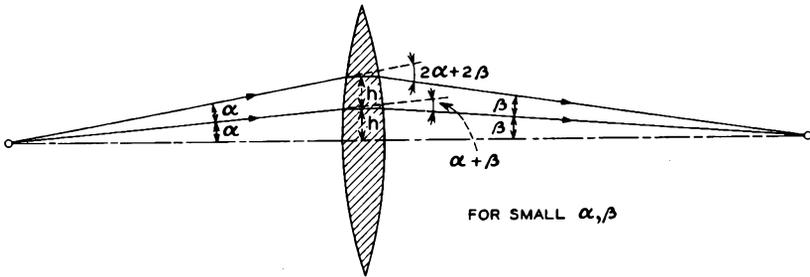


FIG. 3.1    The focusing action of a converging light lens on rays which are close to the axis and nearly parallel to the axis.

axis are given a deflection which is proportional to the distance of the rays from the axis.

An axially symmetric electric or magnetic field, or a combination of the two, acts in a similar manner on the trajectories of electrons traveling through the field. Electrons that enter the field along paths close to the axis and nearly parallel to the axis experience a radial force which is proportional to the distance of the electrons from the axis. The electron trajectories therefore are deflected in proportion to their distance from the axis, and the axially symmetric field acts as a lens. Figures 3-2(a) and 3-2(b) illustrate an electric electron lens and a magnetic electron lens, respectively.

A considerable parallelism exists between the geometric relations that govern the paths of light rays through a light lens and those that govern the trajectories of electrons through an electron lens. However, it will be useful to note several important differences between the two kinds of lenses.



(a)                                             (b)
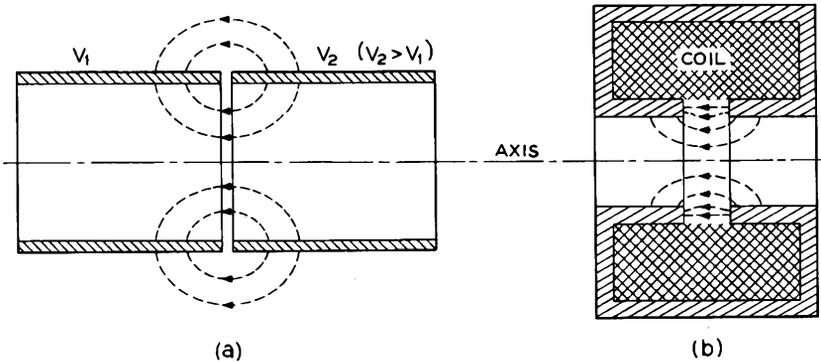
Fig. 3.2   An electric electron lens and a magnetic electron lens.

In the first place, the boundaries of a light lens are usually well defined, whereas an electron lens between field-free regions has no well-defined boundaries, since the field approaches zero asymptotically at the ends of the lens. Rays passing through a light lens, such as that illustrated in Figure 3-1, suffer abrupt changes in direction in passing between the different media that make up the lens, but the electron trajectories in an electron lens change only in a continuous manner. The electron lens has greater versatility than a light lens in that its strength can be varied merely by changing the field intensity. However, we shall find that a charge-free region of axially symmetric electric or magnetic field can act only as a converging lens on a beam of electrons whose path begins and ends in regions of zero field. In this respect there is no counterpart to the diverging lens of light optics. Furthermore, aberrations in electron lenses are generally greater than in light lenses, and correcting for the aberrations is much more difficult. Finally, a magnetic electron lens causes a rotation of the image about the axis of the lens, and there is no counterpart to this in light optics.

It will be convenient to make use of several simplifications in notation in expressing the equations we shall use in this chapter. A single dot over a variable will be used to indicate the first derivative with respect to time, and a double dot will indicate the second derivative with respect to time. Thus $\dot{r} = dr/dt$, and $\ddot{r} = d^2r/dt^2$. Similarly $r' = dr/dz$, and $r'' = d^2r/dz^2$,
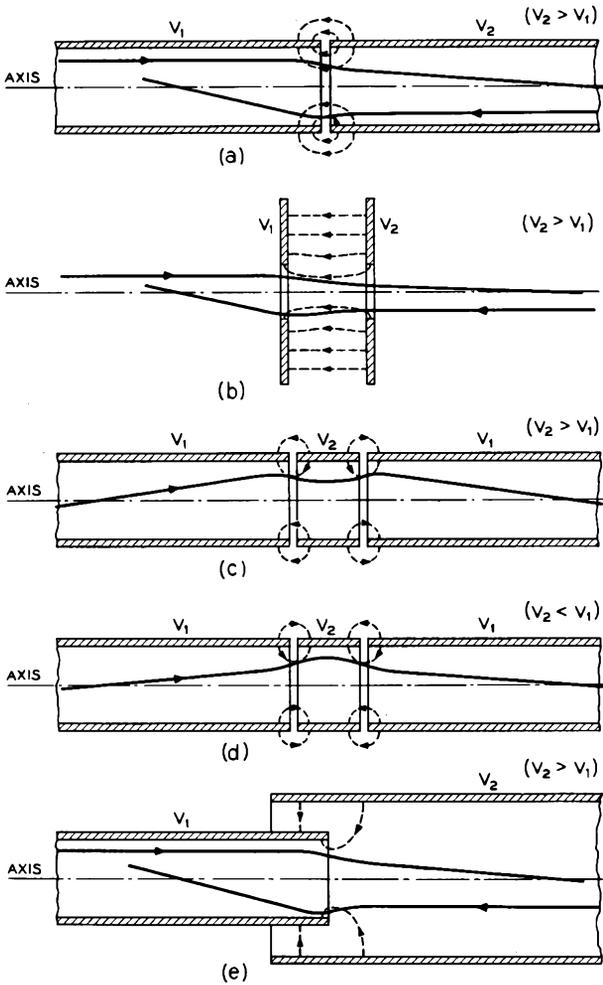
FIG. 3.1-1 Several electric electron lenses. Approximate shapes of trajectories of
electrons passing through the lenses are shown by the solid lines.

where $z$ is the axial coordinate. The ratio $e/m$ appears frequently in the
equations, and we shall denote it by $\eta$.

## 3.1   Electric Lenses

Figure 3.1-1 illustrates several types of electric lenses. The approximate
shapes of trajectories of electrons passing through the lenses are shown in

the figure. The lens shown in Figure 3.1-1(a) is formed by two coaxial cylinders of equal radius, the one at the right being at higher potential than the one at the left. An arrow on the upper trajectory indicates that it is the trajectory of an electron which passes through the lens from left to right. To the left of the gap the electron experiences a radial force tending to deflect it toward the axis, whereas to the right of the gap the radial force is directed away from the axis. However, since the electron is further from the axis to the left of the gap, and since the radial component of the field increases with distance from the axis, the inward force to the left of the gap is stronger. Furthermore, the electron travels more slowly to the left of the gap because it is in a region of lower potential. Consequently, the trajectory receives a net deflection toward the axis, and some distance to the right of the lens the electron crosses the axis.

The lower trajectory shown in Figure 3.1-1(a) is that of an electron which travels from right to left. As the electron enters the field, it is at first deflected away from the axis. However, after passing the gap, the electron travels more slowly, and since it is further from the axis, it experiences a relatively strong inward force. The electron, therefore, receives a net deflection toward the axis in passing through the lens.

Similar reasoning applies to the other electron trajectories shown in Figure 3.1-1. Each trajectory is reversible in the sense that an electron emerging from the lens would follow the same path back through the lens if its direction of travel were reversed without changing the magnitude of its velocity. Clearly, for a given potential difference between the electrodes, the faster an electron is traveling at the time it passes through a lens, the smaller the angle through which it will be deflected.

A particularly interesting lens is that illustrated in Figures 3.1-1(c) and 3.1-1(d). The lens focuses an electron beam for either $V_2 > V_1$ or $V_2 < V_1$. By holding $V_1$ constant and varying $V_2$, the stength of the lens can be varied without changing the electron velocity on either side of the lens. Such a lens is used in many cathode-ray tubes to focus the electron beam. It is often called an einzel lens. The German word "einzel" means "single" and is used in this case to imply that the potential and the electron velocity are the same on either side of the lens.

Let us consider the radial forces acting on an electron in an axially symmetric electric field. In Appendix V it is shown that the potential at radius $r$ from the axis of an axially symmetric potential distribution is given in terms of the potential along the axis by

$$V(z,r) = V(z,0) - \frac{r^2}{4}V''(z,0) + \frac{r^4}{64}V''''(z,0) - \ldots \qquad (3.1\text{-}1)$$

where $V(z,0)$ is the potential along the axis, and the primes indicate differentiation with respect to $z$. By means of Equation (3.1-1) the potential

at all points in an axially symmetric potential distribution can be described in terms of the potential on the axis. For regions close to the axis we can neglect all but the first two terms of this expression, so that

$$V(z,r) = V(z,0) - \frac{r^2}{4}V''(z,0) \qquad (3.1\text{-}2)$$

and the radial gradient of potential is given by

$$\frac{\partial V(z,r)}{\partial r} = -\frac{r}{2}V''(z,0) \qquad (3.1\text{-}3)$$

Since the radial force acting on an electron is given by $-eE_r = e(\partial V/\partial r)$, it follows that for small $r$ the radial force is proportional to the distance of the electron from the axis. If the electron is traveling nearly parallel to the axis, its trajectory is deflected by an amount proportional to the distance of the trajectory from the axis. This therefore explains the lens action of an axially symmetric electric field.

Figure 3.1-2 shows a two-cylinder electron lens in which the spacing between the cylinders is small compared with their radii. An expression[1]
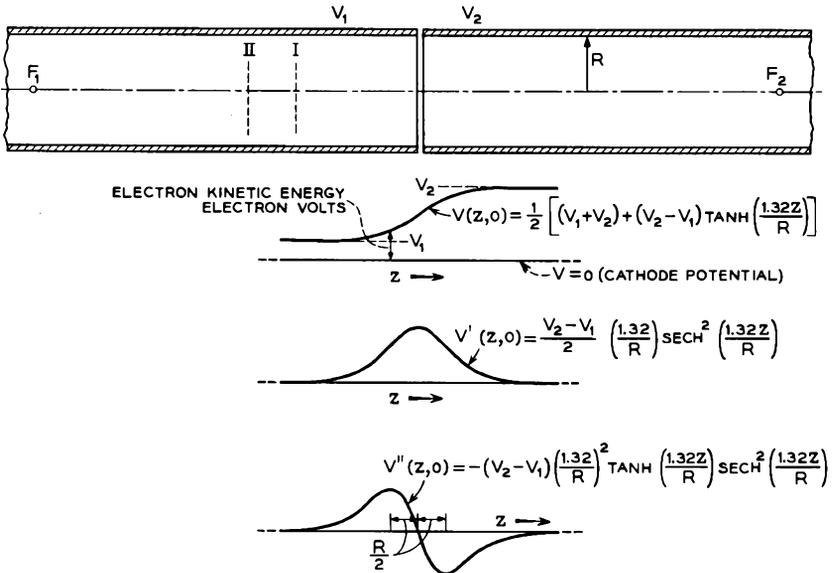


Fig. 3.1-2 A two-cylinder electric lens. The axial potential $V(z,0)$ and its first and second derivatives are shown below the lens. The positions of the principal planes and the focal points for $V_2 = 4V_1$ are indicated (see Figure 3.1-4).

[1]Reference 3.1.

for the potential along the axis $V(z,0)$ is plotted in the figure together with plots of $V'(z,0)$ and $V''(z,0)$. The potential $V(z,0)$ varies only slightly with changes in spacing between the cylinders, provided the spacing remains small compared with the radii of the cylinders. From the foregoing discussion it is evident that the radial force on an off-axis electron is proportional to the product of the distance of the electron from the axis and $V''(z,0)$. To the left of the mid-point between the cylinders the radial force is directed toward the axis, and for a given value of $r$ it reaches a maximum $R/2$ to the left of the mid-point, where $R$ is the radius of the cylinders. To the right of the mid-point the radial force is outward, and for a given value of $r$ it reaches a maximum $R/2$ to the right of the mid-point.

The equation describing the trajectory of an electron that travels nearly parallel to the axis of an axially symmetric electric field and at a small distance $r$ from the axis is known as the paraxial-ray equation. It will be helpful to derive this equation, since we shall use it in later discussion. From Equation (3.1-3), we can express the radial force acting on an electron as

$$m\ddot{r} = e\frac{\partial V(z,r)}{\partial r} = -\frac{er}{2}V''(z,0) \tag{3.1-4}$$

Now

$$\dot{r} = r'\dot{z} \tag{3.1-5}$$

and

$$\ddot{r} = r''(\dot{z})^2 + r'\ddot{z} \tag{3.1-6}$$

If the electron trajectory is nearly parallel to the axis, $\dot{z}$ will be approximately equal to the total velocity of the electron, and $(\dot{z})^2$ can be equated to $2\eta V(z,r)$, where $\eta = e/m$, and $V(z,r)$ is measured relative to cathode potential. The quantity $\ddot{z}$ on the right-hand side of Equation (3.1-6) is equal to the instantaneous acceleration of the electron in the $z$ direction, or $\eta V'(z,r)$. Furthermore, $V(z,r) \approx V(z,0)$ and $V'(z,r) \approx V'(z,0)$, so that Equation (3.1-6) can be rewritten as

$$\ddot{r} = 2\eta V(z,0)r'' + \eta V'(z,0)r' \tag{3.1-7}$$

Combining this with Equation (3.1-4), we obtain

$$r'' + \frac{V'(z,0)}{2V(z,0)}r' + \frac{V''(z,0)}{4V(z,0)}r = 0 \tag{3.1-8}$$

This is the paraxial-ray equation which we set out to derive. Several important conclusions can be drawn from it:

1. If $r_1(z)$ and $r_2(z)$ are two independent solutions of the equation, then $ar_1(z) + br_2(z)$ is also a solution of the equation, and, in fact, any solution $r_3(z)$ can be expressed in the form $r_3(z) = ar_1(z) + br_2(z)$.

2. Since the equation is homogeneous in $V$, increasing the electrode potentials in the same proportion does not change the shape of the trajectory through the lens. Furthermore, the equation is independent of $e$ and $m$, so that an electron or a negatively charged ion accelerated through the same potential rise and entering the lens along the same trajectory would follow the same path through the lens.

Equation (3.1-8) can be expressed in a second useful form by substituting

$$r = SV^{-1/4} \qquad (3.1\text{-}9)$$

where $V = V(z,0)$. This leads to

$$S'' + \frac{3}{16}\left(\frac{V'}{V}\right)^2 S = 0 \qquad (3.1\text{-}10)$$

Let us integrate Equation (3.1-10) along the axis of an axially symmetric field from one region of zero field to another. We obtain

$$S_2' - S_1' = -\frac{3}{16}\int_{z_1}^{z_2}\left(\frac{V'}{V}\right)^2 S\,dz \qquad (3.1\text{-}11)$$

where $z_1$ and $z_2$ are the $z$ coordinates of two points on the axis on opposite sides of the lens and at which the potential gradient is zero. The point $z_1$ is assumed to be on the initial side of the lens, and the point $z_2$ is on the final side. Since the integrand on the right-hand side is always greater than zero, it follows that

$$S_2' - S_1' < 0 \qquad (3.1\text{-}12)$$

Now $S = rV^{1/4}$, and $S' = r'V^{1/4} + rV'/4V^{3/4}$. Consider an electron which approaches the lens along a path that is parallel to the axis but displaced from it. For such an electron

$$S_1' = |\, r'V^{1/4} + rV'/4V^{3/4}\,|_{z_1} = 0$$

since $r' = V' = 0$ at $z = z_1$. It follows from Equation (3.1-12) that $S_2' < 0$. However, at $z = z_2$, $V' = 0$, so that $r' < 0$ at $z = z_2$. Thus the path of the electron is bent toward the axis by the field, and we must conclude that *all charge-free regions of axially symmetric electric field between field-free regions act as converging lenses.*

If an electron approaches a lens along a path that is parallel to the axis but displaced from it, the electron emerges from the lens as though it were deflected at a plane which generally is not at the geometrical center of the lens. This effect is illustrated in Figure 3.1-3 for several trajectories passing through an axially symmetric field at different distances from the axis. The plane at which the trajectories appear to have been deflected is called a *principal plane,* and the point where the electrons ultimately cross the
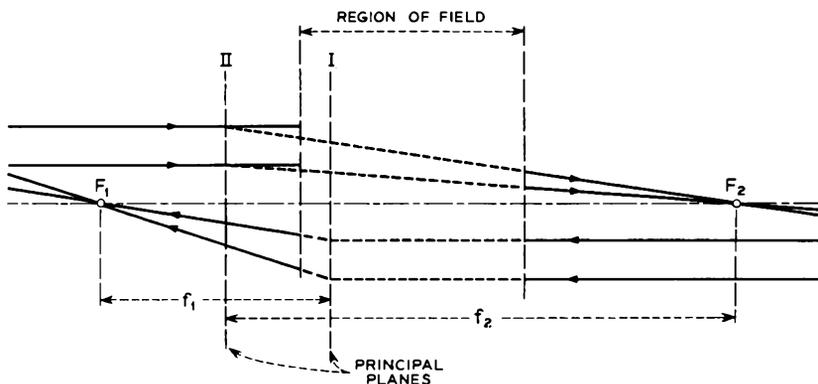
FIG. 3.1-3 The principal planes, the focal lengths, and the focal points of a lens.

axis is called a *focal point*. The distance from the principal plane to the focal point is called the *focal length*. There are two principal planes I and II, two focal points $F_1$ and $F_2$, and two focal lengths $f_1$ and $f_2$, one of each associated with electrons moving in either direction through the lens. If the electrodes and their potentials are symmetrical about the geometrical mid-point of the lens, as in the case of the einzel lens shown in Figures 3.1-1(c) and 3.1-1(d), the focal points and principal planes are also symmetrically located about the mid-point. However, in the case of the lens shown in Figure 3.1-2, where the potentials are not the same on either side of the geometrical mid-point, the principal planes are displaced toward the low-voltage side of the lens, and the focal lengths are not equal. The location of the principal planes and focal points is shown in Figure 3.1-2 for the case in which the potential of the right-hand cylinder is four times that of the left-hand cylinder.

Mathematical expressions for the potential $V(z,0)$ along the axis of a lens are available for only a few electrode configurations, one example being the two-cylinder lens of Figure 3.1-2. Goddard[2] has used the expression for $V(z,0)$ given in Figure 3.1-2 to obtain solutions of the paraxial-ray equation for the case of electrons which approach the lens along paths that are parallel to the axis but displaced from it. In this way the positions of the principal planes and the focal lengths of the lens were determined as functions of $V_2/V_1$. Figure 3.1-4 shows plots of the focal lengths $f_1$ and $f_2$ and the distances $x_1$ and $x_2$ from the mid-point of the lens to the principal planes for a range of values of $V_2/V_1$. The location of the principal planes is found to remain nearly constant for $V_2/V_1$ greater than about 4.

---

[2]Reference 3.2.
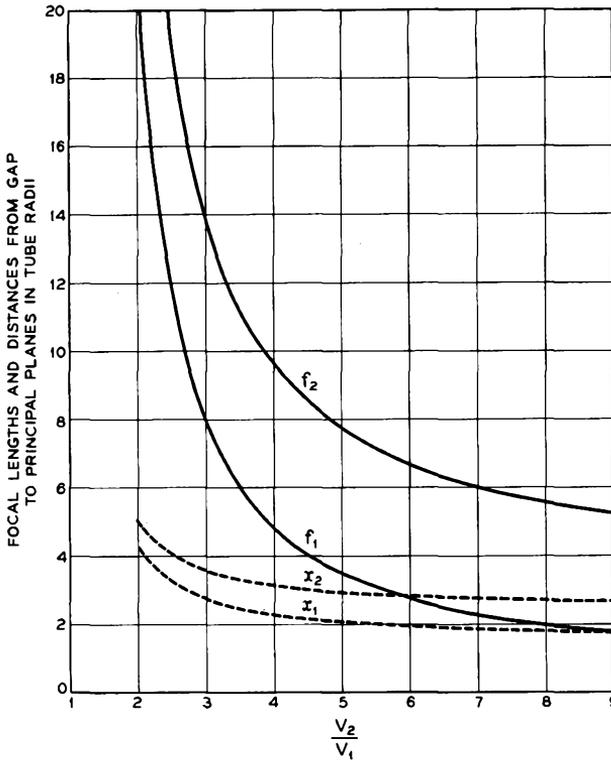
FIG. 3.1-4 The focal lengths $f_1$ and $f_2$ for a two-cylinder lens, such as that shown in Figure 3.1-2, as a function of the ratio of the potentials applied to the cylinders. The potentials $V_1$ and $V_2$ are measured relative to that of the cathode from which the electrons are emitted. The distances $x_1$ and $x_2$ from the gap between the cylinders to the principal planes are also plotted in the figure. (From L. S. Goddard, *Proc. Cambridge Phil. Soc.* **42**, 106, 1946)

In general, expressions for $V(z,0)$ are very complicated, so that an explicit solution of the paraxial ray equation is difficult, if not impossible, to obtain. Furthermore, for many electrode configurations an expression for $V(z,0)$ is not available. When there is no expression for $V(z,0)$, the electrode configuration can be simulated in an apparatus called an electrolytic tank,[3] and the axial potential can be measured experimentally. Approximate solutions to the paraxial-ray equation can then be obtained by breaking the field up into a number of segments in the axial direction

---

[3]See, for instance, Reference 3b, Figure 5.15, p. 67.

and estimating the path of an electron across each segment.[4] In a few specific electron lenses, data concerning the locations of the principal planes and focal lengths have been obtained by direct measurement of the focusing action of the lenses upon electron beams passing through them. Such data are given in Reference 3.5 and in Reference 3e, pp. 369-373.

In Appendix VI the following relations between the object position, the image position, and the focal lengths of an electron lens are derived:

$$\frac{f_1}{u} + \frac{f_2}{v} = 1 \tag{3.1-13}$$

$$\text{magnification} = \frac{\text{image size}}{\text{object size}} = \frac{f_1}{f_2}\frac{v}{u} \tag{3.1-14}$$

and

$$\frac{f_2}{f_1} = \left(\frac{V_2}{V_1}\right)^{1/2} \tag{3.1-15}$$

where the object is located $u$ units to the left of the first focal plane and the image is located $v$ units to the right of the second focal plane. The region to the left of the lens is at potential $V_1$ with respect to the cathode, and the region to the right of the lens is at potential $V_2$. From Equation (3.1-15) we see that for the two-cylinder lens shown in Figure 3.1-2 the focal length $f_2$ is twice $f_1$, when $V_2 = 4V_1$.

In a cathode-ray tube the electron gun directs the beam to a "crossover," and a lens beyond the crossover forms an image of the crossover at the screen of the tube. Using Equations (3.1-13) and (3.1-14), the position of the image and its magnification can be related to the focal lengths of the lens and the position of the crossover.

The concepts of principal planes, of focal points, and of focal lengths have been adopted from light optics, where they are used to describe the paths of light rays through lenses. The arguments that are employed in Appendix VI to derive Equations (3.1-13) and (3.1-14) apply equally well to a light lens, and, in fact, Equations (3.1-13) and (3.1-14) are of principal importance in work with light lenses. It can be shown that the square root of electric potential in the case of an electron lens is analogous to index of refraction in light optics. For a light lens at the surface between two media of different indices of refraction, the ratio of the focal lengths is given by $f_2/f_1 = n_2/n_1$, where $n_1$ and $n_2$ are, respectively, the refractive indices of the media in which the focal points $F_1$ and $F_2$ are located. The two-cylinder lens shown in Figure 3.1-2 is therefore analogous to a light lens

---

[4]Methods for making such computations are discussed in: Reference 3.3; Reference 3.4; Reference 3a, Chapter III; Reference 3b, p. 101; Reference 3e, p. 360.

at the boundary between two media of different refractive indices, whereas
the einzel lens is analogous to a light lens surrounded by a medium of the
same index of refraction.

## 3.2   Magnetic Lenses

Figure 3.2-1 shows a magnetic lens that is formed by a cylindrical per-
manent magnet and two re-entrant pole pieces. Since the magnetic poten-
tial outside the magnetic material satisfies Laplace's Equation, the off-axis
magnetic potential can be expressed in terms of the potential on the
axis using Equation (3.1-1), where $V(z,0)$ is replaced by $\psi(z,0)$, the
magnetic potential on the axis. The magnetic flux density $\mathbf{B}$ in a region
of free space is proportional to the gradient of magnetic potential, and it
follows from the magnetic equivalent of Equation (3.1-3) that the radial
component of $\mathbf{B}$ is directly proportional to $r$ for small $r$. The $z$ component
of $\mathbf{B}$, on the other hand, is nearly constant with $r$ for small $r$, since the
equipotential surfaces are normal to the axis at the points where they cross
the axis.

Consider an electron that enters the lens shown in Figure 3.2-1 from the
left along a path that is initially parallel to the axis but displaced a small
distance from it. To the left of the gap the radial component of $\mathbf{B}$ is
directed toward the axis and, since the force acting on an electron in a
magnetic field is $-e(\mathbf{u} \times \mathbf{B})$, the electron experiences a force that is
directed out of the page. This gives the electron angular momentum about
the axis, so that it crosses the $z$ component of $\mathbf{B}$ as it passes through the
central region of the lens. The $z$ component of $\mathbf{B}$ deflects the electron
toward the axis as it passes through the central part of the lens. Beyond
the center of the lens the lines of $\mathbf{B}$ have a radial component away from the
axis, and the electron loses angular velocity about the axis. We shall find
that when the electron has traveled beyond the region of field, its angular
velocity about the axis is reduced to zero. The electron therefore emerges
from the lens with a radial component of velocity, which is directed toward
the axis, and with no angular velocity. At some point beyond the lens the
electron trajectory passes through the axis.

Suppose that two electrons approach the lens along paths that are parallel
to the axis and lying in a plane containing the axis. One path is twice
as far from the axis as the second, and the radial distance from the axis to
each of the paths is small. As the electrons enter the magnetic field, the
radial component of $\mathbf{B}$ encountered by the outer electron is twice that en-
countered by the inner electron, so that the outer electron acquires twice
as much velocity in the $\theta$ direction. The *angular* velocity of the two electrons
about the axis is therefore the same, and the outer electron crosses the $z$
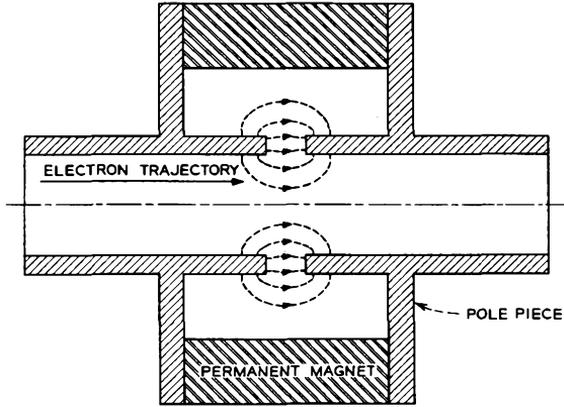
FIG. 3.2-1  A magnetic electron lens.

component of **B** with twice as much velocity and receives twice as much deflection toward the axis. Beyond the center of the lens, the outer electron again experiences twice the radial component of **B**, this time directed away from the axis, and loses twice as much velocity in the $\theta$ direction. Both electrons therefore emerge from the lens with zero angular velocity, and, since the outer electron received twice as much deflection toward the axis, both are directed toward the same point on the axis. Consequently, the trajectories to the right of the lens lie in a plane which contains the axis, but which is rotated about the axis from the plane that contained the trajectories to the left of the lens.

It will be helpful to develop two equations that describe the motion of an
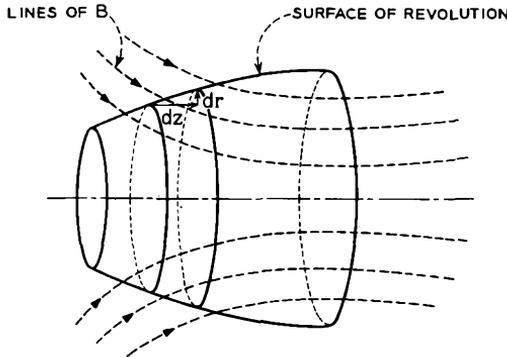


FIG. 3.2-2  A surface of revolution which contains the electron trajectory. The axis of the surface of revolution coincides with that of the field.

electron as it travels through an axially symmetric magnetic field along a path close to the axis and nearly parallel it. The electron is assumed to have zero initial angular velocity about the axis. The first equation relates the instantaneous angular velocity of the electron to the axial magnetic field, and the second describes the radial motion of the electron in passing through the field. The second equation is known as the paraxial-ray equation for magnetic fields.

Figure 3.2-2 shows a portion of a surface of revolution which contains the trajectory of an electron that enters the magnetic field along a path directed away from the axis. The axis of the surface of revolution coincides with that of the field. As the electron crosses the lines of **B**, it experiences a force in the $\theta$ direction, and from Equation (1.2-5) we can write ehat

$$\frac{d(r^2\theta)}{dt} = \eta r(\dot{r}B_z - \dot{z}B_r) \tag{3.2-1}$$

where $\theta = d\theta/dt$ and $\eta = e/m$. Multiplying by $dt$, we obtain

$$d(r^2\theta) = \eta r(drB_z - dzB_r) \tag{3.2-2}$$

Consider an incremental length of trajectory in which the electron advances a distance $dz$ in the $z$ direction and a distance $dr$ in the $r$ direction. The magnetic flux that crosses the portion of the surface of revolution corresponding to the axial length $dz$ can be expressed as

$$d\Phi = 2\pi r(drB_z - dzB_r) \tag{3.2-3}$$

where $d\Phi$ is assumed to be positive if the flux within the surface of revolution increases as $z$ increases. Combining Equation (3.2-2) with Equation (3.2-3), we obtain

$$d(r^2\theta) = \frac{\eta}{2\pi}d\Phi \tag{3.2-4}$$

Integrating this equation along the axis from a point to the left of the lens where $\theta = \Phi = 0$ to a point within the region of field, we obtain

$$r^2\theta = \frac{\eta}{2\pi}\Phi \tag{3.2-5}$$

For small $r$, $\Phi$ is related to the axial magnetic flux density by $B_z = \Phi/\pi r^2$, so that

$$\theta = \frac{\eta B_z}{2} \tag{3.2-6}$$

Thus the angular velocity of the electron at a given point on its trajectory is proportional to the $z$ component of magnetic field at that point, and when the electron has traveled beyond the region of field, its angular velocity is reduced to zero.

From Equation (1.2-4), the radial force acting on the electron is given by

$$\ddot{r} - r(\dot{\theta})^2 = -\eta r \dot{\theta} B_z \tag{3.2-7}$$

Combining this with Equation (3.2-6), we obtain

$$\ddot{r} + \left(\frac{\eta B_z}{2}\right)^2 r = 0 \tag{3.2-8}$$

Now for an electron traveling nearly parallel to the axis of an axially symmetric magnetic field in a region where the electric potential is constant, Equation (3.1-7) reduces to

$$\ddot{r} = 2\eta V r'' \tag{3.2-9}$$

where $V$ is the potential through which the electrons have been accelerated. Combining Equations (3.2-8) and (3.2-9) gives

$$r'' + \frac{\eta}{8V} B_z^2 r = 0 \tag{3.2-10}$$

This is the paraxial-ray equation for electrons traveling in an axially symmetric magnetic field when no electric fields are present. Together with Equation (3.2-6) it describes the trajectory of an electron traveling close to the axis of the field and nearly parallel to the axis. Since Equation (3.2-10) is linear, any solution of the equation can be expressed as a linear combination of any two other independent solutions.

Rewriting Equation (3.2-10) in the form

$$r'' = -\frac{\eta}{8V} B_z^2 r \tag{3.2-11}$$

we see that wherever $B_z$ is different from zero, $r''$ is negative, and the trajectory is curved toward the axis. Hence *all magnetic lenses are converging*.

A "weak" lens is one for which the focal length is long compared with the region of field. Suppose that an electron approaches such a lens along a path that is initially parallel to the axis but displaced a small distance from it. Integrating Equation (3.2-11) along the axis between points on either side of the lens where the field is zero, we obtain

$$r_2' = -\frac{\eta}{8V} \int_{z_1}^{z_2} B_z^2 r \, dz \tag{3.2-12}$$

where $z_1$ and $z_2$ are points on the axis on opposite sides of the lens and beyond the region of field, and $r_2'$ is the slope of the trajectory at $z = z_2$. If the focal length is long compared with the region of field, $r$ will remain nearly constant in the region of field and can be taken outside the integral

in Equation (3.2-12). The focal length $f$ for such a lens is then given by

$$\frac{1}{f} = -\frac{r_2'}{r} = \frac{\eta}{8V}\int_{z_1}^{z_2} B_z^2 dz \qquad (3.2\text{-}13)$$

In a magnetic lens the focal lengths $f_1$ and $f_2$ are equal; and if the magnet and pole pieces are symmetric about a central plane, the principal planes are located equal distances on either side of the central plane. In the weak lens approximation it is generally sufficient to assume that the principal planes coincide with the mid-plane of the lens.
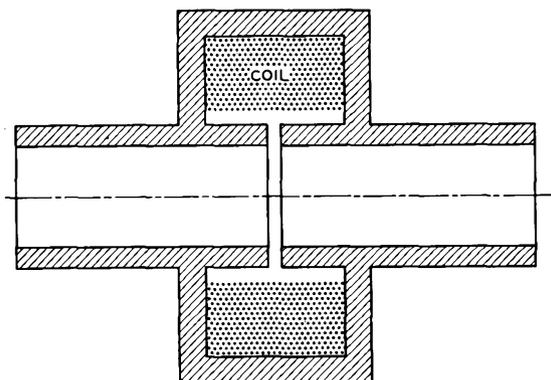


FIG. 3.2-3 An electromagnet lens.

Figure 3.2-3 shows an electromagnet lens with re-entrant pole pieces that almost touch each other. In the cylindrical region of space extending from the axis out to the pole pieces, $\int \mathbf{H}\cdot d\mathbf{l}$ around any closed path is equal to zero, and we can define a magnetic potential within the region such that the magnetic potential difference between two points is equal to $\int \mathbf{H}\cdot d\mathbf{l}$ along any path between them.[5] If the cylindrical part of the pole pieces is made of high permeability steel, so that it acts as a unipotential body, and if the spacing between the pole pieces is small compared with the inside radius of the pole pieces, a plot of magnetic potential along the axis would be of similar shape to the electric potential $V(z,0)$ plotted in Figure 3.1-2. The axial potential therefore would be proportional to tanh $(1.32z/R)$ + constant, where $R$ is the inside radius of the pole pieces. The

---

[5]However, we must confine ourselves to a region that does not surround the coil, since $\int \mathbf{H}\cdot d\mathbf{l}$ along a closed path which surrounds the coil is not zero, and the magnetic potential would not be single-valued.

axial magnetic flux density would be proportional to the gradient of this
and can be expressed as

$$B_z = B_o \operatorname{sech}^2\left(\frac{1.32z}{R}\right) \qquad (3.2\text{-}14)$$

Substituting this expression in Equation (3.2-13), we obtain for the re-
ciprocal of the focal length

$$\frac{1}{f} = \frac{\eta}{8V}B_o^2 \int_{z_1}^{z_2} \operatorname{sech}^4\left(\frac{1.32z}{R}\right)dz \qquad (3.2\text{-}15)$$

Using the relations $\operatorname{sech}^2 z = 1 - \tanh^2 z$ and $\operatorname{sech}^2 z\, dz = d(\tanh z)$, we
obtain

$$\frac{1}{f} = \frac{\eta}{8V}B_o^2 \frac{R}{1.32}\left[\tanh\left(\frac{1.32z}{R}\right) - \frac{\tanh^3}{3}\left(\frac{1.32z}{R}\right)\right]_{z_1}^{z_2}$$

$$= \frac{\eta}{8V}B_o^2 \frac{R}{1.32}\frac{4}{3} \approx \frac{\eta}{8V}B_o^2 R \qquad (3.2\text{-}16)$$

where the points $z_1$ and $z_2$ have been taken to be effectively at $-\infty$ and
$+\infty$, respectively. The focal length $f$ is therefore given by

$$f = \frac{8V}{\eta B_o^2 R} = \frac{V}{2.20 \times 10^{10} B_o^2 R} \qquad (3.2\text{-}17)$$

For $V = 10^4$ volts, $B_o = 10^{-2}$ weber/meter$^2$, $R = 2 \times 10^{-2}$ meter (2 cm),
the focal length $f$ is 0.23 meter or 23 cm. In principle, such a lens might be
used to focus the beam of a television picture tube.

If an electron trajectory on one side of a lens lies in a plane containing
the axis, the trajectory after emerging from the lens will also lie in a plane
containing the axis. However, the second plane is rotated about the axis
from the first plane. From Equation (3.2-6) the angle of rotation between
the planes is given by

$$\theta = \frac{\eta}{2}\int_{t_1}^{t_2} B_z dt = \frac{\eta}{2}\int_{z_1}^{z_2} B_z\frac{dz}{dz/dt} = \sqrt{\frac{\eta}{8V}}\int_{z_1}^{z_2} B_z dz \qquad (3.2\text{-}18)$$

where $t_1$ and $t_2$ are, respectively, the times at which the $z$ coordinate of the
electron is $z_1$ and $z_2$, and where it is assumed that $\dot{z}$ is very nearly constant
through the lens and is equal to $\sqrt{2\eta V}$. If $B_z$ is in the direction of travel of
the electron, $\theta$ is positive. In the case of the lens described above with
$V = 10^4$ volts, $B_z = 10^{-2}\operatorname{sech}^2(1.32z/R)$ weber/meter$^2$, and $R = 2 \times 10^{-2}$
meter, the angle $\theta$ is 0.45 radian, or 26 degrees.

## 3.3   Aberrations and Deflection Defocusing Effects

Like light lenses, electron lenses have aberrations, or imperfections in
their image-forming and focusing characteristics. Figure 3.3-1 illustrates

0 PRINCIPLES OF ELECTRON TUBES
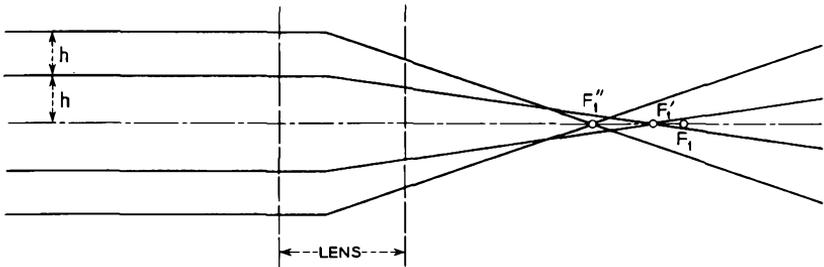


Fig. 3.3-1 Spherical aberration, or aperture defect.

one type of aberration common to all electron lenses and known as spherical aberration, or aperture defect. This is one of the most serious defects of electron lenses. Rays that pass through the lens far from the axis are focused to a different focal point than the paraxial rays. In the figure, rays that enter the lens along paths that are parallel to the axis and very close to it are focused to the point $F_1$. However, rays that are initially displaced
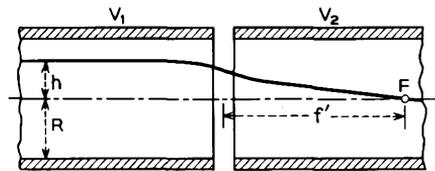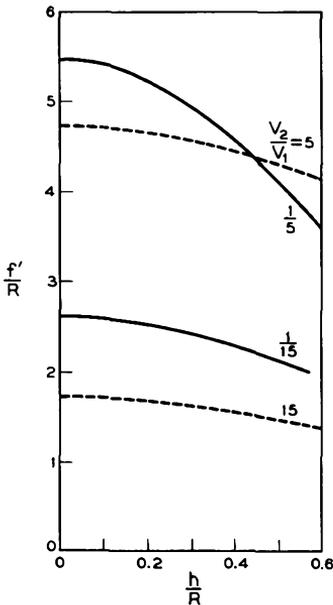


Fig. 3.3-2 Measurements of the spherical aberration in a two-cylinder electric lens. (From O. Klemperer, *Electron Optics*, 2nd Ed., Cambridge University Press, 1953)

an appreciable distance $h$ from the axis are focused to the point $F_1'$, and rays that are initially displaced a distance $2h$ from the axis are focused to the point $F_1''$. In electron lenses the focal point generally moves closer to the lens for rays that are further from the axis. The effect can be accounted for if the theory presented in Sections 3.1 and 3.2 is extended to include higher-order terms in the expressions for the off-axis fields, and if the angles through which the electron trajectories are deflected are no longer assumed to be small.

Figure 3.3-2 shows experimental data concerning the spherical aberration in a two-cylinder electron lens. Rays that approach the lens along paths that are parallel to the axis and displaced a distance $h$ from it are deflected so that they cross the axis a distance $f'$ from the geometrical midpoint of the lens. For a given semi-aperture $h$ and focal length, the spherical aberration is evidently less when the electrons are accelerated in passing through the lens than when they are decelerated. To a first approximation, the axial displacement of the focal point is proportional to the square of the lens semi-aperture $h$. Magnetic lenses are generally found to have less spherical aberration than electric lenses of comparable focal length.

Suppose that in Figure 3.3-1 a screen were placed perpendicular to the axis at $F_1$. If the semi-aperture of the lens were $2h$, the rays would strike the screen over a small circular area. Moving the screen closer to the lens would at first cause the diameter of the circular area to decrease and later to increase, the condition of best focus being that corresponding to minimum diameter of the spot on the screen. The circular spot on the screen at best focus is called the circle of least confusion, as in light optics. As the semi-aperture of the lens is decreased, the diameter of the circle of least confusion decreases, and its axial position approaches the paraxial-ray focal point $F_1$.

A second type of aberration, known as chromatic aberration, is caused by the finite distribution of electron velocities in the beam. The faster electrons in the beam are deflected less by the lens than the slower ones. Additional types of aberrations are encountered when an electron lens of large aperture forms an image of an electron source of appreciable size. Some of these aberrations have counterparts in light optics and are identified with the same names as those used in light optics. They include coma, field curvature, astigmatism, and distortion. Magnetic lenses introduce still other aberrations associated with the rotation of the image. Factors contributing to the various types of aberrations encountered in electron optics are summarized below:

1. Higher-order components in the expressions for the off-axis fields together with geometrical factors relating to the large lens aperture and large deflection angles.

2. The distribution of electron velocities (which leads to chromatic aberration).

3. Space-charge effects in which the electrons are deflected by the electric field associated with the beam itself.

4. Mechanical imperfections in the alignment or shape of the electrodes or pole pieces.

5. In the case of magnetic lenses, inhomogeneities in the magnetic material.

Although much can be done to minimize inherent aberrations in light lenses by causing the geometrical and physical properties of the lens to change with radial distance from the axis, similar corrections in electron lenses are much harder to achieve, since the off-axis fields are directly related to the axial field. Consequently, the resolution that can be achieved with a good electron lens is far less than can be achieved with a good light lens.

Changes in beam shape and focusing also occur when a beam is deflected. Figure 3.3-3 shows an electron beam that passes through a pair of deflection plates and is incident upon a planar screen mounted perpendicular to the axis of the undeflected beam. The undeflected beam is adjusted for best focus on the screen, and in this condition it is incident over a small circular region on the screen. When the beam is deflected, the spot on the screen becomes oval in shape and of area larger than that produced by the undeflected beam. Four rays, which are initially at the outer edge of the beam, are shown in the figure. Ray 1 is closest to the positive deflection plate when the beam passes between the plates, and ray 2 is closest to the negative deflection plate. Rays 3 and 4 are at the sides of the beam. Clearly, electrons in the upper part of the beam are in a region of higher potential as they pass between the plates, and they will remain in the deflecting field
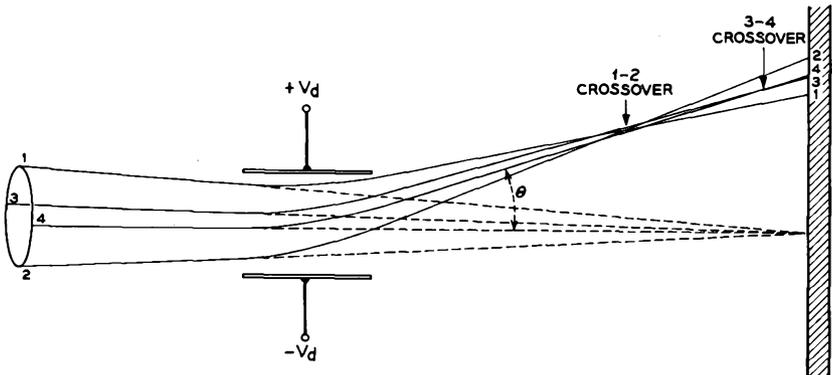


FIG. 3.3-3 The deflection defocusing effect.

for a shorter time than those in the lower part of the beam. Consequently, electrons in the upper part of the beam will be deflected less than those in the lower part of the beam, and rays 1 and 2 will cross at a point well in front of the screen. Rays 3 and 4 cross over somewhat closer to the screen but still in front of it, since the distance to the screen is further when the beam is deflected. The effect of the deflecting field in causing rays 1 and 2 to cross over sooner than rays 3 and 4 is called *deflection defocusing*. For a given mean angle of deflection and a given beam diameter, the difference between the deflection of rays 1 and 2 decreases as the length of the deflection plates increases. In applications where small deflection defocusing is particularly desirable, relatively long deflection plates are used.

Magnetic deflecting fields also cause defocusing effects similar to those illustrated in Figure 3.3-3. However, since the electrons maintain a constant velocity in passing through a magnetic deflecting field, the deflection defocusing for a given angle of deflection is less in a magnetic deflecting field than in an electrostatic deflecting field. When large deflection angles are needed, magnetic deflecting fields are usually employed. Thus in television tubes where the deflection angle (indicated by the angle $\theta$ in Figure 3.3-3) may range as high as 55 degrees, only magnetic deflection will give adequate focus of the beam over the whole screen.

In cathode-ray tubes the deflection angles are usually much smaller than in television tubes, generally less than 15 degrees, and the defocusing resulting from electrostatic deflection is usually not severe. Electrostatic deflection is preferred in cathode-ray tubes for two reasons: (1) Electrostatic deflection requires less driving power,[6] and (2) better linearity between beam deflection and the applied deflection signal can be achieved with electrostatic deflection.

### 3.4 The Spreading of an Electron Beam Because of Its Own Radial Electric Field; Focusing and Confining Beams by Applied Axial Fields

In a number of microwave tubes it is desirable to use a small-diameter electron beam with high axial charge density. Such a beam generates a

---

[6]To illustrate this point, consider the energy per unit volume which must be stored in an electric field and a magnetic field in order to produce a given amount of deflection. If the deflecting force resulting from a magnetic field $B$ is equal to that from an electric field $E$, then $Beu = eE$, and $Bu = E$. The ratio of the energy stored per unit volume in the magnetic field to that stored in the electric field is $(B^2/2\mu_o)/(\varepsilon_o E^2/2) = B^2c^2/E^2 = c^2/u^2$, where $c$ is the velocity of light, and where use has been made of the relations $\mu_o\varepsilon_o = 1/c^2$ and $Bu = E$. Since $c$ is always greater than $u$, more energy per unit volume must be stored in the magnetic field in order to produce a given amount of deflection. Furthermore, the deflecting coils are generally outside the tube so that the volume in which the energy is stored is appreciably greater with magnetic deflection. These two factors combine to require much higher driving powers in the case of magnetic deflection than with electrostatic deflection.

radial electric field, which in the absence of other applied fields causes the beam to spread, the off-axis electrons being deflected away from the axis. Usually it is desirable to prevent the beam from spreading, and this can be accomplished in several ways: (1) By directing the beam into a region of uniform axial magnetic field of sufficient intensity, (2) by directing the beam along the axis of a series of equally spaced magnetic or electric lenses of suitable strength, or (3) by directing the beam along the axis of a bifilar helix with the two windings at different potentials. We shall first describe the spreading of an electron beam because of its radial electric field. Later we shall consider the use of axial magnetic and electric fields to prevent the beam from spreading.

The axial linear charge density of a beam of current $I_o$ amperes and electron velocity $u_z$ meters per second is $I_o/u_z$ coulombs per meter. From Equation (1.4-5) the radial electric field intensity at the surface of the beam is

$$E_r = -\frac{I_o}{2\pi\varepsilon_o r u_z} \tag{3.4-1}$$

where $r$ is the beam radius. For a beam current of 10 ma, a beam diameter of 1 mm, and a beam voltage[7] of 1000 volts, Equation (3.4-1) indicates a radial electric field intensity at the surface of the beam of 19 kv/meter, or 19 volts/mm.

If the beam in the above example passed concentrically within a conducting cylinder of inside diameter 2 mm, the potential at the surface of the beam would be 6.6 volts less than that of the cylinder; and if the charge density across the beam were uniform, the potential at the center of the beam would be 11.4 volts less than that of the cylinder. However, in practice, the beam generates positive ions as a result of collisions between the electrons in the beam and molecules of residual gas in the tube. Since the radial field of the beam acts on the ions with a force directed toward the axis, the ions are entrapped by the beam. (The kinetic energy of the ions at the time they are generated is usually a small fraction of an electron volt, and this is not sufficient to overcome the potentials resulting from the radial field of the beam.) The trapping of ions by the beam in turn reduces the net axial charge density and thereby reduces the radial electric field. Generally, the ions tend to "drain" in the axial direction, since in most cases there is a region of lower potential at at least one end of the beam. The extent to which the beam charge is neutralized is therefore determined in part by the potential gradients along the axis of the beam and in part by the residual gas pressure within the tube.

Hines *et al.*[8] describe experimental measurements of the ion neutralization

---

[7] The net voltage through which the electrons have been accelerated.
[8] Reference 3.6.

of a beam having a current of 14.5 ma, a length of 17 cm, and a beam voltage of 950 volts. An axial magnetic field of 0.075 weber/meter$^2$ (750 gauss) was used to focus the beam. (See later in this section for focusing with magnetic fields.) The ions drained toward one end of the beam only, the potential at the other end being higher than that of the main portion of the beam. It was concluded that the beam was about 14 per cent neutralized with ions at a tube pressure of $10^{-7}$ mm of Hg, 50 per cent neutralized at a tube pressure of $10^{-6}$ mm of Hg, and nearly fully neutralized at a pressure of a few times $10^{-6}$ mm of Hg. Pressures of the order of $10^{-7}$ to $10^{-5}$ mm of Hg might be typical of those attained in a beam-type microwave tube.

If there is no neutralization of the electron space charge by ions, the radial motion of the electrons at the outer edge of the beam as a result of the radial electric field intensity is described by the equation

$$m\frac{d^2r}{dt^2} = mu_z{}^2\frac{d^2r}{dz^2} = -eE_r = \frac{eI_o}{2\pi\epsilon_o r u_z} \tag{3.4-2}$$

If $u_z$ is constant, this equation can be solved with the aid of tabulated functions.[9] The results are plotted in Figure 3.4-1 for the case of a beam in
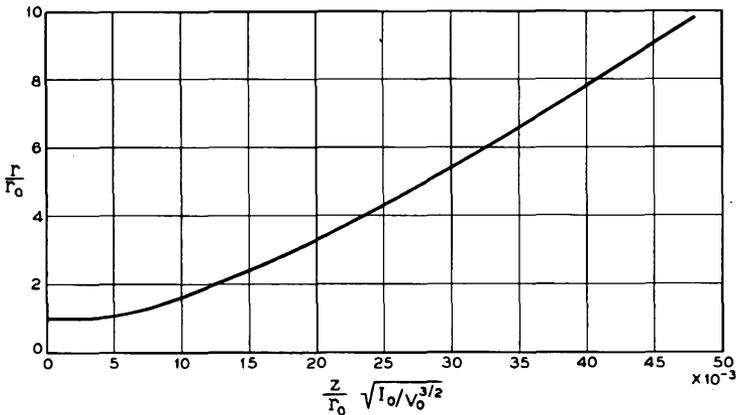


Fig. 3.4-1 The universal beam spread curve.

which the electron trajectories are assumed to be initially parallel to the axis. The plot shows the radius $r$ of the beam as a function of distance $z$ along the beam, the initial radius being $r_o$. The curve is sometimes called the universal beam spread curve. If the current density over the beam cross section is initially uniform, the radial field will deflect the trajectories

[9]Reference 3.7, p. 443.

of electrons in the interior of the beam by an amount proportional to their initial distance from the axis, and at points further along the beam, the current density over the beam cross section will still be uniform.

For a 10-ma, 1000-volt beam of initial diameter 1 mm in which the trajectories are parallel to the axis at $z = 0$ and in which no ion neutralization takes place, the beam diameter would be 1.8 *mm* one centimeter further along the axis and 9 mm four centimeters along the axis. With partial ion neutralization the spreading would be less.

In traveling-wave tubes the electron beam must travel inside a long cylindrical region defined by the slow-wave circuit of the tube with essentially no interception of the beam by the slow-wave circuit. Often the slow-wave circuit consists of a wire helix of length perhaps 70 to 250 times its inside diameter. For the beam to travel inside such a slow-wave circuit, additional fields must be applied to prevent the beam from spreading. Several methods for doing this, involving the use of axial electric or magnetic fields, are considered under separate headings below.[10]

### (a)  A Uniform Axial Magnetic Field

Figure 3.4-2 shows a magnetic circuit which produces a long region of uniform magnetic field parallel to its axis. An electron gun is located within the left-hand pole piece and, because the pole piece acts as a magnetic shield, there is essentially zero magnetic field in the region of the gun. We shall assume that the transition along the axis from the region of zero magnetic field to the full magnetic field takes place over a very short axial distance. Suppose a single electron approaches the transition region from the side of zero magnetic field along a path which is initially parallel to the axis but displaced a distance $r_o$ from it. In passing through the transition region the electron acquires an angular velocity about the axis, which from Equation (3.2-6) is given by

$$\theta = \frac{\eta B_z}{2} \qquad\qquad (3.4\text{-}3)$$

---

[10]At first thought it might seem that the beam diameter could be adequately limited by establishing a high enough gas pressure in the tube that the electron charge would be almost fully neutralized by ion charge. However, there would always be a small excess of electrons in the beam and hence a small radial field, since otherwise the ions would be free to escape. This small radial field would cause too much spreading of the beam for most traveling-wave tube applications. Furthermore, higher gas pressures result in greater ion bombardment of the cathode and shorter cathode life. High ion densities also result in mechanical oscillation of large numbers of the ions within the potential well formed by the electron beam. The ion motion modulates the beam and thereby causes a type of noise, called ion oscillation noise, to appear in the output of the tube.
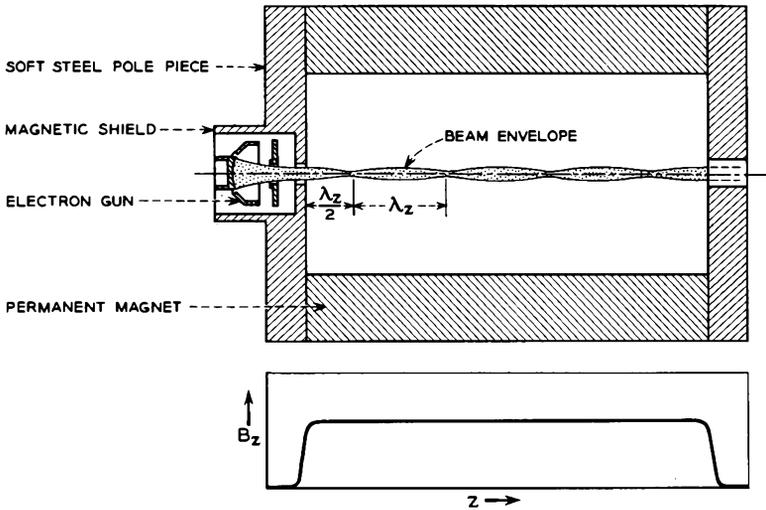
FIG. 3.4-2 An electron beam directed into a region of uniform magnetic field

where $B_z$ is the uniform axial magnetic field, and $\eta = e/m$. If the electron passes through the transition region sufficiently quickly, it will still be at distance $r_o$ from the axis upon entering the region of uniform axial field. Its velocity in the $\theta$ direction therefore will be

$$u_\theta = r_o\dot\theta = r_o\frac{\eta B_z}{2} \qquad (3.4\text{-}4)$$

This transverse velocity causes the electron to cross the lines of axial magnetic field, so that the motion of the electron in the transverse plane is circular with radius

$$r = \frac{u_\theta}{\eta B_z} = \frac{r_o}{2} \qquad (3.4\text{-}5)$$

The electron therefore travels through the uniform magnetic field in a helical path of radius $r_o/2$, and since it initially started at distance $r_o$ from the axis, with velocity only in the $\theta$ and $z$ directions, it periodically passes through the axis and returns to its original radius $r_o$. Interestingly enough, this result is independent of the magnitude of the axial magnetic field, the initial electron velocity, or the initial distance of the electron from the axis. The time taken for the electron to complete one turn of its helical path is $\pi r_o/u_\theta$, so that the points at which the electron passes through the axis are separated by an axial distance given by

$$\lambda_z = u_z\frac{\pi r_o}{u_\theta} = u_z\frac{2\pi}{\eta B_z} \qquad (3.4\text{-}6)$$

Suppose a cylindrical electron beam of very low axial charge density is directed along the axis of the magnetic circuit shown in Figure 3.4-2. We shall assume that the axial charge density is sufficiently small that the radial electric field of the beam exerts a much weaker transverse force on the off-



POSITION OF ELECTRON AT
THE TIME OF ENTRY INTO
THE MAGNETIC FIELD

BEAM ENVELOPE AT THE
POINT OF ENTRY TO THE
MAGNETIC FIELD
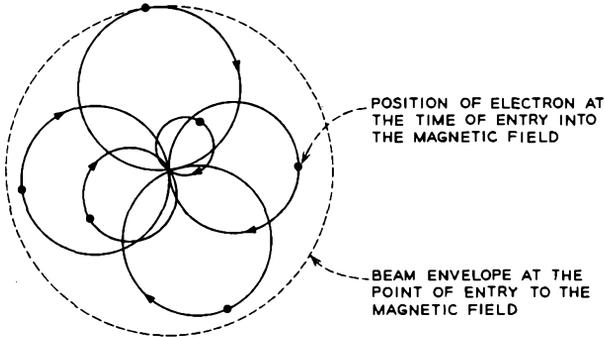
FIG. 3.4-3 The broken line shows the beam envelope at the time the beam is launched into the magnetic field. The motion of individual electrons in the transverse plane after entering the magnetic field is shown by the solid lines.

axis electrons than the force exerted by the magnetic field. In such a beam, electrons which travel along paths that are parallel to the axis just before entering the magnetic field follow helical paths in the region of uniform magnetic field with one side of the helical path touching the axis. Each electron passes through the axis at points spaced by a distance $\lambda_z$, the first point being $\lambda_z/2$ beyond its point of entry into the magnetic field. Figure 3.4-3 shows the motion of the electrons in the transverse plane. Since all the electrons pass through (or close to) the axis at nearly the same points, the beam envelope necks down from its initial radius to a very small radius at a point $\lambda_z/2$ beyond the point of entry into the magnetic field and each $\lambda_z$ thereafter. The beam envelope therefore resembles that shown in Figure 3.4-2 and is said to be "scalloped." For a 1000-volt beam and $B_z = 0.05$ weber/meter$^2$, $\lambda_z = 1.3$ cm.

Since electrons that enter the magnetic field with large $r_o$ have greater kinetic energy in the transverse plane than those that enter with small $r_o$ and since all electrons in the beam have essentially the same total kinetic energy, the outer electrons will have slightly smaller axial velocity in the region of uniform magnetic field. The total kinetic energy of an electron in the region of uniform magnetic field can be expressed as

$$eV_o = \tfrac{1}{2}m(u_\theta^2 + u_z^2) = \tfrac{1}{2}m[(r_o\eta B_z/2)^2 + u_z^2] \qquad (3.4\text{-}7)$$

where $V_o$ is the voltage through which the electrons have been accelerated. Clearly, an electron that enters the magnetic field with large $r_o$ will have smaller $u_z$, and hence smaller $\lambda_z$, in the region of uniform magnetic field than one that enters with small $r_o$. Consequently, the outer electrons in the beam will gradually slip behind the inner electrons, and the envelope will be only quasi-periodic in the axial direction.

As the axial charge density of the beam is increased, so that the transverse force from the radial electric field becomes comparable with that resulting from the axial magnetic field, the electron motion is appreciably modified. Brillouin[11] has described an exact solution for electron flow in an axial magnetic field in which the outward force resulting from the radial electric field of the beam is balanced by the inward force of the axial magnetic field. The conditions required for Brillouin flow are difficult to achieve in practice, but the solution defines a value of magnetic flux density in terms of the beam current, the beam voltage, and the beam diameter, and it is often helpful to measure the field actually needed to confine a beam to a given diameter in terms of this field.

To obtain Brillouin flow, the following conditions must apply at the point of entry of the beam to the region of uniform magnetic field:

1. The beam must have a uniform current density across its diameter.
2. The electron trajectories must be parallel to the axis just before entering the magnetic field.
3. The transition from zero axial magnetic field to the full field must occur over a very short axial distance.
4. The beam axis must coincide with that of the magnetic field.

In addition, there must be no trapping of ions by the electron beam.

In Brillouin flow an electron which enters the magnetic field at distance $r_o$ from the axis experiences a radial force which is just sufficient to keep it moving in a helical path of radius $r_o$ about the axis of the beam. The transverse force of the magnetic field must then be sufficient to account for the radial acceleration of the electron when moving in a helical path of radius $r_o$ plus the force resulting from the radial electric field at radius $r_o$. The axial magnetic field is therefore determined by the relation

$$B_z e u_\theta = \frac{m u_\theta^2}{r_o} - e E_{r_o} \qquad (3.4\text{-}8)$$

If the beam radius is $a$ and if the current density is uniform across the beam cross section, we can use Equation (3.4-1) to express the second term on the

---

[11]Reference 3.8.

right as

$$eE_{r_o} = -\frac{r_o^2}{a^2}\frac{eI_o}{2\pi\varepsilon_o r_o u_z} = -\frac{r_o}{a^2}\frac{eI_o}{2\pi\varepsilon_o(2\eta V_o)^{1/2}} \qquad (3.4\text{-}9)$$

where $V_o$ is the voltage through which the electrons have been accelerated. Combining Equations (3.4-4), (3.4-8), and (3.4-9), we obtain

$$B_z^2 = \frac{\sqrt{2}\,I_o}{\pi\varepsilon_o\eta^{3/2}V_o^{1/2}a^2} = \frac{0.69 \times 10^{-6}I_o}{V_o^{1/2}a^2} \qquad (3.4\text{-}10)$$

This equation gives the magnetic flux density required for Brillouin flow. Since it is independent of the radius $r_o$, the same magnetic field applies to all electron trajectories for which $r_o \leq a$. For a 10-ma, 1000-volt beam of diameter 1 mm, the Brillouin magnetic flux density is $2.95 \times 10^{-2}$weber/meter², or 295 gauss.

In Brillouin flow the beam envelope maintains a constant diameter through the region of longitudinal magnetic field, the individual electrons following helical trajectories which are concentric with the beam axis, and the beam as a whole twisting about its axis with angular velocity $\dot{\theta} = \eta B_z/2$. In a thin "cross-sectional slab" of the beam the individual electrons maintain their positions relative to each other, and the slab as a whole rotates about the axis with angular velocity $\dot{\theta}$.

However, in practice, the axial charge density of the beam will be partially neutralized with ions. In this case the transverse force resulting from a magnetic field equal to the Brillouin field would predominate, so that the electrons would periodically pass near to the axis, and the beam envelope would be scalloped. Furthermore, most convergent electron guns[12] give rise to sufficiently high transverse velocities that the maxima in the diameter of the scallops would be somewhat larger than the beam diameter at the point of entry into the magnetic field. (This point is further discussed in Reference 3g.) However, it is found that by increasing the magnetic field, the maximum diameter of the scallops can be reduced. Often a magnetic field equal to $1\frac{1}{2}$ to 3 times the Brillouin field is used.

As the magnetic field is increased appreciably above the Brillouin value, the transverse force resulting from the magnetic field becomes the principal transverse force acting on the electrons. Harker[13] and Ashkin[14] have concluded on the basis of experimental measurements that with a magnetic field greater than, or equal to, about three times the Brillouin field, the effects of the radial electric field can be neglected, and a majority of the

---

[12]Guns which generate a beam of smaller diameter than that of the cathode.
[13]Reference 3.10.
[14]Reference 3.11.
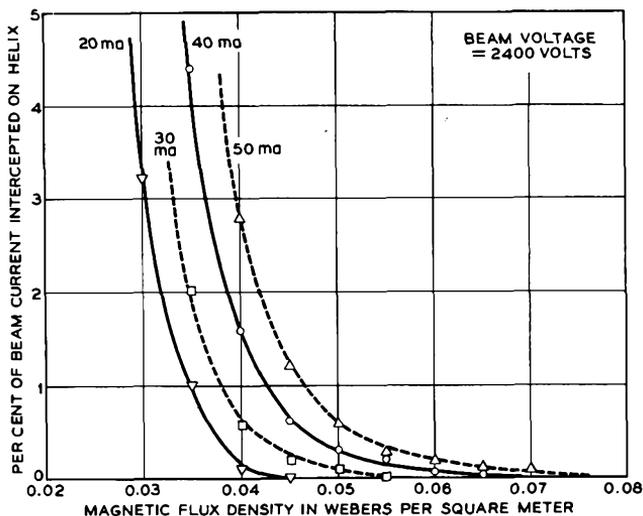
Fig. 3.4-4 Measurements of the beam interception by the slow-wave circuit of a traveling-wave amplifier as a function of the applied axial magnetic flux density. The beam was generated by the electron gun illustrated in Figure 4.5-1(a). (From J. P. Laico *et al.*, *Bell System Tech. J*. **35**, 1285, 1956. Reprinted by permission of American Telephone and Telegraph Company)

electrons pass through, or very close to, the beam axis. In this case an electron that enters the magnetic field at distance $r_o$ from the axis travels in a nearly helical path of radius $r_o/2$ and periodically passes through or close to the axis.

Figure 3.4-4 shows measurements of the fraction of the beam current intercepted on a helix-type slow-wave circuit of a traveling-wave amplifier as a function of the applied axial magnetic flux density. The data are plotted for several values of beam current. The helix had an inside radius of 1 mm and a length of 17 cm. The electron gun was similar to that shown in Figure 4.5-1(a). A plot of current density vs. radius for the electron beam at the point of entry into the magnetic field is shown in Figure 4.5-4. An electron emitted from the edge of the cathode with zero emission velocity in the direction parallel to the cathode surface arrives at the point of entry into the magnetic field at a radius of 0.45 mm from the beam axis. However other electrons emitted from the edge of the cathode with relatively high emission velocity parallel to the cathode surface arrive at the point of entry into the magnetic field at distances from the beam axis as high as 0.7 to 0.8 mm.

Figure 3.4-4 shows that with increasing beam current, a higher magnetic field was required to prevent interception of the beam by the helix, as would
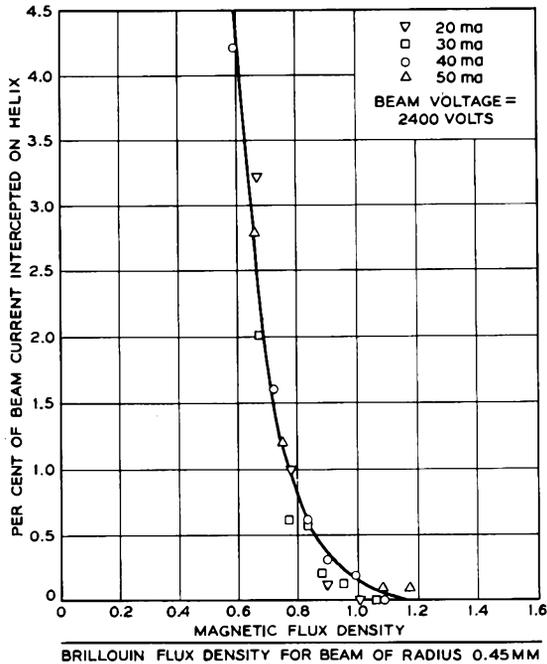
FIG. 3.4-5 The data for Figure 3.4-4 are plotted vs. the ratio of the actual magnetic flux density to the Brillouin flux density for a beam of uniform current density and a radius of 0.45 mm.

be expected from Equation (3.4-10). Figure 3.4-5 shows the same data plotted as a function of the ratio of the actual magnetic flux density to the Brillouin flux density for a beam of uniform current density and radius 0.45 mm.

## (b)  Confined Flow[15]

A type of electron flow, known as confined flow, is achieved with the electron gun entirely immersed in the magnetic field. Often a uniform axial magnetic field is used. The cathode of the electron gun might consist of a planar disc which is perpendicular to the field, whereas the accelerating electrode would have an aperture somewhat larger than the cathode diameter. An electron gun that is used with confined flow (and has several accelerating electrodes) is illustrated subsequently in Figure 4.5-1(c).

---

[15]Reference 3b, p. 161.

If the magnetic field lines are parallel to the beam axis over the entire length of the beam, starting at the cathode, the beam diameter obtained with confined flow is always larger than the cathode diameter, but it decreases and asymptotically approaches the cathode diameter as the magnetic field intensity is increased. With increasing magnetic field the electrons increasingly tend to follow the field lines, and the motion of an individual electron in the transverse plane becomes limited to a smaller and smaller area, the motion being nearly circular.

With confined flow the magnetic field required to confine a given beam current to a given diameter is always greater than that needed when the beam is generated outside the magnetic field and injected into it, as described in Section (a) above. Confined flow has found its chief application in low-noise microwave amplifier tubes, where the magnetic field in the region of the potential minimum reduces the transverse motion of the electrons and thereby effects a reduction in the noise generated by the beam.

Confined flow also can be achieved with a convergent electron gun by establishing in the region of the gun a magnetic field that converges in the same manner as the electron trajectories in the absence of the magnetic field. In this case the electrons "follow the magnetic field lines" through the accelerating region of the gun, and in the region beyond the gun their motion is much as described above.

## (c)  *Focusing with Periodic Magnetic Fields*

A series of equally spaced lenses also can be used to focus an electron beam. In this case the off-axis electrons experience a radial impulse, which is directed toward the axis, as they pass each lens. The impulses deflect the electrons toward the axis, but between lenses the beam again spreads because of the radial electric field due to the space charge. For a particular condition of lens strength and spacing and for a particular average beam radius, the impulses from the lenses just balance the integrated radial outward force resulting from the space charge of the beam, and the beam diameter at successive lenses remains constant. The shape of the beam envelope is then somewhat as illustrated in Figure 3.4-6. Focusing an electron beam
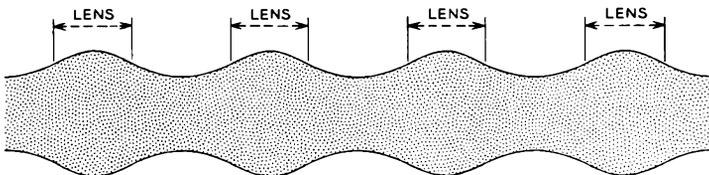


FIG. 3.4-6 The focusing action of a series of equally spaced lenses.
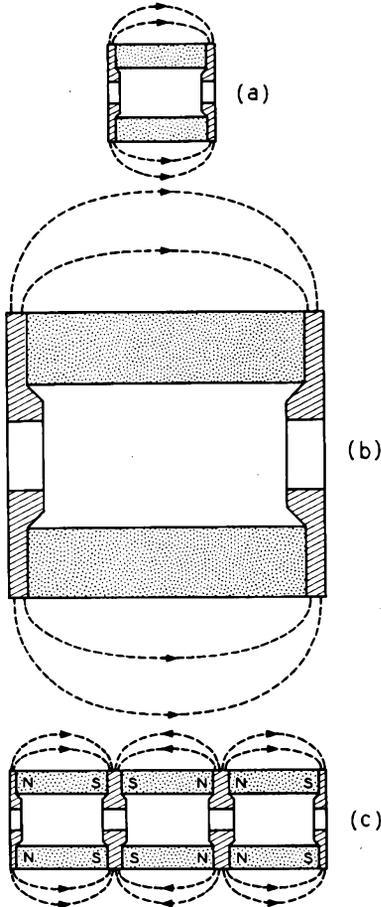
FIG. 3.4-7 The use of a periodic focusing
structure reduces the weight of magnetic
material needed to produce a given mag-
netic field over a given axial distance.
(From J. T. Mendel *et al., Proc. IRE* **42**,
800, 1954)

with a series of equally spaced lenses
is called periodic focusing, since the
axial field varies periodically in the
$z$ direction.

The stability of periodic focusing
can be made plausible by noting that
if the beam radius increases above
that needed to obtain the balance
condition, the radial impulses re-
ceived from the lenses predominate,
and the off-axis electrons receive a
net deflection toward the axis. On
the other hand, if the beam radius
becomes less than that required for
balance, the radial outward force
predominates and the beam expands.

Periodic focusing can be achieved
with both electric lenses and mag-
netic lenses. When magnetic lenses
are used, the axial fields of succes-
sive lenses are usually reversed in
direction, and in this way a substan-
tial reduction in magnet weight can
be achieved over that of a permanent
magnet or electromagnet which
would produce a uniform axial focus-
ing field.[16] To explain this, we might
first note that the magnetic circuit
shown in Figure 3.4-2 establishes a
magnetic field throughout a far
larger volume than that occupied by
the beam, and, since the total weight
of the magnet material is closely
related to the magnetic energy
stored in the space surrounding the
magnet, much of the weight of the
magnet would appear to be wasted.

Figure 3.4-7 illustrates how weight can be saved using a periodic perma-
nent magnet circuit. The magnetic circuit shown in Figure 3.4-7(a) is as-
sumed to produce a uniform axial magnetic field over the length of the

---

[16]Reference 3.12.

magnet. By increasing all the linear dimensions of the circuit by a factor of 3, as in the magnetic circuit shown in part (b) of the figure, the length of the axial magnetic field is increased by three times, but the magnitude of the axial field remains unchanged. The larger magnet weighs $3^3 = 27$ times as much as the smaller magnet, and the energy stored in the space surrounding it is 27 times as great. On the other hand, three of the smaller magnets placed end for end with like poles together (i.e., north beside north and south beside south), as in the assembly shown in Figure 3.4-7(c), have 1/9 the weight of the larger magnet and produce approximately the same axial field over the same axial distance, but with two reversals in direction. The energy stored in the space surrounding the three magnets in Figure 3.4-7(c) is approximately 1/9 that stored in the space around the larger magnet, since the leakage fields extend only 1/3 as far from the axis.

If the larger magnet were replaced with $n$ smaller permanent magnets of the same total overall length and axial magnetic field, the weight of the periodic circuit would be $1/n^2$ times that of the larger magnet. However, in practice the reversals of the axial field are not really abrupt and, in order to achieve adequate focusing of the beam, a somewhat higher peak magnetic field must be used. This requires the magnets of the periodic structure to be somewhat heavier, and consequently the weight of the periodic circuit needed to focus a given beam is between $1/n^2$ and $1/n$ that of a single permanent magnet which would focus the beam with a uniform axial field.

Periodic structures also have the advantage of much smaller leakage fields and hence less likelihood of interference with nearby devices or equipment.

Let us now examine the electron motion in a periodic magnetic field. Equations (3.1-4) and (3.2-8) can be combined to give an equation that describes the radial motion of an electron in the presence of both an axial magnetic field and a radial electric field, namely

$$\frac{d^2r}{dt^2} + \left(\frac{\eta B_z}{2}\right)^2 r - \eta\frac{dV}{dr} = 0 \qquad (3.4\text{-}11)$$

Suppose the axial magnetic field varies as a cosine function, so that

$$B_z = B_o \cos\frac{2\pi z}{L} \qquad (3.4\text{-}12)$$

where $L$ is the magnet period, or twice the center-to-center distance between adjacent pole pieces. Substituting for $dV/dr = -E_r$ from Equation (3.4-1), setting $z = u_z t$, and combining Equations (3.4-11) and (3.4-12), we obtain the following equation for the motion of an electron at the surface of the beam:

$$\frac{d^2r}{dz^2} + \left(\frac{\eta B_o}{2u_z}\cos\frac{2\pi z}{L}\right)^2 r - \frac{\eta I_o}{2\pi\varepsilon_o u_z^3}\frac{1}{r} = 0 \qquad (3.4\text{-}13)$$

It is convenient to make a change of variables and rewrite this equation in the form

$$\frac{d^2\sigma}{dT^2} + \alpha(1 + \cos 2T)\sigma - \frac{\beta}{\sigma} = 0 \qquad (3.4\text{-}14)$$

where

$$\sigma = \frac{r}{a}, \qquad T = \frac{2\pi z}{L}$$

$$\alpha = \frac{1}{2}\left(\frac{\eta B_o L}{4\pi u_z}\right)^2 = 2.79 \times 10^8 \frac{B_o^2 L^2}{V_o}$$

$$\beta = \frac{\eta I_o L^2}{8\epsilon_o \pi^3 u_z^3 a^2} = \frac{385 I_o L^2}{V_o^{3/2} a^2}$$

and where use has been made of the relations $u_z^2 = 2\eta V_o$ and $2\cos^2 T = 1$



(a)
INSUFFICIENT
MAGNETIC FIELD
$\alpha = 0.15$
$\beta = 0.2$

(b)
CORRECT
MAGNETIC FIELD
$\alpha = 0.2$
$\beta = 0.2$

(c)
EXCESS
MAGNETIC FIELD
$\alpha = 0.4$
$\beta = 0.2$
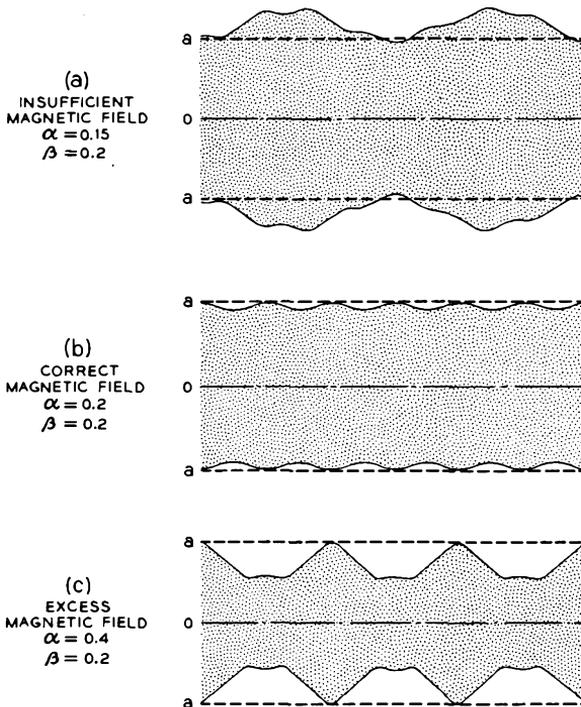
FIG. 3.4-8 The shape of the envelope of a beam for three conditions of the magnetic field parameter $\alpha$. The small ripples on the beam are associated with the pole piece spacing $L/2$. (From J. T. Mendel *et al.*, *Proc. IRE* **42**, 800, 1954)

$+ \cos 2T$. The beam radius at the point of entry into the magnetic field is assumed to be $a$.

Solutions to Equation (3.4-14) have been obtained with the aid of an analog computer[17] for the case in which: (1) The electron flow is "laminar," that is, the electron trajectories do not cross one another; and (2) the current density across the beam cross section is uniform. (Also implicit in Equations (3.4-13) and (3.4-14) is the assumption of no ion neutralization.) The shape of the beam envelope as determined by the computer for three values of the axial magnetic field is shown in Figure 3.4-8. The computer results show that minimum beam ripple is obtained when $\alpha = \beta$ or

$$B_{\text{rms}}{}^2 = 0.69 \times 10^{-6} \frac{I_o}{V_o{}^{1/2} a^2} \qquad (3.4\text{-}15)$$

where $B_{\text{rms}} = B_o/\sqrt{2}$. It will be noted that the right-hand side of this equation is the same as that of Equation (3.4-10), which gives the Brillouin field needed to focus a beam of current $I_o$, voltage $V_o$, and radius $a$. Thus for a sinusoidally varying field the rms value of the magnetic field must equal the Brillouin field. This result is perhaps not surprising, since the radial force resulting from the axial magnetic field is proportional to $B_z{}^2$, and with a sinusoidal field such that $B_{\text{rms}} = B_{\text{Brillouin}}$, the average radial force from the magnetic field is the same as with Brillouin flow.

By setting $\beta = 0$ in Equation (3.4-14), the equation reduces to a form of Mathieu Equation[18] that is characterized by solutions for $\sigma$ which are periodic in $T$ for certain ranges of $\alpha$, and which are unstable for other ranges of $\alpha$. Figure 3.4-9 shows the ranges of $\alpha$ for which the solutions are stable.
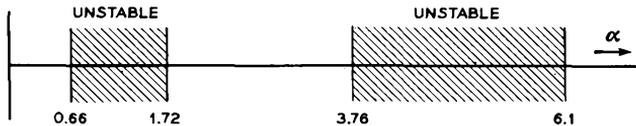


FIG. 3.4-9 The regions in which Equation (3.4-14) is unstable when $\beta = 0$.

The significance of this is that, if we reduce the beam current $I_o$ to a vanishingly small value, so that $\beta \to 0$, but maintain constant beam voltage, there will be some values of the parameters $B_o$, $L$, and $V_o$ for which the beam will be focused by the lenses and others for which $\sigma = r/a$ will be unstable and the beam will become divergent. Furthermore, it is found[19] that even with higher beam currents the periodic structure transmits practically no current in the regions marked "unstable" in Figure 3.4-9.

[17]Reference 3.13.
[18]Reference 3.14.
[19]Reference 3.13.

In practice, most periodic circuits are designed to operate in the first
"pass band" in Figure 3.4-9, corresponding to $\alpha < 0.66$. To determine the
value of $B_o$ that should be used, the beam is first focused with a solenoid
which produces a uniform axial field, and the minimum value of $B_z$ which
gives good beam transmission is measured. The value of $B_o$ for the periodic
circuit is then taken to be approximately $\sqrt{2}$ times this field (assuming that
the axial field is to vary in a nearly sinusoidal manner). The period $L$ of
the periodic circuit is then determined so that $\alpha$ is less than 0.66, perhaps
30 per cent less. For $B_o = 0.08$ weber/meter$^2$ (or 800 gauss), and a beam
voltage of 1500 volts, a period $L$ of 1.92 cm gives an $\alpha$ of 0.44. Reference
3.15 describes the design of the pole pieces and permanent magnets for a
periodic circuit.

### (d)  Periodic Focusing with Electric Fields

Tien[20] has described the focusing of an electron beam using a periodic
electric field. Such a field might be obtained with a series of ring electrodes
as illustrated in Figure 3.4-10(a) or a bifilar helix such as that illustrated in
Figure 3.4-10(b). In both cases the outer electrons experience a relatively
strong force toward the axis when they are close to the electrodes at the



(a) $V_2 > V_1$                              (b) $V_2 > V_1$
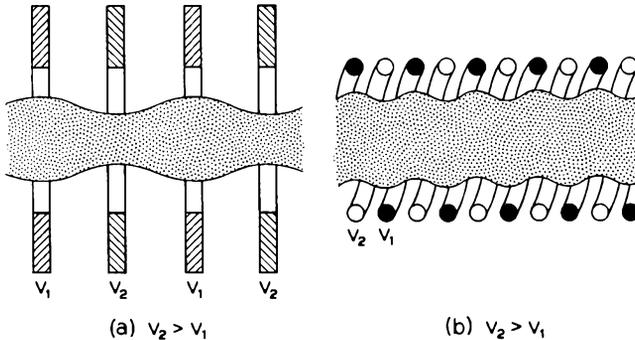
FIG. 3.4-10  Periodic focusing of a beam with electric fields: (a) with a series of ring
electrodes, and (b) with a bifilar helix.

lower potential and a somewhat weaker outward force when they are
opposite the electrodes at the higher potential. Also, their axial velocity is
less when they experience the inward force than when they experience the
outward force. Consequently, there is a net focusing effect that can be
used to balance the outward force of the radial electric field of the beam.
As in the case of periodic focusing with magnetic fields, the beam radius is

---

[20]Reference 3.16.

found to be stable for some values of the focusing parameters, whereas for others it becomes divergent.

Tien pointed out that the bifilar helix also can be used as the slow-wave circuit of a traveling-wave amplifier. Such a tube has been developed by RCA.[21] The helix structure of the RCA tube is illustrated in Figure 3.4-11. The length of the helix is 22 cm. Figure 3.4-12 shows measurements of the per cent beam current intercepted by the helix as a function of the voltages applied to the helix. It can be seen that good focusing is achieved over a range of helix voltages.
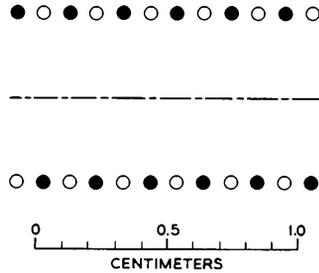
FIG. 3.4-11 The bifilar helix used in an electrostatically focused traveling-wave amplifier developed by RCA. (Courtesy Radio Corporation of America)
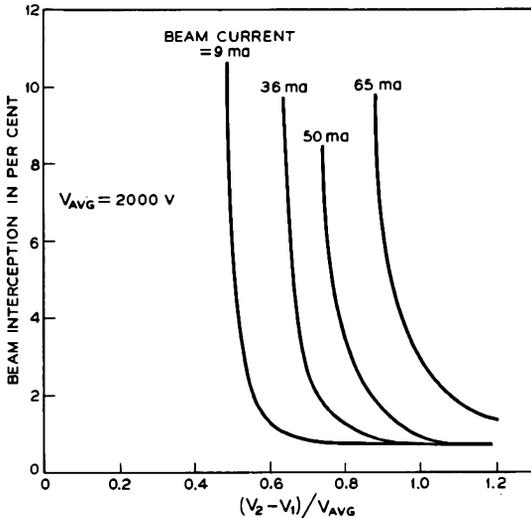
FIG. 3.4-12 Beam interception on the helix structure illustrated in Figure 3.4-11 vs. $(V_2 - V_1)/V_{avg}$, where $V_1$ and $V_2$ are the voltages applied to the two helices and $V_{avg} = (V_1 + V_2)/2$. (From D. J. Blattner and F. E. Vaccaro, *Electronics* 32, No. 1, 46, 1959. Copyright by *Electronics*, a McGraw-Hill Publication)

[21]Reference 3.17.

*PROBLEMS*
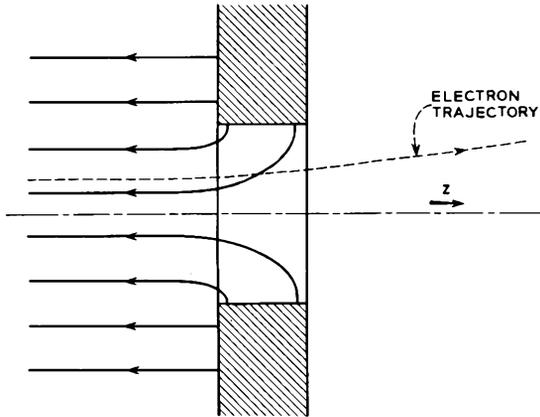


Problem 3.1

3.1 The figure shows an electrode with a small aperture in it. To the left of the electrode there is a uniform potential gradient $E_z$. An electron approaches the aperture from the left along a path parallel to the axis of the aperture but displaced a small distance from it. As the electron passes through the aperture, the diverging field deflects it away from the axis. Show that the field in the region of the aperture acts as a diverging lens with focal length approximately equal to $4V_o/E_z$, where $V_o$ is the potential through which the electron has been accelerated at the time it passes through the aperture. Assume that the electron's velocity is sufficiently large when it reaches the aperture that the electron remains at nearly constant distance from the axis as it passes through the aperture, and the effect of the radial field is to give the electron an outward impulse.

3.2 A single turn of wire which conducts a current $I_o$ generates an axially symmetric field which can be used as a magnetic lens. Using the expression given in Equation (3.2-13) for the focal length of a weak lens, show that

$$f = \frac{256V_oR}{3\pi\eta\mu_o{}^2I_o{}^2} \approx \frac{98V_oR}{I_o{}^2}$$

for such a lens, where $R$ is the radius of the turn, and $V_o$ is the beam voltage.

3.3 Sketch a magnetic lens that produces no net rotation of the beam.

3.4 Figure 3.3-3 shows a cylindrical beam of electrons that passes between two parallel deflection plates and is deflected through a mean angle $\theta$. However, because of deflection defocusing effects, electrons at the upper side of the beam are deflected through a slightly smaller angle, which we shall assume to be $\theta - \Delta\theta$, and electrons at the lower side of the beam are deflected through an angle $\theta + \Delta\theta$. Show that for a given beam diameter and given angle $\theta$, the incremental angle $\Delta\theta$ is inversely proportional to the length of the deflection plates. Assume that the field between the deflection plates is uniform, and that the effects of fringing fields

at the ends of the plates can be neglected. (Actually the fringing field at the exit end of the deflecting plates has the effect of reducing the deflection defocusing.)

3.5 Two apertured electrodes, one at higher potential than the other, form an electric lens. An electron beam passes through the lens in the direction of increasing potential. The electrode at lower potential has a wire grid across its aperture and is in contact with the grid. The wires of the grid are laid in two directions at right angles so as to produce a fine mesh. Show qualitatively that a beam passing through the lens experiences a *diverging* action. Note that this does not contradict the statements made in section 3.1 about *axially symmetric* fields acting as converging lenses.

3.6 Show that with Brillouin flow all the electrons of the beam have the same axial velocity, equivalent to that produced by an accelerating potential equal to the potential on the beam axis.

## REFERENCES

Several general references pertaining to the material covered in this chapter are:

3a. O. Klemperer, *Electron Optics*, 2nd Ed., Cambridge University Press, Cambridge, England, 1953.

3b. J. R. Pierce, *Theory and Design of Electron Beams*, 2nd Ed., D. Van Nostrand Co., Inc., Princeton, N. J., 1954.

3c. V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope*, John Wiley and Sons, Inc., New York, 1945.

3d. V. K. Zworykin and G. A. Morton, *Television*, 2nd Ed., John Wiley, and Sons, Inc., New York, 1945.

3e. K. R. Spangenberg, *Vacuum Tubes*, Chapters 13, 14, 15, McGraw-Hill Book Co., Inc., New York, 1948.

3f. P. Grivet, *Advances in Electronics*, Vol. II, p. 47, 1950.

3g. G. R. Brewer, "Some Characteristics of a Magnetically Focused Electron Beam — Parts I and II, "*Technical Memoranda No. 495* and *No. 528*, Research Laboratories, Hughes Aircraft Co., Culver City, Calif.

3h. P. A. Lindsay, "Velocity Distribution in Electron Streams," *Advances in Electronics and Electron Physics* **XIII**, 182, 1960.

Other references covering specific subjects discussed in Chapter 3 are:

3.1 S. Bertram, *Proc. IRE* **28**, 418, 1940.

3.2 L. S. Goddard, *Proc. Cambridge Phil. Soc.* **42**, 106, 1946.

3.3 K. Spangenberg and L. M. Field, *Proc. IRE* **30**, 138, 1942.

3.4 G. Liebmann, *Advances in Electronics* **II**, 102, 1950.

3.5 K. R. Spangenberg and L. M. Field, *Elec. Comm.* 21, 194, 1943.

3.6 M. E. Hines, G. W. Hoffman, and J. A. Saloom, *J. Appl. Phys.* **26**, 1157, 1955.

3.7 K. R. Spangenberg, *Vacuum Tubes*, McGraw-Hill Book Co., Inc., 1948.

3.8 L. Brillouin, *Phys. Rev.* **67**, 260, 1945.

3.9 J. P. Laico, H. L. McDowell, and C. R. Moster, *Bell System Tech. J.* **35**, 1285, 1956.

3.10 K. J. Harker, *J. Appl. Phys.* **28**, 645, 1957.

3.11 A. Ashkin, *J. Appl. Phys.* **29**, 1595, 1958.

3.12  J. R. Pierce, *J. Appl. Phys.* **24**, 1247, 1953.
3.13  J. T. Mendel, C. F. Quate, and W. H. Yocom, *Proc. IRE* **42**, 800, 1954.
3.14  N. W. McLachlan, *Theory and Applications of Mathieu Functions*, Oxford University Press, Oxford, England, 1947.
3.15  J. E. Sterrett and H. Heffner, *Trans. IRE* **ED-5**, 35, 1958.
3.16  P. K. Tien, *J. Appl. Phys.* **25**, 1281, 1954.
3.17  D. J. Blattner and F. E. Vaccaro, *Electronics* **32**, No. 1, 46, 1959.

# Chapter 4

## DIODES AND ELECTRON GUNS

The simplest vacuum tube is the two-electrode tube, or diode. In its usual form, this tube consists of a thermionic cathode and an anode. When the anode is at a positive potential with respect to the cathode, electrons emitted from the cathode are drawn to the anode, and when the anode is negative with respect to the cathode, virtually no electrons reach it. The diode therefore serves as a one-way current device, and as such it finds its chief application.

The diode geometry that lends itself most readily to analysis of the electron behavior is that in which the cathode and anode are planar, parallel, and of linear dimensions large compared with the spacing between them. We shall find that the significant laws that describe the operation of diodes with this type of geometry apply also to diodes with more complicated geometries.

If a small aperture is made in the anode of a planar diode, some of the electrons emitted from the cathode pass through the aperture into the space beyond. The device therefore acts as a crude sort of electron gun. Most electron guns use at least one additional electrode which helps to shape the field between the cathode and anode. With a suitable choice of geometry for this electrode, and with a suitable shape of cathode and anode, it is possible to cause essentially all the current drawn from the cathode to pass through the anode aperture.

Electron guns are a basic element in many types of electron tubes. Many microwave tubes make use of high-current-density, cylindrical beams of electrons. To obtain these beams, electrode geometries must be devised that accelerate the electrons to the required velocity and focus them to the required beam diameter. The electron guns used in cathode-ray tubes and storage tubes focus the beam to a "crossover," and an electron lens beyond the crossover forms an image of the crossover on the screen or storage

surface of the tube. Several designs of electron guns for microwave tubes, cathode-ray tubes, and storage tubes are described in Section 4.5.

Because the electrons emitted from the cathode of an electron gun have a finite component of velocity parallel to the cathode at the time of emission, the electrons tend to drift across the beam while being accelerated away from the cathode. If the electron beam is focused to a smaller diameter than that of the cathode, the transverse velocities of the electrons in the beam increase as the beam diameter is decreased. Ultimately, if the beam is focused to a crossover, the maximum current density that can be obtained at the crossover for a given cathode current density and beam voltage is limited by the emission velocities at the cathode. These effects are discussed in Section 4.4.

In this chapter and in subsequent chapters we shall use the subscript $o$ to designate dc electrode voltages and currents. Thus $V_{ao}$ and $I_{ao}$ are the dc anode voltage and current.

## 4.1   The Planar Diode

Here we consider a diode consisting of two planar, parallel electrodes of linear dimensions large compared with the spacing between them. We assume that the effects of fringing fields at the edges of the electrodes can be neglected, and that the fields between the electrodes are everywhere normal to the electrodes.

Figure 4.1-1 shows qualitatively the fields and potential distribution[1] in the interelectrode space of such a tube for several conditions of applied anode voltage and cathode emission current. When electrons are present in the interelectrode space, electric field lines extend from induced positive charges on the electrodes to the electrons, and *the net positive charge on the electrodes is equal to the total negative charge in the interelectrode space.* In Figure 4.1-1(a), the anode is held at cathode or ground potential while appreciable electron emission from the cathode takes place. (We assume that both electrodes have the same work function, so that the effects of contact potential difference can be neglected.) In this case, electric field lines extend from induced positive charges on both electrodes to the electrons in the interelectrode space, with the result that the potential in the space between the electrodes is less than ground potential and reaches a minimum at some

---

[1]Notice that in Figures 2.1-2 and 2.1-3 of Chapter 2 we have plotted the potential that applies to a negative unit charge, whereas in Figure 4.1-1 we plot the potential of a positive unit charge. In the field of atomic physics, convention calls for using the potentials that apply to negative charges, whereas in electron-tube work the potentials that apply to positive charges are more frequently used.
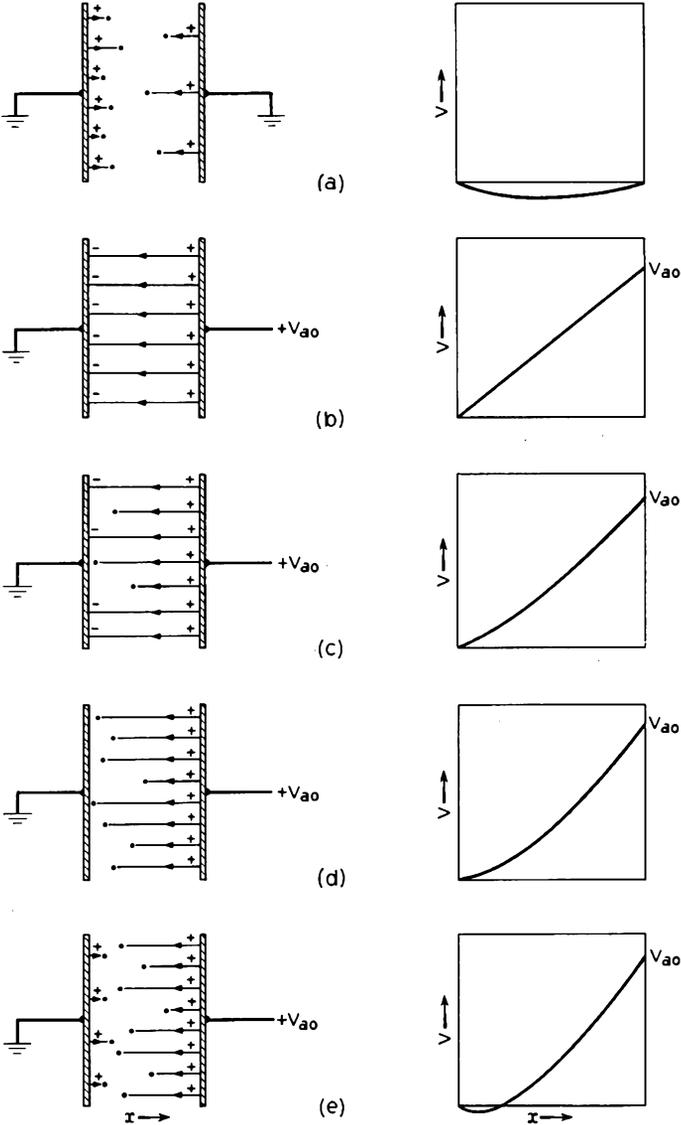
FIG. 4.1-1 The field lines and potential distribution in the interelectrode space of a planar diode for several conditions of applied anode voltage and cathode emission current.

point between the electrodes. Since the electrons emitted from the cathode have a range of velocities, the faster electrons are able to overcome the potential minimum and pass on to the anode, whereas those emitted with relatively small velocity travel only part way out to the potential minimum before being returned to the cathode.

To the right of the potential minimum, all the electrons move from left to right, whereas to the left of the potential minimum there are additional electrons that travel part way out to the potential minimum and return to the cathode. This means that the electron density to the left of the potential minimum is greater than the electron density at points of equal potential to the right of the potential minimum. Consequently, the potential gradient is greater to the left of the potential minimum, and the position of the potential minimum is displaced to the left of the mid-point between the electrodes. Later we shall find that the potential difference between the cathode and the potential minimum is directly proportional to the average electron emission energy in the direction normal to the cathode. (See Equation (4.1-1).)

In Figure 4.1-1(b), a voltage $V_{ao}$ is assumed to be applied to the anode, but no cathode emission takes place. In this case, field lines extend from positive charges on the anode to negative charges on the cathode, and the potential varies linearly from the cathode to the anode. In Figure 4.1-1(c), a small cathode emission is also assumed to take place. In this case, the emitted electrons experience a field which draws them toward the anode, so that the entire cathode emission current reaches the anode. The current drawn from the cathode is then said to be temperature-limited, since its magnitude is determined by the cathode temperature and shows little variation with changes in positive anode potential. The density of field lines leaving the anode in this case is greater than in Figure 4.1-1(b), whereas the density of field lines arriving at the cathode is less than in Figure 4.1-1(b), the same anode voltage being applied in each case. (The density of field lines at a given point is, of course, proportional to the potential gradient at that point.)

As the cathode temperature is raised so that more electrons are emitted, more field lines originating on the anode terminate on electrons in the interelectrode space, and the electric field intensity at the cathode surface decreases correspondingly. At a sufficiently high cathode temperature, the field lines extending from the anode to electrons in the interelectrode space have sufficient density to account for the potential drop $V_{ao}$, and the electric field intensity at the cathode is zero. This condition is illustrated in Figure 4.1-1(d). With still greater cathode emission (Figure 4.1-1(e)), a potential minimum forms in front of the cathode, and some of the emitted electrons are returned to the cathode.

When a potential minimum is present in front of the cathode, changes in cathode temperature serve only to raise or lower the potential at the minimum and have very little effect on the net current drawn from the cathode. The current drawn from the cathode in this case is said to be space-charge-limited and is determined largely by the voltage applied to the anode. Increasing the anode voltage requires a greater density of field lines at the anode to account for the potential difference between the anode and the potential minimum. This means that more field lines extend from the anode to the electrons in the interelectrode space, and more of the emitted electrons are drawn to the anode. Thus with increasing anode voltage, the current drawn from the cathode increases, and the potential at the minimum rises closer to cathode potential. At a sufficiently high anode potential, the potential minimum disappears, and the current drawn from the cathode becomes temperature-limited.
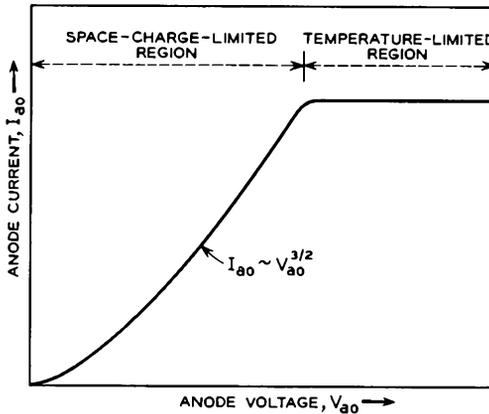


Fig. 4.1-2 An idealized plot of anode current vs. anode voltage for a planar diode. The cathode temperature is assumed to be fixed.

Figure 4.1-2 shows an idealized plot of anode current vs. anode voltage for a planar diode. The regions in which the current drawn to the anode is space-charge-limited and temperature-limited are shown in the figure. A planar diode with a tungsten cathode would exhibit a current-voltage relationship that would closely approximate this curve.

From the discussion of emission energies given in Section 2.4, it will be recalled that the fraction $F$ of the emitted electrons that can overcome a retarding voltage of $V$ volts is given by $F = \epsilon^{-\epsilon V/kT}$. Thus, when the current drawn from the cathode of a diode is space-charge-limited, the fraction $F$ of the total emission current that is drawn to the anode is given by $F =$

$\epsilon^{-eV_m/kT}$, where $-V_m$ is the potential at the potential minimum measured relative to cathode potential. Hence

$$V_m = -\frac{kT}{e}\ln F = -\frac{T}{11,600}\ln F \text{ volts} \qquad (4.1\text{-}1)$$

For $F = 1/\epsilon$ and $T = 1160°K$, $V_m = 0.1$ volt. Under most operating conditions, $V_m$ is of the order of a few tenths of a volt or less. It is noteworthy that $V_m$ is directly proportional to $kT$, the average emission energy in the direction normal to the cathode. If the emission energy were zero, the potential minimum would vanish.

The plane of the potential minimum is often called the virtual cathode, since all the electrons that pass this plane ultimately reach the anode. Let us now proceed to determine the current-voltage relationship for a space-charge-limited planar diode. We shall assume that the electrons pass the virtual cathode with zero velocity. The potential at the virtual cathode is taken to be zero, and distance $x$ is measured from the virtual cathode toward the anode. If $V(x)$ is the potential at a point $x$ meters from the virtual cathode, and $u(x)$ is the electron velocity at that point, the boundary conditions at $x = 0$ are $V = 0$, $u = 0$, and $dV/dx = 0$. The equations relating the parameters of interest are:

Poisson's Equation

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\varepsilon_o} \qquad (4.1\text{-}2)$$

the energy equation

$$\tfrac{1}{2}mu^2 = eV \qquad (4.1\text{-}3)$$

and the current density relation

$$J = -\rho u \qquad (4.1\text{-}4)$$

where $\rho$ is the volume charge density.
In these equations, $\rho$, $u$, and $V$ are assumed to be functions of $x$, whereas from the equation of continuity it follows that $J$ is independent of $x$. The charge density $\rho$ is negative, and $u$ is positive. Eliminating $\rho$ and $u$ from the above equations, we obtain

$$\frac{d^2V}{dx^2} = \frac{J}{\varepsilon_o\sqrt{2(e/m)V}} \qquad (4.1\text{-}5)$$

Next, both sides of this equation can be multiplied by $dV/dx$ and integrated with respect to $x$ from $x = 0$ to $x$, giving

$$\left|\frac{dV}{dx}\right|^2 = \frac{4JV^{1/2}}{\varepsilon_o\sqrt{2(e/m)}} + C_1 \qquad (4.1\text{-}6)$$

Since $V = dV/dx = 0$ at $x = 0$, the constant $C_1$ is zero. Taking the square root of both sides and integrating once more, we obtain

$$(4/3)V^{3/4} = 2\sqrt{J/\epsilon_o}(m/2e)^{1/4}x + C_2 \tag{4.1-7}$$

Since $V = 0$ at $x = 0$, $C_2$ is also zero. Finally, this equation can be solved for the current density $J$, giving

$$J = \frac{4}{9}\epsilon_o\sqrt{2(e/m)}\frac{V^{3/2}}{x^2} \tag{4.1-8}$$

If the experimental values of $e$, $m$, and $\epsilon_o$ are substituted in this expression, it becomes

$$J = 2.33 \times 10^{-6}\frac{V^{3/2}}{x^2} \text{ amps/meter}^2 \tag{4.1-9}$$

Here $V = V(x)$ is the potential at a point $x$ meters from the virtual cathode. If the applied anode voltage is $V_{ao}$ volts above that of the potential minimum, and if the distance from the virtual cathode to the anode is $d$ meters, the current density is given by

$$J = 2.33 \times 10^{-6}\frac{V_{ao}^{3/2}}{d^2} \text{ amps/meter}^2 \tag{4.1-10}$$

If $V_{ao} >> V_m$, the voltage $V_{ao}$ can be taken to be the anode-to-cathode voltage. Similarly, the distance from the cathode to the potential minimum is usually small compared with the distance from the potential minimum to the anode, so that $d$ can be taken to be the anode-to-cathode distance. Hence, to a good approximation, the current drawn from the cathode under space-charge-limited conditions varies as the 3/2 power of the anode voltage divided by the square of the anode-to-cathode distance. This result is known as the Child-Langmuir Law. Langmuir[2] also developed more accurate equations for the planar diode which take into account the distribution of electron emission velocities and which show the location of the potential minimum. However, for most purposes, Equation (4.1-10) gives a sufficiently accurate expression for the current density $J$, and the potential minimum can be assumed to be very close to the cathode.

## 4.2   Diodes with Other Electrode Geometries

Two further conclusions concerning Equations (4.1-9) and (4.1-10) are of interest:

1. If $A$ is the cathode area of a planar diode, the current drawn to the anode under space-charge-limited conditions is given by $JA = 2.33 \times 10^{-6} V_{ao}^{3/2}A/d^2$. Because of the factor $A/d^2$ in this expression, it is evident that,

---

[2]Reference 4.1.

if the linear dimensions of a planar diode are increased by a factor $k$, the same current flows to the anode for the same applied voltage.

2. Combining Equations (4.1-9) and (4.1-10), we obtain

$$V(x) = \frac{x^{4/3}}{d^{4/3}} V_{ao} \qquad (4.2\text{-}1)$$

Hence, despite the fact that increasing the anode voltage increases the current drawn from the cathode, the potential at points between the electrodes remains directly proportional to the applied anode voltage.

Equations describing the space-charge-limited flow of electrons between concentric cylinders and concentric spheres have also been derived.[3] In each of these cases, it is found that: (1) The current drawn to the anode is proportional to the 3/2 power of the applied anode voltage; (2) if two diodes differ by a factor $k$ in their linear dimensions, the same current flows to the anode when the same anode voltage is applied; and (3) the potential
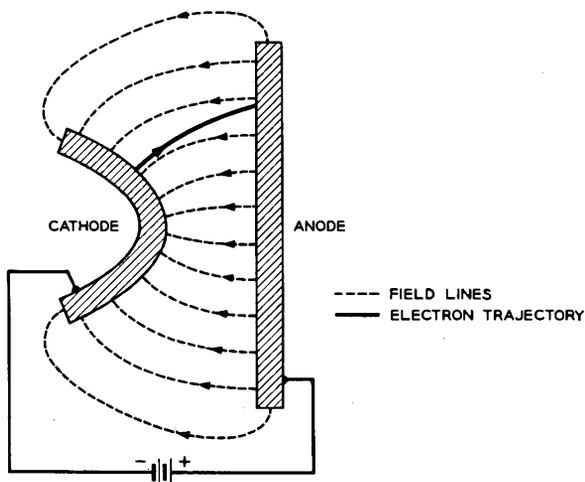


FIG. 4.2-1 A diode with a parabolic-shaped cathode and a planar anode.

at points between the electrodes is directly proportional to the applied anode voltage when space-charge-limited conditions apply.

In the planar, cylindrical, and spherical diodes, the field lines are straight, and the electron trajectories are parallel to the field lines. However, in diodes with other electrode geometries, this is not the case; when the field lines are curved, the electron trajectories cross the field lines. This effect is

---

[3]Reference 4.1, p. 245, and References 4.2 and 4.3.

illustrated in Figure 4.2-1 for a diode consisting of a parabolic-shaped
cathode and a planar anode. The trajectory of a single electron is shown
in the figure. Because the electron has inertia, the path it follows does not
bend as sharply as the field lines.

The question then arises as to whether in a space-charge-limited diode
with arbitrarily shaped electrodes the electron trajectories change their
shape when the anode voltage is varied. To answer this question, suppose
that $V = V_1(x,y,z)$ and $\rho = \rho_1(x,y,z)$ are steady-state solutions for the
potential and charge in the interelectrode space of a particular diode. These
solutions meet the following boundary conditions: $V = V_{ao}$ at the anode
surface, $V = dV/dn = 0$ at the potential minimum (which we shall
assume coincides with the cathode surface). Here $d/dn$ is the derivative
in the direction normal to the cathode surface. From Poisson's equation
it follows that, if the anode voltage is now changed to $kV_{ao}$, solutions
of the potential and charge distribution which meet the new boundary
conditions are given by $V = kV_1(x,y,z)$ and $\rho = k\rho_1(x,y,z)$. Furthermore,
it is shown in Appendix VII that only one steady-state solution of Poisson's
Equation will meet the boundary conditions for a space-charge-limited
diode. It follows, therefore, that the potential in the interelectrode space
of a space-charge-limited diode with arbitrarily shaped electrodes is directly
proportional to the applied anode voltage. This being the case, we can
invoke the same arguments that were used in Section 1.1 to show that the
electron trajectories are not affected by changes in positive anode voltage.
This conclusion, in fact, is experimentally verified, apart from effects
arising from the finite emission velocity of the electrons.

Next let us consider how the current density $J = -\rho u$ in the inter-
electrode space of a diode with arbitrarily shaped electrodes varies with the
applied anode voltage when space-charge-limited conditions prevail.
From the relationship,

$$\tfrac{1}{2}mu^2 = eV(x,y,z) \tag{4.2-2}$$

we see that $u = u(x,y,z)$ is proportional to the square root of $V(x,y,z)$, and
hence it is proportional to the square root of the applied anode voltage.
The charge density $\rho = \rho(x,y,z)$ is related to the potential $V(x,y,z)$ by

$$\nabla^2 V(x,y,z) = -\frac{\rho}{\varepsilon_o} \tag{4.2-3}$$

Since $V(x,y,z)$ is proportional to the applied anode voltage, and since the
Laplacian operator is linear, it follows that $\rho$ is directly proportional to the
applied anode voltage. Consequently, $J = -\rho u$ is proportional to the 3/2
power of the applied anode voltage. Thus the current drawn to the anode

of a space-charge-limited diode having arbitrarily shaped electrodes is proportional to the 3/2 power of the applied anode voltage.

By similar reasoning we can deduce that, if the dimensions of a diode with arbitrarily shaped electrodes are scaled by a constant factor and if the tube is operated under space-charge-limited conditions, the total current drawn to the anode for a given applied anode voltage is unchanged. Let us suppose that the linear dimensions of a diode are increased by the factor $k$. The potential at corresponding points between the electrodes remains the same for the same applied anode voltage, so that the electron velocity $u$ at corresponding points between the electrodes remains the same. However, $\partial V/\partial x$ is changed by $1/k$, and $\partial^2 V/\partial x^2$ is changed by $1/k^2$. Since $\rho$ is proportional to $\nabla^2 V(x,y,z)$, it follows that $\rho$ is changed by $1/k^2$. Consequently, the current density $J = -\rho u$ is changed by $1/k^2$; and since the electrode area is $k^2$ times its previous value, the current drawn to the anode for the same applied anode voltage remains unchanged.

## 4.3   Two Examples of Diode Rectifiers

### The 412A

Figure 4.3-1 shows the construction of the Western Electric 412A full-wave diode rectifier. The tube consists of two diodes with indirectly



SECTION A–A

HEATER

CERAMIC
INSULATOR

ANODE

CATHODE

ENLARGED VIEW
OF DIODE
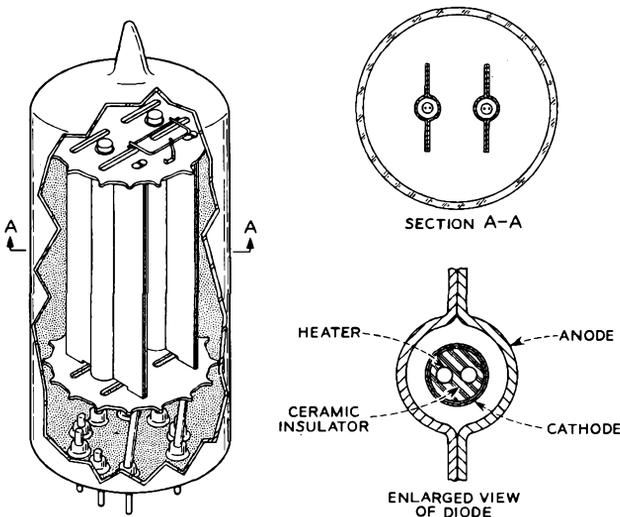
FIG. 4.3-1 The construction of the Western Electric 412A full-wave diode rectifier. The overall height of the tube is 6.7 cm.

heated cathodes enclosed in a common envelope. A ceramic insulator sep-
arates the heater of each diode from the cathode. The cathodes are cy-
lindrical sleeves of nickel with a "double-carbonate" oxide coating on the
outer side. The cathode-anode spacing is 0.5 mm. The anodes are made of
nickel which is coated with fine carbon particles in order to increase the heat
radiation from the outer surface. This in turn enables the anodes to operate
at a lower temperature for a given power dissipation.

Figure 4.3-2 shows measurements of anode current vs. anode voltage for
the 412A for several heater voltages. The normal heater operating voltage



FIG. 4.3-2 Anode current vs. anode voltage for the 412A for several heater operat-
ing voltages $V_h$. The normal heater operating voltage is 6.3 volts.

is 6.3 volts. At $V_h = 6.3$ volts, the anode current is space-charge-limited
over the range of anode voltages for which the data are plotted, and the
anode current increases very nearly as the 3/2 power of the anode voltage.
At the other filament voltages for which data are plotted in the figure, the
anode current is space-charge-limited at lower anode voltages and "tem-
perature-limited" at higher anode voltages. In the region of "temperature-
limited" operation, the anode current actually increases with increasing
anode voltage rather than remaining constant, as it would in an ideally
temperature-limited diode. It is thought that this can be attributed to

the rough and porous nature of the oxide-cathode emitting surface. At the onset of temperature-limited operation only the current drawn from the outermost parts of the cathode surface is temperature-limited, whereas the current drawn from the re-entrant parts and the entrances to the pores is still space-charge-limited. Thus, with increasing anode voltage, the current drawn from regions that are still space-charge-limited continues to increase, but the total area from which space-charge-limited current is drawn decreases. Operation of the tube for an extended time in the temperature-limited region is found to be harmful to cathode life.
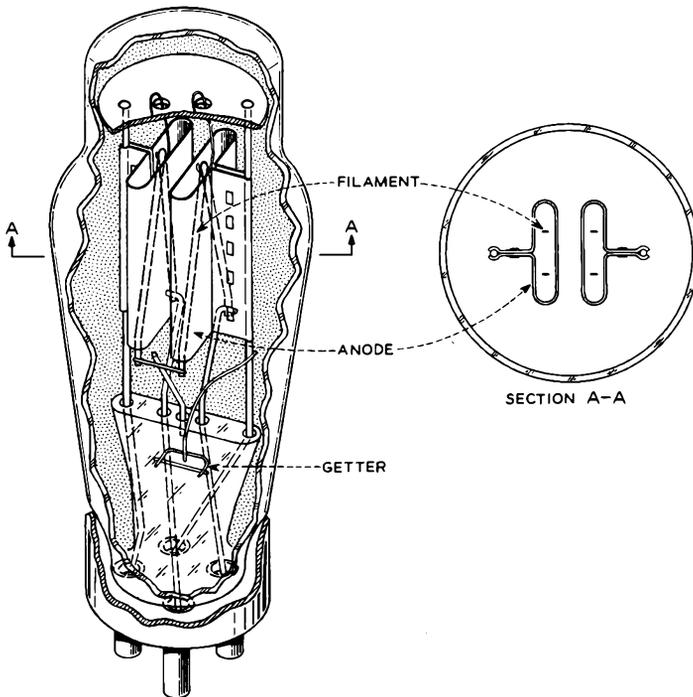


FIG. 4.3-3 The construction of the Western Electric 274B full-wave diode rectifier. The over-all height of the tube is 13.8 cm.

Maximum ratings for the 412A are given in Table 4.3-1. The use of a ceramic insulator between the heater and cathode permits operation of the diodes with as much as 450 volts potential difference between the heater and cathode.

<div align="center">Table 4.3-1</div>

<div align="right">*Maximum Rating*</div>

| | |
|---|---|
| Peak inverse voltage*, volts | 1250 |
| Peak anode current per anode, ma | 300 |
| DC output current (when operated as a full-wave rectifier), ma | 100 |
| DC heater-cathode potential, volts | 450 |

*Maximum negative voltage applied to the anode with respect to cathode voltage.

## The 274B

The construction of the Western Electric 274B full-wave diode rectifier is illustrated in Figure 4.3-3. The tube consists of two diodes with filamentary cathodes enclosed in a common envelope. The filaments are made from a nickel alloy and have a "double-carbonate" oxide coating. The nickel alloy contains the following elements in addition to nickel:

| Element | Per Cent by Weight | Element | Per Cent by Weight |
|---|---|---|---|
| Co | 0.5 to 0.75 | C | 0.04 to 0.07 |
| Cu | < 0.10 | Si | < 0.03 |
| Fe | < 0.15 | Mg | 0.04 to 0.08 |
| Mn | < 0.20 | Ti | < 0.03 |

These small amounts of impurities in the nickel increase its resistivity and mechanical strength. The elements in the right-hand half of the table
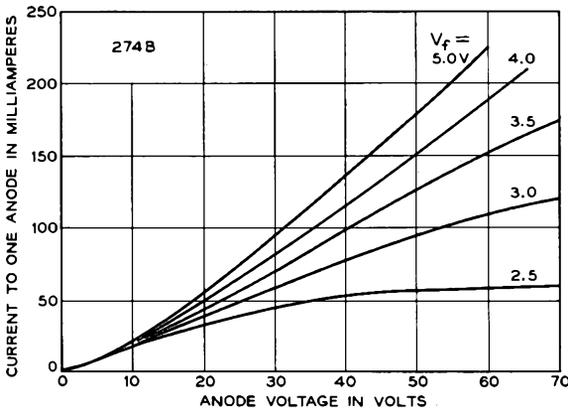


FIG. 4.3-4 Anode current vs. anode voltage for the 274B for several filament operating voltages $V_f$. The normal filament operating voltage is 5.0 volts.

also serve as the reducing agents which maintain the activity of the oxide coating. The filaments are connected in series within the envelope of the tube. The anodes are made of carbonized nickel, as in the 412A.

Figure 4.3-4 shows measurements of the anode current vs. anode voltage for the 274B. The normal filament operating voltage is 5.0 volts. The shape of the curves in the "temperature-limited" region is further complicated for this tube by the facts that the distance from the filament surface to the anode varies over the filament surface and that there is a voltage drop along the length of the filament. Consequently, the onset of temperature-limited operation occurs at different anode voltages for different parts of the filament surface.

Maximum ratings for the tube are given in Table 4.3-2.

<div align="center">TABLE  4.3-2</div>

<div align="right">*Maximum Rating*</div>

| | |
|---|---|
| Peak inverse voltage, volts................................. | 1500 |
| Peak anode current per anode, ma......................... | 675 |
| DC output current (when operated as a full-wave rectifier), ma... | 225 |

## 4.4  Some Effects of Thermal Emission Velocities

Appendix IV summarizes the relations pertaining to the velocity distribution, energy distribution, and angular distribution of the electrons emitted from a thermionic cathode, as discussed in Section 2.4.

Because the electrons are emitted with a finite component of velocity parallel to the cathode surface, they tend to drift across the beam while being drawn away from the cathode by the applied field. In consequence of this, the electron beams generated by electron guns are always larger than they would be if the electron emission velocity were zero.

As a simple example to illustrate the sideways drift of electrons in an accelerating field, let us consider the electron trajectories in the planar, parallel diode shown in Figure 4.4-1. (In the illustrations used in this section, it will be convenient to identify the trajectories of electrons emitted from the cathode with zero kinetic energy with solid lines, and the trajectories of electrons having finite kinetic energy at the time of emission with broken lines. The former electrons will be called *nonthermal* electrons, and the latter will be called *thermal* electrons.) Suppose the anode-to-cathode spacing of the diode shown in Figure 4.4-1 is 1 cm, and a voltage of $+10$ volts is applied to the anode. We shall assume that the cathode emission is very small and temperature-limited and that the electric field between the electrodes is uniform. Consider an electron which is emitted from the cathode with 1/10 electron volt of kinetic energy parallel to the
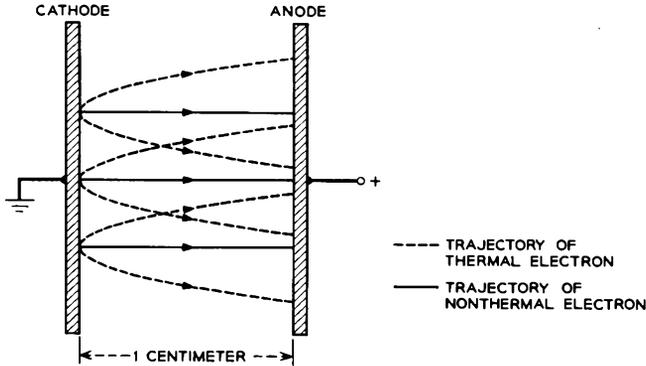
FIG. 4.4-1 Some electron trajectories in the interelectrode space of a planar diode.

cathode surface and no kinetic energy normal to the surface. At the time the electron strikes the anode, the ratio of its transverse energy to its energy in the direction of the accelerating field is 1/100. The corresponding ratio of velocities is equal to the square root of this, or 1/10. Since the final velocity in the direction of the accelerating field is twice the average velocity, the ratio of the transverse velocity of the electron to its average velocity in the direction of the field is 2/10. Consequently, the electron does not strike the anode at a point directly opposite its point of emission, but 2 millimeters to one side.

If the cathode temperature in the above example were 1160°K, the electron we have considered would have had average transverse energy, since $W_T = T/11{,}600$ electron volts is the average transverse energy. Other electrons would be emitted with appreciably greater transverse energies. Furthermore, if the current drawn from the cathode were space-charge-limited, the time taken by an electron to reach the anode would be at least 3/2 as great (Problem 4.1), so that the electrons would drift even farther to the side. Of course, by increasing the accelerating voltage, the amount of sideways drift is reduced. If the anode voltage in the above example were increased to 1000 volts, the electron would drift only 0.2 mm to the side for the case of the uniform accelerating field.

Let us now consider the distribution of points of arrival on the anode of electrons emitted from a single point on the cathode of a planar diode. Suppose an electron emitted from a point on the cathode with transverse velocity $\sqrt{kT/m}$ drifts a distance $\sigma$ to the side in traveling from the cathode to the anode. If the anode voltage is large compared with $kT/e$, we can neglect the effects of emission velocities normal to the cathode (to a first approximation) and assume a constant time for electrons to travel from the cathode to the anode. In this case, an electron emitted from the cathode

with transverse velocity $u_t$ would drift a distance $r$ to the side while traveling from the cathode to the anode, where

$$r = \frac{u_t}{\sqrt{kT/m}}\sigma \qquad (4.4\text{-}1)$$

Rearranging this equation, we obtain

$$u_t = \frac{r}{\sigma}\sqrt{kT/m} \qquad (4.4\text{-}2)$$

Substituting for $u_t$ in Equation (2) of Appendix IV from the above equation, we find that the probability that an electron, which is incident upon the anode, is displaced a distance in the range $r$ to $r + dr$ from the point on the anode directly opposite its point of emission is given by

$$dP(r) = \frac{1}{2\sigma^2}\epsilon^{-r^2/2\sigma^2}dr^2 \qquad (4.4\text{-}3)$$

Equation (4.4-3) implies a Gaussian distribution in two dimensions, and $\sigma$ can be identified as the standard deviation of the distribution. Thus electrons emitted from a single point on the cathode will be incident upon the anode at points whose density is given by a two-dimensional Gaussian function with standard deviation $\sigma$.

An electron emitted from the cathode of a planar diode with a component of velocity parallel to the cathode follows a curved trajectory which bends increasingly toward the normal to the electrodes. For this reason, the current density of electrons arriving at the anode per unit solid angle in the direction normal to the anode is far higher than the cathode emission current density per unit solid angle in the direction normal to the cathode. From Equation (7) of Appendix IV it follows that the latter quantity is $J_o/\pi$, where $J_o$ is the total cathode emission current density. Let us proceed now to obtain the current density of electrons arriving at the anode per unit solid angle in the direction normal to the anode. We shall use this quantity in later discussion.

We shall assume, as before, that fringing fields at the edge of the diode can be neglected, and that an electron emitted from the cathode with a component of velocity in the direction parallel to the cathode will maintain this component of velocity throughout its travel from the cathode to the anode. In this case, an electron which is emitted from the cathode in a direction making an angle $\theta_1$ with the normal to the cathode, and which has kinetic energy $eV_1$ joules at the time of emission, will be incident upon the anode at an angle $\theta_2$ with respect to the normal such that

$$\sin\theta_2 = \frac{V_1^{1/2}}{(V_1 + V_2)^{1/2}}\sin\theta_1 \qquad (4.4\text{-}4)$$

where $V_2$ is the cathode-anode potential difference. Differentiating both sides of this gives

$$\cos\theta_2 \, d\theta_2 = \frac{V_1^{1/2}}{(V_1 + V_2)^{1/2}} \cos\theta_1 \, d\theta_1 \qquad (4.4\text{-}5)$$

From Equations (5) and (6) of Appendix IV, the current density of electrons emitted from the cathode with kinetic energy between $eV_1$ and $e(V_1 + dV_1)$ joules and with emission velocities lying in the angular range $\theta_1$ to $\theta_1 + d\theta_1$ with respect to the normal is

$$dJ(V_1, \theta_1) = J_o \frac{eV_1}{kT} \epsilon^{-eV_1/kT} \frac{de V_1}{kT} 2 \sin\theta_1 \cos\theta_1 \, d\theta_1 \qquad (4.4\text{-}6)$$

where $J_o$ is the total cathode emission current density. This is also the current density at the anode due to electrons which are emitted from the cathode with kinetic energy in the range $eV_1$ to $e(V_1 + dV_1)$ joules, and which arrive at the anode with angles of incidence in the range $\theta_2$ to $\theta_2 + d\theta_2$, where $\theta_2$ is related to $\theta_1$ by Equation (4.4-4). Thus the current density arriving at the anode per unit solid angle at an angle $\theta_2$ with respect to the normal, and which is composed of electrons emitted from the cathode with kinetic energy in the range $eV_1$ to $e(V_1 + dV_1)$ joules, is

$$\frac{dJ(V_1, \theta_1)}{2\pi \sin\theta_2 \, d\theta_2} = \frac{J_o}{\pi} \frac{e(V_1 + V_2)}{kT} \epsilon^{-eV_1/kT} \frac{de V_1}{kT} \cos\theta_2 \qquad (4.4\text{-}7)$$

where we have substituted from Equations (4.4-4) and (4.4-5) for $\sin\theta_1$ and $\cos\theta_1 \, d\theta_1$. By setting $\cos\theta_2 = 1$ in the right-hand side of Equation (4.4-7), we obtain the current density which arrives at the anode per unit solid angle in the direction *normal* to the anode and which is composed of electrons emitted from the cathode with kinetic energy in the range $eV_1$ to $e(V_1 + dV_1)$. Then, by integrating this quantity with respect to $V_1$ from zero to infinity, the total current density incident upon the anode per unit solid angle in the direction normal to the anode is found to be

$$J_a(\theta_2 = 0) = \frac{J_o}{\pi}\left(\frac{eV_2}{kT} + 1\right) \qquad (4.4\text{-}8)$$

This is the expression we set out to derive. It compares with a cathode emission current density per unit solid angle in the direction normal to the cathode of $J_o/\pi$. In the discussion that follows we shall use Equation (4.4-8) to obtain an approximate expression for the maximum current density which can be obtained at a crossover.

Figure 4.4-2 shows a planar diode in which the anode has a small circular aperture. Two additional electrodes located behind the anode combine with the anode to form an einzel lens. (See Figures 3.1-1(c) and 3.1-1(d).) The beam of electrons passing through the lens is focused to a crossover $L$
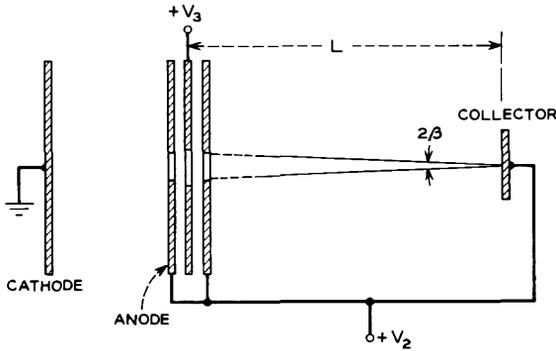
Fig. 4.4-2 A planar diode with a small circular aperture in the anode. Two aper-
tured electrodes behind the anode combine with the anode to form an einzel lens.
The electrons passing through the anode aperture are focused to a crossover at
the collector electrode.

units from the lens. A collector electrode at the crossover intercepts the
beam. Let us now estimate the current density at the center of the cross-
over. We shall assume that Equation (4.4-8) gives the current density
incident upon the anode aperture per unit solid angle in the direction normal
to the aperture. We shall further assume that electrons emitted from the
cathode with zero transverse velocity and incident upon the anode aperture
from the cathode side are deflected by the lens in the direction of the center
of the crossover. An element of area $dA$ in the center of the crossover and
normal to the beam axis will subtend a solid angle $dA/L^2$ at a point on the
aperture. Therefore, unit area of the aperture will transmit a current
$\dfrac{J_o}{\pi}\left(\dfrac{eV_2}{kT} + 1\right)\dfrac{dA}{L^2}$ to the element of area $dA$, and the current density incident
upon $dA$ from unit area of the aperture will be $\dfrac{J_o}{\pi}\left(\dfrac{eV_2}{kT} + 1\right)\dfrac{1}{L^2}$. If the radius
of the aperture is $R$, the total current density at the center of the crossover
will be

$$J_o\left(\frac{eV_2}{kT} + 1\right)\frac{R^2}{L^2} \tag{4.4-9}$$

More accurately, it can be shown with the aid of statistical mechanics[4]
that the maximum current density obtainable at a crossover with *any* lens
system is given by

$$J_o\left(\frac{eV_2}{kT} + 1\right)\sin^2\beta \tag{4.4-10}$$

_____

[4]Reference 4.4. See also Reference 4.5.

where $\beta$ is the half angle subtended at the crossover by the aperture diameter. Moreover, it can be shown that this limiting current density can be approached only when the aperture of the lens system passes a small part of the total current drawn from the cathode, and when the lens system is essentially aberration-free. The limiting current density given by the above expression applies both at a crossover and at an image of a crossover formed by a subsequent lens system.

In a cathode-ray tube the half angle $\beta$ of the cone of trajectories incident upon the screen is often of the order of $1/100$ radian or smaller, whereas $eV_2$ may be $5 \times 10^4$ times $kT$. Thus the maximum current density obtainable at the screen of the tube is often of the same order of magnitude as the cathode current density $J_o$. However, because of aberrations, the actual current density at the screen is usually less than $J_o$.

Suppose that in the device shown in Figure 4.4-2 an electron emitted from the cathode with transverse velocity $\sqrt{kT/m}$ passes through the anode aperture and strikes the collector at a point $\sigma$ units from the center of the crossover. If we assume that the transit time from the cathode to the collector is the same for all electrons reaching the collector and that aberrations in the einzel lens are small, an electron that leaves the cathode with transverse velocity $u_t$ will be displaced a distance $r = u_t\sigma/\sqrt{kT/m}$ from the beam axis by the time it reaches the crossover. We then can use the same arguments that were presented in connection with Equation (4.4-3) to show that the current density incident upon the collector is proportional to

$$\epsilon^{-r^2/2\sigma^2} d\frac{r^2}{2\sigma^2}$$

where $r$ is the distance measured along the surface of the collector from the beam axis.

If the lens in Figure 4.4-2 is made stronger, $L$ decreases and $\beta$ increases. However, $\sigma$ decreases, since the transit time from the aperture to the crossover is smaller. Similarly, decreasing the strength of the lens reduces $\beta$ and increases $\sigma$. (Ultimately the beam will diverge at the lens.) Thus an electron beam can be focused to a crossover of small diameter and large angle of convergence $\beta$, or a large diameter and small angle of convergence, but not simultaneously to a small diameter and small angle of convergence.

Next let us consider some effects of thermal emission velocities in convergent beams such as are used in many microwave tubes. Figure 4.4-3 shows a diode with a cathode emitting surface and anode which are portions of spheres, both concentric about the point $P$. Nonthermal electrons in such a diode travel in radial lines from the cathode to the anode, since the forces acting on them are directed toward the point $P$. The trajectory of one thermal electron is shown in the figure. Suppose this electron is emitted
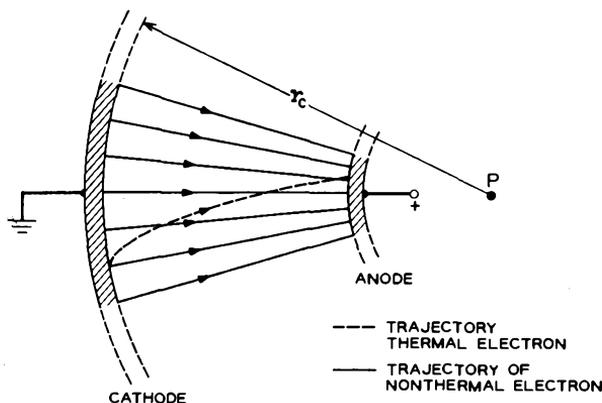
Fɪɢ. 4.4-3 Trajectories of a thermal electron and several nonthermal electrons in a spherical diode.

with velocity $u_{tc}$ parallel to the cathode surface. Since the forces acting on it are directed toward the point $P$, angular momentum about the point $P$ is conserved. When the electron reaches radius $\bar{r}$ from the point $P$, its component of velocity $u_{tr}$ *transverse to the radial direction* is given by

$$mu_{tr}\bar{r} = mu_{tc}\bar{r}_c \qquad (4.4\text{-}11)$$

or

$$u_{tr} = \frac{\bar{r}_c}{\bar{r}}u_{tc} \approx \frac{d_c}{d}u_{tc} \qquad (4.4\text{-}12)$$

where $\bar{r}_c$ is the radius of the cathode emitting surface measured from the point $P$ to the cathode emitting surface, $d_c$ is the cathode diameter, and $d$ is the diameter of the beam at radius $\bar{r}$ from point $P$. Consequently, as the thermal electron travels from the cathode toward the anode, its component of velocity *transverse to the nonthermal electron trajectories* increases, and at a given point it is inversely proportional to the beam diameter at that point. This result applies when the current drawn from the cathode is space-charge-limited as well as when it is temperature-limited.

The foregoing is a particular example of a quite general relationship which applies to paraxial electron beams. This relationship states that, if the diameter of an electron beam is reduced from $d_1$ to $d_2$ by the action of axially symmetric fields and if the fields acting on the electrons are directly proportional to the distance from the beam axis, the transverse velocities of the thermal electrons measured relative to the trajectories of the nonthermal electrons are increased by the ratio $d_1/d_2$. The discussion presented below will develop this more general relationship.

It will be convenient to introduce a radial coordinate $\mu$ which varies linearly from zero on the beam axis to unity at the beam edge. Let $r_e(z)$ be the beam radius (corresponding to $\mu = 1$). We shall assume that the non-thermal electrons travel in laminar paths, that is, paths that do not cross one another, and that the radial forces within the beam are directly proportional to distance from the axis. The path traveled by a nonthermal electron is therefore one of constant $\mu$. If there is appreciable space charge in the beam, the assumption that the radial forces are proportional to distance from the axis implies a uniform current density over the beam cross section. (See Equation (3.4-9).)

Consider a thermal electron whose radial coordinate is given by

$$r = \mu r_e \qquad (4.4\text{-}13)$$

where both $\mu$ and $r_e$ are functions of $z$. Differentiating this equation twice with respect to time gives

$$\ddot{r} = \ddot{\mu}r_e + 2\dot{\mu}\dot{r}_e + \mu\ddot{r}_e \qquad (4.4\text{-}14)$$

where

$$\dot{r}_e = r_e'\dot{z} \qquad (4.4\text{-}15)$$

and

$$\ddot{r}_e = r_e''(\dot{z})^2 + r_e'\ddot{z} \qquad (4.4\text{-}16)$$

Similar expressions hold for $\dot{\mu}$ and $\ddot{\mu}$. (The notation used here is similar to that of Chapter 3.) Since the radial forces acting on the thermal electron are directly proportional to its distance from the axis, the radial acceleration of the thermal electron is $\mu$ times the radial acceleration of a nonthermal electron at the edge of the beam, or

$$\ddot{r} = \mu\ddot{r}_e \qquad (4.4\text{-}17)$$

Comparing this equation with Equation (4.4-14), we see that

$$\ddot{\mu}r_e + 2\dot{\mu}\dot{r}_e = 0 \qquad (4.4\text{-}18)$$

or

$$\frac{d}{dt}(\dot{\mu}r_e^2) = 0 \qquad (4.4\text{-}19)$$

Integrating this, we find that

$$\dot{\mu}r_e^2 = \text{constant} \qquad (4.4\text{-}20)$$

a relationship which applies over the whole length of the beam. Consider two points on the electron's trajectory such that $\mu = \mu_1$ at one point and

$\mu = \mu_2$ at the second point. Let the beam radius be $r_{e1}$ at the first point and $r_{e2}$ at the second point. Then, from Equation (4.4-20),

$$\frac{\mu_2 r_{e2}}{\mu_1 r_{e1}} = \frac{r_{e1}}{r_{e2}} \tag{4.4-21}$$

Now the quantity $\mu_1 r_{e1}$ is the transverse velocity of the thermal electron measured relative to the trajectories of the nonthermal electrons at the point where the beam radius is $r_{e1}$. We shall denote it by $u_{t1}$. Similarly we shall set $u_{t2} = \mu_2 r_{e2}$. Equation (4.4-21) therefore can be written as

$$u_{t2} = \frac{r_{e1}}{r_{e2}} u_{t1} = \frac{d_1}{d_2} u_{t1} \tag{4.4-22}$$

where $d_1 = 2r_{e1}$, and $d_2 = 2r_{e2}$. This is the relation we set out to obtain.

The case we have just considered applied to a thermal electron whose initial transverse velocity is in the *radial* direction only. If the thermal electron also has an initial component of velocity in the $\theta$ direction, that is, the direction both perpendicular to the beam axis and the radial direction, we can use a Cartesian coordinate system in the transverse plane to describe the transverse motion of the electron. For an axially symmetric electric field in which $E_r = ar$, the $x$ and $y$ components of the radial field can be expressed as $E_x = ax$ and $E_y = ay$. Furthermore we can rewrite the foregoing equations replacing $r$ by $x$ or $y$ and $\mu$ by $\mu_x$ or $\mu_y$. In this way it is easily shown that Equation (4.4-22) applies for any direction of the initial transverse velocity $u_{t1}$.
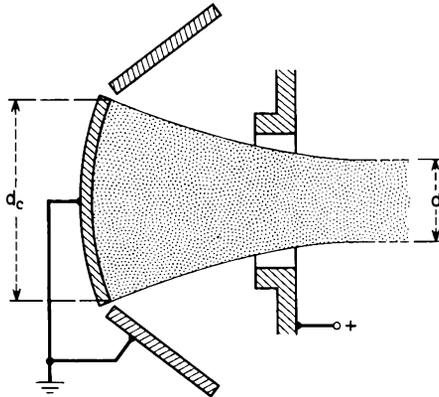


Fig. 4.4-4 A convergent electron gun which generates a beam of diameter $d$ from a cathode of diameter $d_c$.

The foregoing results can be applied to an electron gun which generates a beam of diameter $d$ and has a cathode of diameter $d_c$. Such a gun is illustrated in Figure 4.4-4. The average transverse velocity of the electrons in the region where the beam diameter is $d$ will be $d_c/d$ times the average transverse velocity at the cathode surface, and from Equation (2) of Appendix IV the probability that an individual electron has transverse velocity in the range $u_t$ to $u_t + du_t$ at a point where the beam diameter is $d$ will be

$$dP(u_t) = \frac{mu_t}{kT}\left(\frac{d}{d_c}\right)^2 \epsilon^{-mu_t^2(d/d_c)^2/2kT} du_t \qquad (4.4\text{-}23)$$

Highly convergent electron guns (guns with high $d_c/d$) therefore generate electron beams with high transverse velocities.

## 4.5  Electron Guns

It will be convenient to consider separately the electron guns used in microwave tubes and those used in cathode-ray tubes and storage tubes, since the principles involved in the two cases are quite different.

### a.  Electron Guns Used in Microwave Tubes

Figure 4.5-1 shows three electron guns that are used in beam-type microwave tubes. The first two have a relatively large cathode area in order to draw the required total emission current, and electrodes in front of the cathode focus the beam to a cross section much smaller than the cathode area. In this way electron beams of current density far greater than the cathode emission density can be obtained. The third electron gun is operated in a uniform axial magnetic field, so that the electron beam cross section is just a little larger than the cathode area, and the electron motion

TABLE  4.5-1

| Gun | a | b | c |
|---|---|---|---|
| Beam voltage, volts................... | 2600 | 500 | 570 |
| Beam current, amp................... | 0.040 | 0.066 | 0.0005 |
| Perveance, amp/volts$^{3/2}$............... | $0.30 \times 10^{-6}$ | $5.9 \times 10^{-6}$ | $3.7 \times 10^{-8}$ |
| Cathode current density, amp/cm$^2$....... | 0.21 | 0.19 | 0.16 |
| Average beam current density, amp/cm$^2$.. | 5 | 7 | $\sim 0.1$ |
| Angle of convergence*, degrees.......... | 25 | 145 | — |

*The angle subtended by a diameter of the cathode at the center of curvature of the cathode emitting surface.
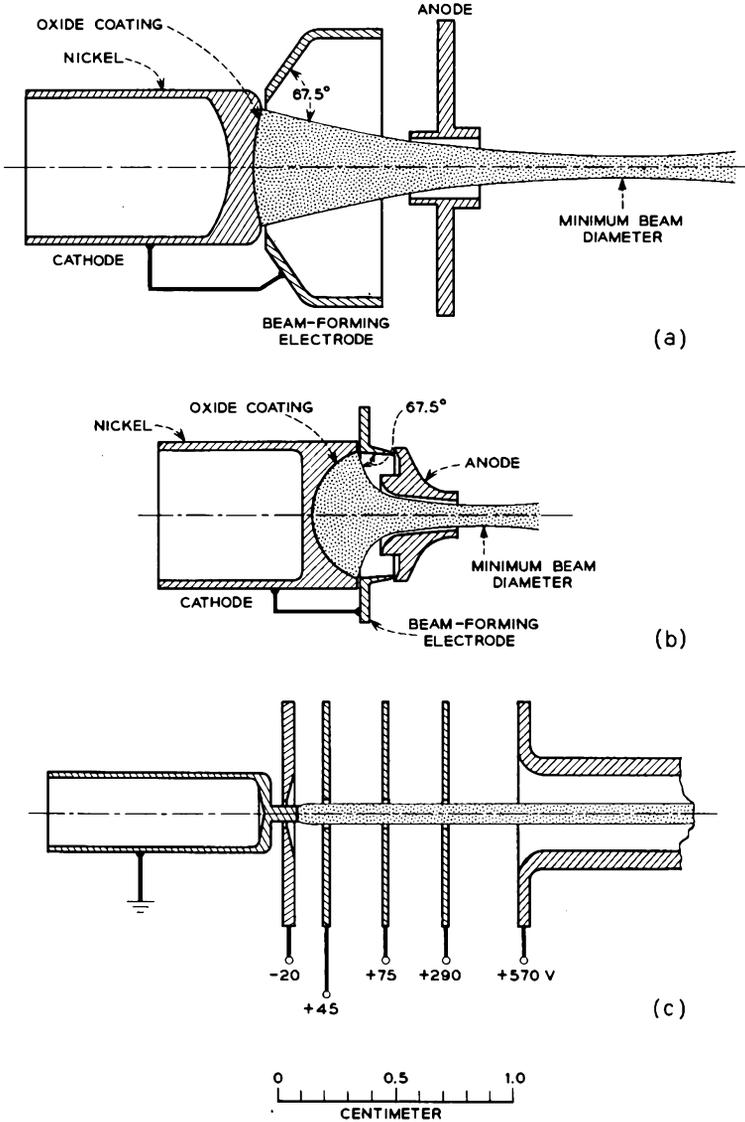
FIG. 4.5-1 Three electron guns used in microwave tubes.

is similar to that described in Section 3.4(b) under the heading of Confined Flow. Table 4.5-1 lists several important characteristics of these guns. The electron gun shown in Figure 4.5-1(a) consists of a cathode, a "beam-

forming electrode," which is operated at cathode potential, and an anode. The cathode emitting surface is concave and spherical in shape. The gun is called a Pierce electron gun, after J. R. Pierce,[5] who first put the design of convergent electron guns on a firm basis.
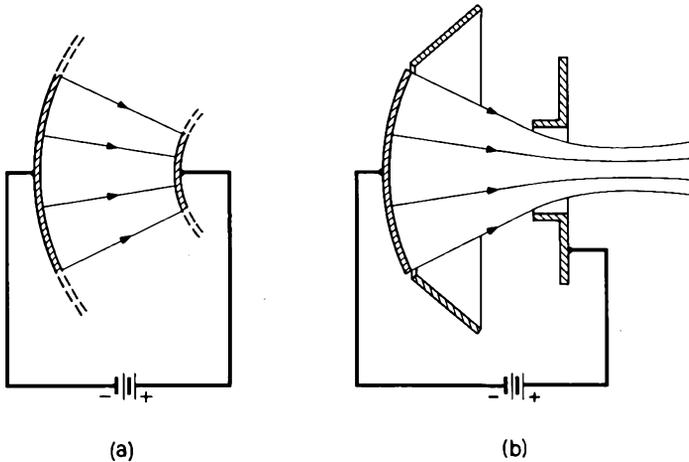


FIG. 4.5-2 The evolution of a Pierce electron gun from a spherical diode.

To understand the choice of shape for the electrodes, we might first imagine a diode consisting of portions of two concentric spheres, such as the one illustrated in Figure 4.5-2(a). The cathode and anode are assumed to be defined by the intersection between the spherical surfaces and a right-circular cone with apex at the common center of the spheres. With such a device we can make a convergent beam of electrons. However, the nonthermal edge electrons travel in radial lines only if they experience a radial electric field and no transverse field. A little consideration shows that this will be the case only if the potential just outside the beam varies with radius $\bar{r}$ (measured from the common center of the spheres) in the same way that it does inside the beam. The beam-forming electrode, therefore, is designed to create a potential along the edge of the beam which matches as nearly as possible that inside the beam. Finally, since we do not want to intercept the electrons, a hole must be made in the anode so that the convergent beam will pass on through. The resulting shape of the electron gun is similar to that shown in Figure 4.5-2(b).

The beam-forming electrode has its principal effect close to the cathode, where the electrons are moving more slowly and the transverse fields are

[5]Reference 4.6.

able to deflect the electron trajectories much more. Let us examine the shape of the beam-forming electrode in this region more closely. Figure 4.5-3 shows a much expanded view of the edge of the beam close to the
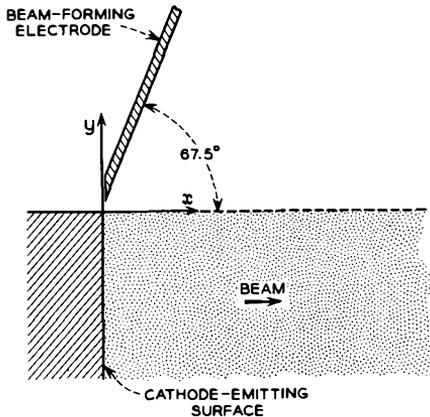


Fig. 4.5-3 An expanded view of the edge of the beam in the region of the cathode.

cathode emitting surface. If the dimensions of the portion of the cathode shown in the figure are assumed to be small compared with the overall cathode dimensions, we can further assume that the portion of the emitting surface shown in the figure is *planar* and that the beam extends a considerable distance above and below the page. The potential problem then reduces to a *two-dimensional* one and is thereby simplified. In the portion of the beam we are considering, the potential will vary approximately as in the planar diode and will be given by

$$V = Ax^{4/3} \tag{4.5-1}$$

where $A$ is a constant, and $x$ is the distance measured from the emitting surface.

Appendix VIII considers two-dimensional potentials which are symmetric about an axis. It is shown there that, if the $x$ axis of a Cartesian coordinate system is the axis of symmetry and if $V = f(x)$ is the potential on the $x$ axis, $V = \mathrm{Re}\, f(x + jy) = \frac{1}{2}[f(x + jy) + f(x - jy)]$ is the potential throughout the $x$-$y$ plane. Furthermore, symmetry of potential about the $x$ axis implies that $\partial V/\partial y = 0$ at $y = 0$, since the potential and its derivatives are continuous in a charge-free region.

Suppose we were to establish in the region just above the beam in Figure 4.5-3 a potential given by

$$V = \mathrm{Re}\, A(x + jy)^{4/3} = \frac{A}{2}[(x + jy)^{4/3} + (x - jy)^{4/3}] \tag{4.5-2}$$

This potential has the property that for $y = 0$ it reduces to $V = Ax^{4/3}$, and furthermore that $\partial V/\partial y = 0$ at $y = 0$, so that there would be no transverse force on the electrons at the edge of the beam.

To establish such a potential, the beam-forming electrode must be shaped to conform to an equipotential defined by Equation (4.5-2) and must be operated at that potential. A convenient potential is that of the cathode, since no additional biasing supply is needed in this case. The equipotential corresponding to cathode potential is obtained by setting $V = 0$ in Equation (4.5-2) and is given by

$$y = (1 + \sqrt{2})x \qquad\qquad (4.5\text{-}3)$$

This is the equation of a straight line making an angle of 67.5 degrees with the $x$ axis.

Thus, close to the electron beam, the beam-forming electrode makes an angle of 67.5 degrees with the beam edge, since in this region the approximations of a two-dimensional potential and a planar cathode are reasonably valid. The shape of the beam-forming electrode further from the beam and the shape of the anode are so chosen that they produce a potential along the edge of the beam which matches the potential that is characteristic of electron flow between concentric spheres. Often an electrolytic tank[6] is used to determine experimentally suitable electrode contours.

In the region of the anode aperture there is a component of electric field directed toward the axis of the beam, and this acts as a diverging lens. If the anode aperture is small compared with the anode-to-cathode distance, the focal length of this lens is[7] $4V_{ao}/V'$, where $V_{ao}$ is the anode voltage, and $V'$ is the potential gradient on the cathode side of the anode aperture. (See Problem 3.1.) The lens causes the off-axis electrons to receive a small deflection away from the axis. Beyond the anode aperture, the radial electric field of the beam causes a further deflection of the off-axis electrons away from the axis, with the result that the beam ultimately reaches a minimum diameter and then diverges. If an axial magnetic field is used to confine the beam, the beam would normally be launched into the field near the point of its minimum diameter.

The discussion of Section 4.2 concerning the relationship between anode current and anode voltage for a space-charge-limited diode applies equally well to electron guns such as those illustrated in Figures 4.5-1(a) and 4.5-1(b). Over the range of cathode currents for which space-charge-limited conditions prevail, the beam current varies as the 3/2 power of the

---

[6]Reference 4a, p. 180.

[7]The effect of the finite size of the anode aperture in Pierce electron guns has been considered by Danielson et al., Reference 4.7, who conclude that in a typical case the focal length given by the above expression should be divided by about 1.1.

anode voltage. Similarly, if the linear dimensions of an electron gun are
scaled by a constant factor, the same beam current is obtained for the same
applied anode voltage, and the beam dimensions scale with the other
linear dimensions of the gun, provided space-charge-limited conditions
prevail.

The ratio of beam current to (beam voltage)$^{3/2}$ is a constant for any
particular electron gun design over the range of beam currents for which
space-charge-limited conditions prevail. The ratio is known as the per-
veance of the gun and is a measure of the amount of beam current the gun
can generate for a given applied voltage. If two guns have the same geome-
try, but differ in their linear dimensions by a constant factor, they both
have the same perveance.
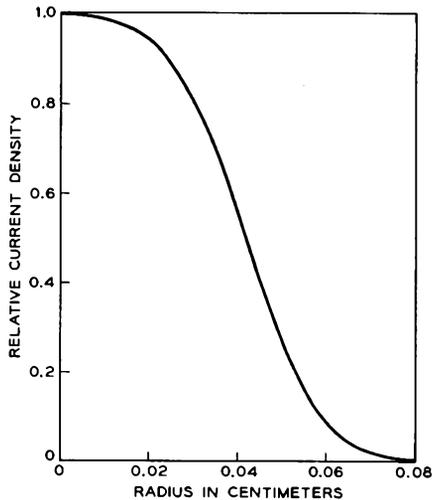


FIG. 4.5-4 Beam current density vs. radial distance from the axis at the point of
minimum beam diameter for the electron gun illustrated in Figure 4.5-1(a).

The effects of thermal velocities in Pierce electron guns have been con-
sidered by Cutler and Hines,[8] and later by Danielson et al.[9] The parameters
that apply to the beam at its point of minimum diameter have been sum-
marized in a family of curves by Herrmann.[10] Data concerning the distribu-
tion of current density across the beam cross section, the size of the beam at
its point of minimum diameter, and the location of the point of minimum

---

[8]Reference 4.8.
[9]Reference 4.7.
[10]Reference 4.9. See also Reference 4.10.

diameter with respect to the anode aperture can be obtained by reference to
the papers by these authors. Figure 4.5-4 shows the calculated current
density vs. radial distance from the beam axis at the point of minimum
beam diameter for the electron gun illustrated in Figure 4.5-1(a). The
curves given by Herrmann also indicate that a nonthermal electron emitted
from the edge of the cathode of this electron gun would arrive at the plane
containing the minimum beam diameter at a radius of 0.045 cm from the
axis. Notice that Figure 4.5-4 indicates that some of the thermal electrons
are far beyond this radius when they reach the plane of minimum beam
diameter. The beam radius at the cathode is 0.24 cm.

A measure of the distribution of transverse velocities in the beam at the
point of minimum diameter can be obtained by assuming that the radial
fields acting on the beam between the cathode and the point of minimum
diameter are directly proportional to distance from the beam axis.[11] We
then can use Equation (4.4-23) to express the probability that an electron
at the point of minimum diameter has transverse velocity in the range $u_t$ to
$u_t + du_t$ as

$$dP(u_t) = \frac{mu_t}{kT}\left(\frac{r_{\min}}{r_c}\right)^2 \epsilon^{-mu_t{}^2(r_{\min}/r_c)^2/2kT} du_t \qquad (4.5\text{-}4)$$

where $r_c$ is the beam radius at the cathode, and $r_{\min}$ is the beam radius at the
point of minimum diameter. This result is sometimes interpreted by saying
that the effective "temperature" of the beam generated by the gun is
$(r_c/r_{\min})^2 T$. For the gun shown in Figure 4.5-1(a) and for a cathode temper-
ature of 1000°K, we can obtain a first-order estimate of the beam tempera-
ture at the point of minimum diameter by setting $r_c = 0.24$ cm and $r_{\min} =$
0.045 cm, or the minimum distance from the beam axis to the trajectory of a
nonthermal electron emitted from the edge of the cathode. The resulting
beam temperature is $(0.24/0.045)^2 \times 1000 = 28{,}000°K$.

The high transverse velocities in a beam generated by a convergent elec-
tron gun increase the difficulty of focusing the beam by any of the several
methods described in Section 3.4. Higher focusing fields are required to
confine the beam to a given diameter than would be predicted by simple
theory which assumes laminar electron flow.

In most convergent electron guns, the total beam current determines the
cathode area. On the one hand, the cathode emitting surface is character-
ized by a maximum emission current density consistent with long life of the

---

[11]This assumption implies that the beam current density is uniform over the beam
cross section. However, from Figure 4.5-4 it is evident that, in fact, the beam current
density at the point of minimum diameter falls off rapidly from a radius about equal to
one third the beam radius. Consequently, the estimate of the beam temperature which
follows Equation (4.5-4) can only be considered as a first-order estimate.

emitter, so that the total beam current determines the minimum cathode area consistent with long cathode life. On the other hand, a larger cathode area than necessary would be wasteful of heater power, and the transverse velocities in the beam for a given minimum beam diameter would be unnecessarily high.

Some reflex klystron oscillators require high-current, high-current-density beams at relatively low voltages, often a few tens of milliamperes at a few hundred volts. Electron guns which produce these beams are of much higher perveance than the gun illustrated in Figure 4.5-1(a). One way to increase the perveance of an electron gun is to reduce the anode-to-cathode spacing. This, in turn, necessitates opening the anode aperture in order to pass the beam. However, it is found that, if the diameter of the anode aperture approaches half the anode-to-cathode spacing, the potential at the center of the aperture falls sufficiently below anode potential that the spherical diode is no longer approximated. In this case, the current density drawn from the edge of the cathode is greater than that drawn from the center of the cathode, and spherical aberration in the accelerating field introduces relatively high transverse velocities in the beam. Furthermore, it can be shown that, if the ratio $\bar{r}_c/\bar{r}_a$ = (radius of cathode emitting surface measured from the center of curvature of the emitting surface)/(distance from anode aperture to the center of curvature of the cathode emitting surface) is reduced below 1.4, the lens at the anode aperture becomes sufficiently strong that the beam beyond the anode is divergent.

Higher perveance also can be obtained by maintaining a relatively large $\bar{r}_c/\bar{r}_a$ and by increasing the angle of convergence (i.e., the angle subtended by a diameter of the cathode at the center of curvature of the cathode emitting surface). This is the approach used in the electron gun shown in Figure 4.5-1(b). The angle of convergence in this case is 145 degrees, or nearly 6 times that of the gun shown in Figure 4.5-1(a). However, the anode aperture is still relatively small, and, in fact, the anode is shaped to follow the beam contour. Although an appreciably higher perveance is obtained in this manner, spherical aberration in the accelerating field causes many of the electrons emitted from the edge of the cathode to cross the axis of the beam near the point of minimum diameter. As a consequence of this, the electron flow is far from laminar, and the transverse velocities are large. The beam is therefore difficult to confine with a magnetic field. The use of even higher angles of convergence would lead to still greater transverse velocities, and few applications could use such a gun.

Figure 4.5-5 shows a plot of beam current vs. anode voltage for the gun shown in Figure 4.5-1(b). The plot is made on "two-thirds power" paper in which the ordinate scale is proportional to the 3/2 power of linear distance measured up the page from the origin, while the abscissa scale is linear. The
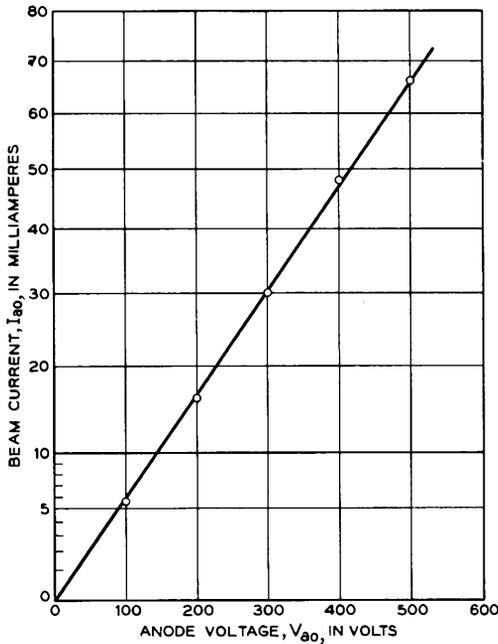
FIG. 4.5-5 A plot of beam current vs. anode voltage for the gun shown in Figure 4.5-1(b).

straight-line relationship between $I_{ao}$ and $V_{ao}$, when plotted on this paper, implies that $I_{ao}$ is proportional to $V_{ao}^{3/2}$, as we would expect from our earlier discussion.

Figure 4.5-1(c) illustrates an electron gun which is used in a low-noise traveling-wave amplifier. The gun is operated in a uniform axial magnetic field of 0.06 weber/meter², and the electron motion is similar to that described in Section 3.4(b) under the heading of Confined Flow. Several apertured accelerating electrodes are provided. The potentials of these electrodes are adjusted to minimize the amplitude of noise signals excited in the electron beam by statistical fluctuations in the electron emission velocity and current at the cathode. (The reduction of noise in an electron beam by this method is described in Chapter 13.) The uniform axial magnetic field might be provided by the permanent magnet circuit illustrated in Figure 1.5-6.

The cathode of the gun illustrated in Figure 4.5-1(c) is planar and of diameter 0.63 mm. A beam current of 0.5 milliamp is drawn from the cathode, and the cathode current density is 160 ma/cm². As the beam

leaves the cathode, its diameter shows a slight increase because of transverse emission velocities and radial fields in the accelerating region. However, the axial magnetic field confines the beam diameter sufficiently that the beam can pass through a helix-type slow wave circuit of inside diameter 1.3 mm and length 13 cm with less than 0.5 microamp interception.

### b.  Electron Guns for Cathode-Ray Tubes and Storage Tubes

These electron guns focus the beam to a crossover which is imaged onto the screen or storage surface by a lens beyond the crossover. Often an apertured electrode between the crossover and the lens passes only the central portion of the beam, so that the effects of aberrations in the gun and lens are small.

Generally the beam currents incident upon the screen or storage surface are lower than those used in microwave tubes. Storage tubes that make use of secondary emission from insulating materials often employ beams of a few microamperes at one or two thousand volts; cathode-ray tubes frequently employ beam currents of a few tens of microamperes at several thousand volts, perhaps 2 to 6 thousand volts; whereas the beam incident upon the screen of a television tube often amounts to a few hundred microamperes at 15 to 20 kv.
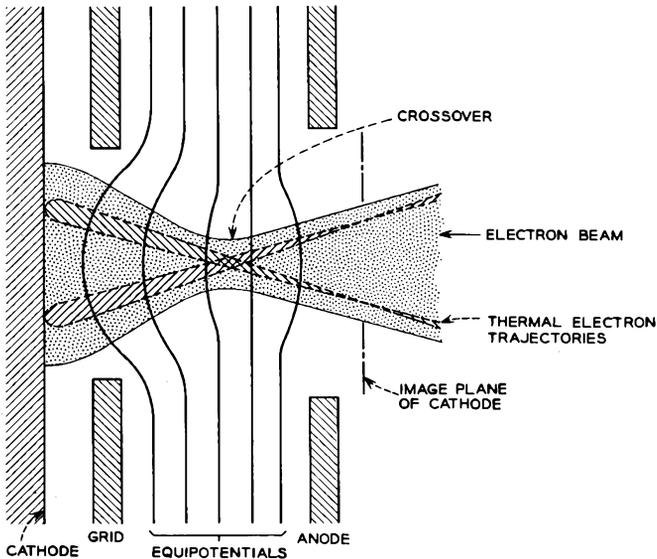


FIG. 4.5-6  A triode electron gun such as is used in a cathode-ray tube or storage tube.

The crossover is formed by a "triode" system consisting of a cathode, a "grid," and an anode, the grid and anode each having a single aperture. Such a structure is illustrated in Figure 4.5-6. The grid is always biased negatively with respect to the cathode, and consequently the current drawn from the cathode comes from a small circular area opposite the grid aperture. Equipotentials plotted in the figure show that electrons passing through the grid aperture experience a field which is both accelerating and convergent. The convergent field causes the nonthermal electrons to cross the axis between the cathode and the anode, and in this way the crossover is formed. Beyond the anode, the paths of the thermal electrons emitted from a single point on the cathode cross one another, and an image of the cathode is formed. The system is sometimes called an *immersion* lens, since the cathode is "immersed" in the accelerating field.

In the region of the crossover, the beam diameter reaches a minimum. The size of the minimum beam diameter is affected by three principal factors: Thermal emission velocities at the cathode, the accelerating potential, and aberrations in the convergent field which forms the crossover. Space charge may also affect the beam diameter at the crossover if the beam current is high and the beam voltage is low. If the convergent field were aberration-free, and if space-charge effects were small, the nonthermal electrons emitted from all parts of the cathode surface would cross the axis at essentially the same point. In this case, we might further assume that a thermal electron passing the crossover would be displaced from the axis by a distance proportional to its initial transverse velocity and independent of its point of emission on the cathode. Suppose that an electron emitted from the cathode with transverse velocity equal to $\sqrt{kT/m}$ were displaced a distance $\sigma$ from the axis by the time it reached the crossover. Then, using the arguments presented in connection with Equations (4.4-2) and (4.4-3), it is easily shown that the current density in the region of the crossover would be proportional to $\epsilon^{-r^2/2\sigma^2}$, where $r$ is the distance from the axis to the point where the beam current density is determined. If the lens system beyond the crossover is aberration-free, and if the beam is focused to a second crossover at the screen, the current density incident upon the screen is also of this form, but with a $\sigma$ increased by the magnification of the lens.

If the grid is made sufficiently negative, the beam current is cut off. Clearly the cutoff condition will prevail when the off-cathode potential gradient at a point on the cathode surface directly opposite the center of the grid aperture is zero or negative. Figure 4.5-7(b) shows a plot of the grid cutoff voltage vs. anode voltage for the triode shown in Figure 4.5-7(a). The straight-line relationship can be explained by noting that the net off-cathode potential gradient is a superposition of that caused by the grid and that caused by the anode, so that doubling the anode voltage requires

double the grid voltage in order to keep the off-cathode potential gradient at the center of the cathode equal to zero. The slope of the line, of course, is dependent upon the electrode dimensions and spacings.

The difference between the applied grid voltage and the cutoff voltage is called the grid drive voltage. As the grid is made more positive than cutoff, the area of the region of the cathode from which current is drawn increases, and the current density drawn from regions of the cathode surface which are already contributing to the beam current increases. Figure 4.5-7(c) shows
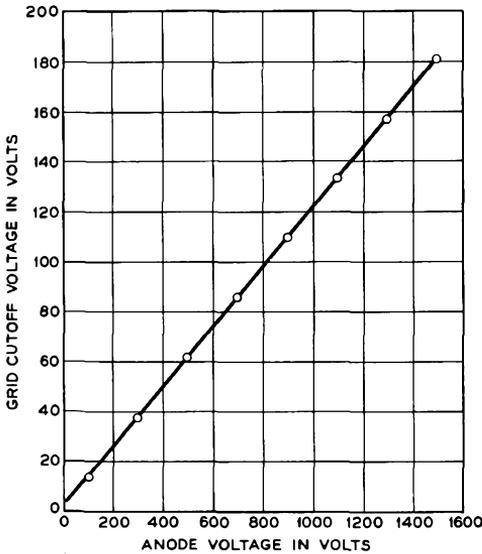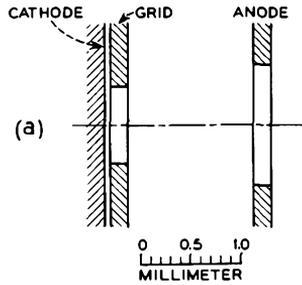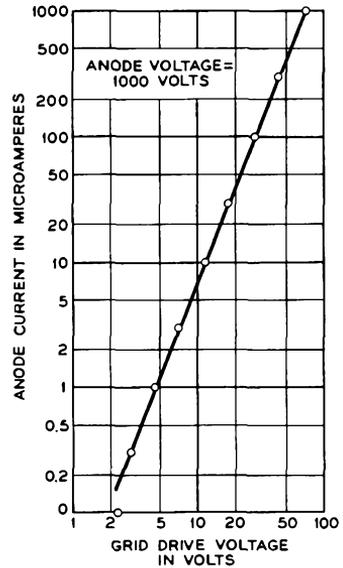


FIG. 4.5-7 Plots of grid cutoff voltage vs. anode voltage and beam current vs. grid drive for the triode gun structure shown in part (a) of the figure.

a plot of beam current vs. grid drive voltage for the triode shown in Figure 4.5-7(a). The slope of the straight-line relationship indicates that the beam current of this structure increases as about the 2.4 power of the grid drive voltage over the range of values for which the data are plotted.

Many experimental data concerning relationships between design parameters and the electrical performance of electron guns for cathode-ray tubes and storage tubes are presented in an informative paper by Hilary Moss.[12]

### PROBLEMS

4.1 For the conditions of applied anode voltage and cathode emission illustrated in Figure 4.1-1(d), show that the time taken for an electron to travel from the cathode to the anode of a planar diode is 3/2 as great as the time taken when no space charge is present. Assume zero emission velocity and $dV/dx = 0$ at the cathode.

4.2 The 3/2 power law of anode current vs. anode voltage does not apply to a diode operating under temperature-limited conditions. Explain why the arguments presented in Section 4.2 are not applicable in this case.
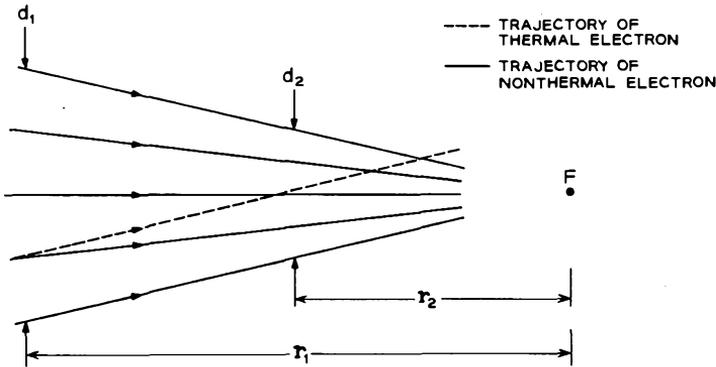


Problem 4.3

4.3 The figure shows a beam of electrons which is convergent upon the point $F$. The trajectories of several nonthermal electrons and one thermal electron are shown in the figure. The beam current density is assumed to be small, and no external fields are applied in the region of the beam. From geometrical considerations show that the thermal electron crosses the nonthermal electron trajectories with a transverse component of velocity which varies inversely as the beam diameter, and hence that

$$u_{t2} = \frac{\bar{r}_1}{\bar{r}_2} u_{t1} = \frac{d_1}{d_2} u_{t1}$$

---

[12]Reference 4e.

where $u_{t1}$ and $u_{t2}$ are, respectively, the transverse velocity of the thermal electron relative to the nonthermal electron trajectories at the points where the beam diameter is $d_1$ and $d_2$, respectively.

## REFERENCES

Three general references on electron guns are:

4a. J. R. Pierce, *Theory and Design of Electron Beams*, 2nd Ed., D. Van Nostrand Co., Inc., Princeton, N. J., 1954.

4b. C. Suskind, *Advances in Electronics and Electron Physics* **VIII**, 363, 1956.

4c. O. Klemperer, *Electron Optics*, 2nd Ed., Chapter 9, Cambridge University Press, Cambridge, England, 1953.

The following two papers describe electron guns for cathode-ray tubes. Many data of a practical nature are given in the second paper.

4d. H. Moss, *J. British IRE* **5**, 10, 1945.

4e. H. Moss, *J. British IRE* **6**, 99, 1946.

Other references covering specific items discussed in the chapter are:

4.1   I. Langmuir and K. T. Compton, *Revs. Modern Phys.* **3**, 191, 1931.

4.2   I. Langmuir and K. Blodgett, *Phys. Rev.* **22**, 347, 1923.

4.3   I. Langmuir and K. Blodgett, *Phys. Rev.* 24, 49, 1924.

4.4   J. R. Pierce, *J. Appl. Phys.* **10**, 715, 1939.

4.5   D. B. Langmuir, *Proc. IRE* **25**, 977, 1937.

4.6   J. R. Pierce, *J. Appl. Phys.* **11**, 548, 1940.

4.7   W. E. Danielson, J. L. Rosenfeld, and J. A. Saloom, *Bell System Tech. J.* **35**, 375, 1956.

4.8   C. C. Cutler and M. E. Hines, *Proc. IRE* **43**, 307, 1955.

4.9   G. Herrmann, *J. Appl. Phys.* **28**, 474, 1957.

4.10  P. T. Kirstein, *IEEE Trans. on Electron Devices*, Vol. ED-10, 69, 1963.

# *Chapter 5*

## GRID-CONTROLLED  TUBES —
## STATIC  CHARACTERISTICS

If a grid is placed in front of a thermionically emitting cathode and if the current drawn from the cathode is space-charge-limited, the voltage applied to the grid can be used to control the current drawn from the cathode. A triode vacuum tube consists of a cathode, a control grid, and an anode. Usually, a dc bias voltage is applied to the control grid to make it negative with respect to the potential minimum and thereby reduce the interception of the electron beam by the grid. By superimposing a small ac signal on the dc bias voltage, the beam current can be modulated with little expenditure of power. AC power amplification is then obtained by causing the ac current flowing in the anode circuit to pass through a load resistance or impedance of suitable size.

Additional grids also may be inserted between the control grid and anode. Generally, these are held at fixed potentials, but in some cases their control action on the beam is used to mix signals from independent sources. A tetrode has a control grid and a screen grid, whereas a pentode has a control grid, a screen grid, and a suppressor grid.

Usually the screen grid in a tetrode is biased at a fixed positive potential with respect to the cathode. Its shielding action between the anode and control grid reduces the capacitance between these electrodes and hence the coupling between the output circuit and the input circuit. In addition, the current reaching the anode of a tetrode is determined largely by the voltages applied to the control grid and screen grid and is nearly independent of anode voltage over a wide range of positive anode voltages. This is an advantage when high-voltage amplification per stage is required.

Many tetrodes are constructed with a large spacing between the screen grid and anode so that space charge in the interelectrode space will depress the potential between the electrodes and prevent secondary electrons

149

emitted from one electrode from reaching the other. The large spacing also reduces the output capacitance of the tube, and this is an advantage in high-gain, broadband amplifiers.

In the pentode, the supressor grid is inserted between the screen grid and anode. It is of a very coarse mesh, and usually it is biased at cathode potential. The suppressor grid depresses the potential between the screen grid and anode and thereby prevents secondary-electron exchange between these electrodes.

In the present chapter we consider the dc behavior of triodes, tetrodes, and pentodes. The mechanical construction and performance of one example of each of these tubes is described. In Chapter 6 the use of grid-controlled tubes in simple low-frequency amplifier circuits is described, and in Chapter 7 the problems and limitations of grid-controlled tubes when operated at very high frequencies are discussed.

As in the previous chapter, we shall use the subscript $o$ to designate dc electrode voltages and currents. Thus $V_{ao}$ and $I_{ao}$ are the dc anode voltage and current.

## 5.1  A Particular Triode and its Electric Field in the Absence of Space Charge

In this section we first describe the electrode geometry and construction of a particular triode, the Western Electric 417A. Then we consider the electric fields in the interelectrode space of this tube when various potentials are applied to the electrodes, and when no space charge is present. The electrical characteristics of the 417A with space-charge-limited operation are described in Section 5.2.

The construction of the 417A is shown in Figure 5.1-1. This is an example of a triode in which the grid is mounted very close to the cathode to increase the effectiveness of the grid voltage in controlling the current reaching the anode. The tube is used in the input stage[1] of a broadband amplifier which amplifies signals with frequencies varying from 58 to 90 Mc/sec.

Table 5.1-1 summarizes the important dimensions of the 417A electrodes. The cathode area is 0.38 cm².

The cathode consists of a short length of nickel tubing that is flattened to provide two planar emitting surfaces. The wall thickness of the tubing is

---

[1]Actually two tubes are used in a "cascode" stage. Triodes are preferred for the input stages of high-gain amplifiers because they generate less noise than tetrodes and pentodes. Since the noise generated by the input stages is amplified by all the remaining stages, the noise output of a high-gain amplifier is much reduced by using low-noise tubes in the input stages.
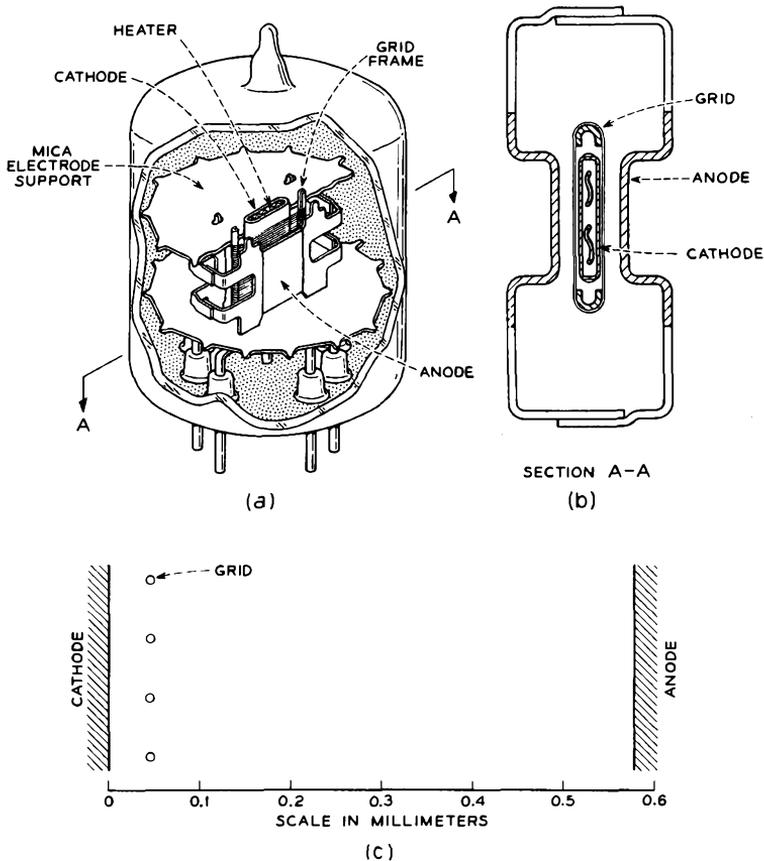
FIG. 5.1-1  The construction of the Western Electric 417A triode.  The overall
height of the tube is 4.4 cm.

TABLE 5.1-1

|  |  | *Millimeters* |
|---|---|---|
| Grid wire diameter | ......................................... | 0.0074 |
|  |  | (or 0.00029 inch) |
| Grid pitch, $P$ (or center-to-center spacing of the grid wires) | ....... | 0.065 |
| Cathode-to-grid spacing, $d_{cg}$ | ................................... | 0.045 |
| Cathode-to-anode spacing, $d_{ca}$ | ................................ | 0.58 |

0.075 mm.  A "double-carbonate" oxide coating is applied to the emitting
surfaces.

The grid is made by winding tungsten wire onto a molybdenum frame and
then brazing the wire to the frame with a small amount of gold.  The high

tensile strength of tungsten permits winding the grid wire onto the molybdenum frame while it is under appreciable tension.[2] This ensures that the resonant frequency of the grid wires is high and minimizes the tendency for mechanical excitation when the tube is vibrated. Vibration of grid wires in grid-controlled tubes is a principal source of "microphonics." After brazing the grid wires to the frame, the grid assembly is gold plated to raise its work function. This reduces thermionic emission from the grid wires, which would otherwise take place when the grid is heated by thermal radiation from the cathode.[3]

The anode, or "plate," of the 417A is made of nickel which is coated with fine carbon particles in order to increase the heat radiation from the outer surface and thereby reduce the anode operating temperature. Under typical operating conditions, a power of 3.5 watts is dissipated in the anode, and the anode temperature is between 500° and 600°C.

Let us consider now the potential obtained in the interelectrode space of the 417A when various voltages are applied to the electrodes and when no space charge is present. It is convenient to think of the potential as being a linear combination of two separate potential functions, which we shall denote $F_1(x,y,z)$ and $F_2(x,y,z)$. $F_1(x,y,z)$ is the potential obtained in the interelectrode space when the grid is at $+1$ volt, and the cathode and anode are at ground potential. $F_2(x,y,z)$ is the potential obtained when the anode is at $+1$ volt and the grid and cathode are grounded. Clearly the functions $F_1$ and $F_2$ satisfy Laplace's Equation, and so does any linear combination of them. In particular, the linear combination given by $V(x,y,z) = V_{go}F_1 + V_{ao}F_2$ satisfies the boundary conditions for the case in which the cathode is at ground potential, and the grid and anode are at $V_{go}$ and $V_{ao}$ volts, respectively. Since it also satisfies Laplace's Equation, it must be the potential function actually obtained with these boundary conditions.

Let us look more closely now at the functions $F_1$ and $F_2$. Figure 5.1-2(a) shows a plot of equipotential contours of the function $F_1$ in a portion of the interelectrode space of the 417A.[4] Figure 5.1-2(b) shows plots of $F_1$ along two lines running from the cathode to the anode; one line passes through the center of a grid wire, and the other passes midway between grid wires. Figures 5.1-2(c) and (d) show similar plots for the function $F_2$.

---

[2]About half the breaking tension of the tungsten wire is used.

[3]Grid emission tends to bias the grid in the positive direction. This increases the beam current and the power dissipation in the tube, which in turn raises the grid operating temperature and further aggravates the situation. In an extreme case, with a very high resistance in the grid circuit, a tube with high grid emission can be destroyed by excessive power dissipation in its electrodes.

[4]Plots such as this can be made with the aid of an electrolytic tank. See, for instance, Reference 4a, p. 180. Analytic expressions for $F_1$ and $F_2$ are given in Reference 5.1, Equations 1 to 4. Approximate expressions are derived in Appendix IX.

THE FUNCTION $F_1$



(a)



(b)

THE FUNCTION $F_2$



(c)



(d)
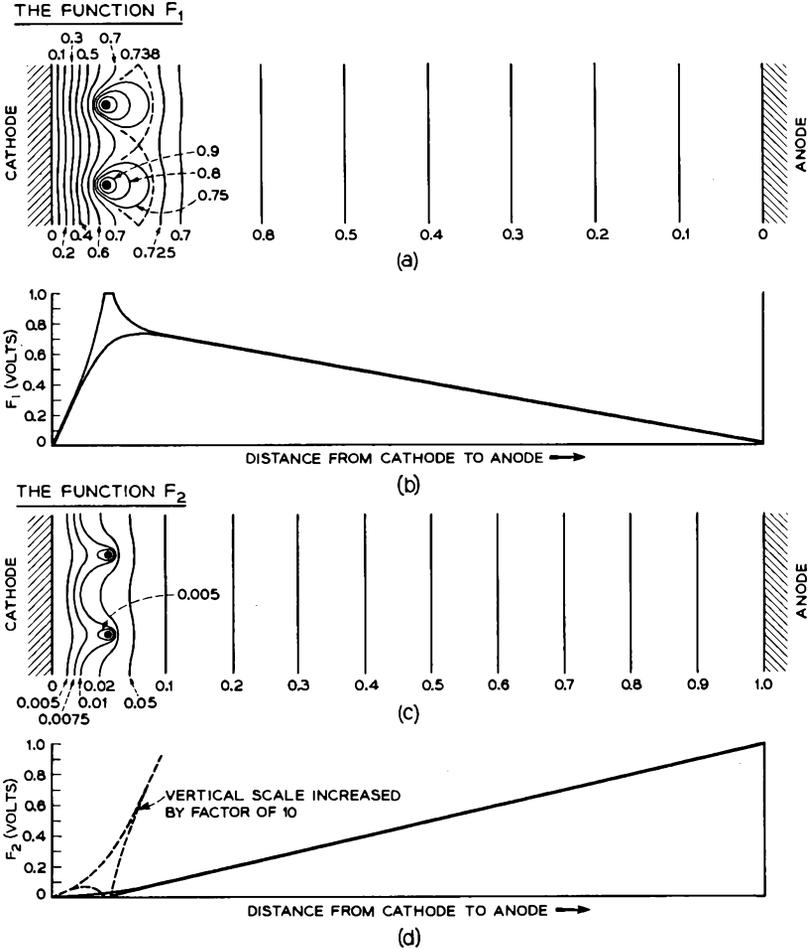
FIG. 5.1-2  Plots of the functions $F_1$ and $F_2$ for the interelectrode space of the 417A.

At the cathode the function $F_1$ has a slope of 176 volts/cm for the 417A, whereas the function $F_2$ has a slope of 3.8 volts/cm. The ratio of these electric fields is called the electrostatic amplification factor and is designated $\mu_{es}$. Thus

$$\mu_{es} = \frac{\left.\dfrac{\partial F_1}{\partial x}\right|_{x=0}}{\left.\dfrac{\partial F_2}{\partial x}\right|_{x=0}} \tag{5.1-1}$$

where $x$ measures distance in the direction normal to the cathode and is zero at the cathode. For the 417A, $\mu_{es} = 176/3.8 = 46$.

The electrostatic amplification factor measures the relative effectiveness of the grid and anode in creating an electric field at the cathode surface. In fact, a second definition of $\mu_{es}$ which follows directly from Equation (5.1-1) is that $\mu_{es}$ is minus the ratio of the anode voltage to grid voltage which gives zero electric field at the cathode. Finally, since electric fields can be superposed, we can write

$$\mu_{es} = -\frac{dV_{ao}}{dV_{go}}\bigg|_{\text{constant electric field at the cathode}} \qquad (5.1\text{-}2)$$

where $dV_{ao}$ and $dV_{go}$ are incremental changes in the anode and grid voltages which give zero change in the electric field at the cathode.

Figure 5.1-3 shows plots of $V = V_{go}F_1 + V_{ao}F_2$ in the region of the cathode and grid of the 417A for an anode voltage of 100 volts and three values of $V_{go}$. For the particular geometry of this tube, a grid voltage of $-2.2$ volts gives nearly zero electric field at the cathode. Using the second definition of the electrostatic amplification factor, given above, we find that $\mu_{es} = 100/2.2 = 45.5$, or approximately 46, in agreement with the value previously obtained using Equation (5.1-1). If the grid voltage is changed by 1 volt with constant anode voltage, the electric field at the cathode surface changes by

$$\frac{\partial F_1}{\partial x}\bigg|_{x=0} = 176 \text{ volts/cm}$$

Thus, we can expect that the control action of the grid voltage upon the current drawn from the cathode with space-charge-limited operation will be considerable.

From the foregoing discussion it is evident that the electrostatic amplification factor is entirely a function of the geometry of the electrodes. In Appendix IX it is shown that an approximate expression for the electrostatic amplification factor of a planar triode is given by

$$\mu_{es} = -\frac{2\pi d_{ga}}{P \ln\left(2\sin\dfrac{\pi R}{P}\right)} \qquad (5.1\text{-}3)$$

where $d_{ga}$ is the grid-to-anode distance, $P$ is the grid pitch, and $R$ is the grid-wire radius. The expression is valid when the cathode-to-grid spacing $d_{cg} \geq P$ and when $R \leq P/20$. The electrostatic amplification factor is independent of the area of the electrodes, but it increases as the grid-to-anode distance is increased. It also increases if the grid-wire radius and grid pitch are decreased in such a manner that the ratio of wire radius to

Fig. 5.1-3  Plots of $V = V_{go}F_1 + V_{ao}F_2$ in the region of the cathode and grid of the 417A for $V_{ao} = 100$ volts and $V_{go} = -3.4$, $-2.2$, and $-1.0$ volts.

Top panel labels:
$V_{go} = -3.4 \text{ V}$
$V_{ao} = 100 \text{ V}$
$V = -3.4 F_1 + 100 F_2$

LOCATION OF GRID WIRES

Middle panel labels:
$V_{go} = -2.2 \text{ V}$
$V_{ao} = 100 \text{ V}$
$V = -2.2 F_1 + 100 F_2$

LOCATION OF GRID WIRES

TO ANODE ⟶

POTENTIAL RELATIVE TO CATHODE IN VOLTS

Bottom panel labels:
$V_{go} = -1.0 \text{ V}$
$V_{ao} = 100 \text{ V}$
$V = -1.0 F_1 + 100 F_2$

LOCATION OF GRID WIRES

DISTANCE FROM CATHODE IN MILLIMETERS ⟶
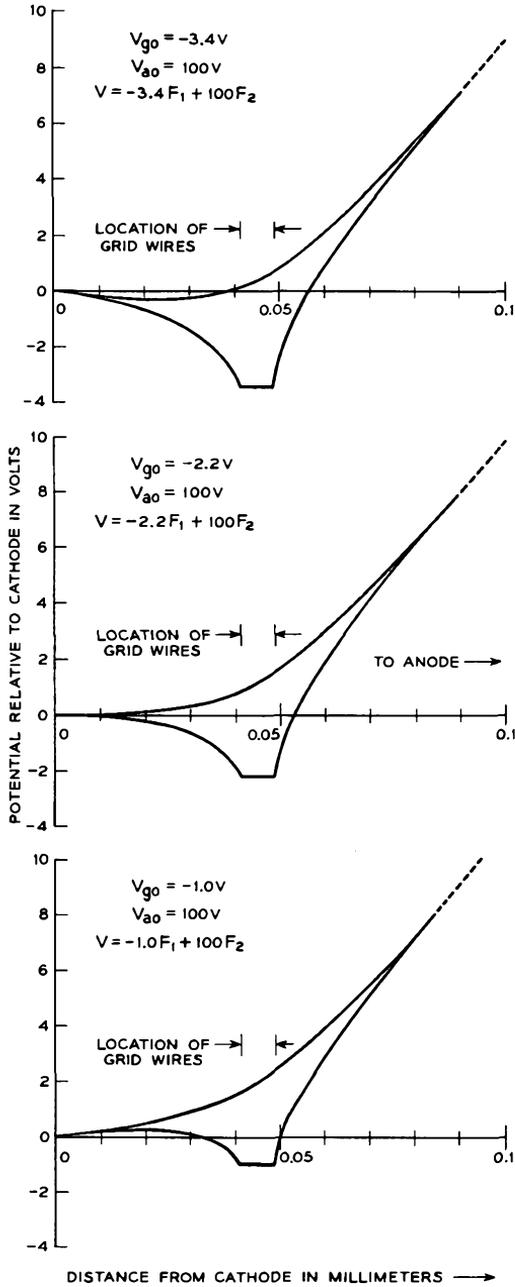
pitch is maintained constant. (The ratio $2R/P$ is sometimes called the *screening fraction* of the grid because it indicates the fraction of the electrode area which is screened or shadowed by the grid.) Note that Equation (5.1-3) indicates that $\mu_{es}$ is independent of the cathode-to-grid spacing. Also given in Appendix IX are expressions for the functions $F_1$ and $F_2$ which are valid when $d_{cg} \geq P$ and when $R \leq P/20$.

Since $V = V_{go}F_1 + V_{ao}F_2$, we can express the gradient of potential in the $x$ direction as

$$
\frac{\partial V}{\partial x} = V_{go}\frac{\partial F_1}{\partial x} + V_{ao}\frac{\partial F_2}{\partial x}
$$

$$
= \frac{\partial F_1}{\partial x}\left( V_{go} + V_{ao}\frac{\dfrac{\partial F_2}{\partial x}}{\dfrac{\partial F_1}{\partial x}} \right) \qquad (5.1\text{-}4)
$$

At the cathode, the potential gradient is therefore given by

$$
\left.\frac{\partial V}{\partial x}\right|_{x=0} = \left.\frac{\partial F_1}{\partial x}\right|_{x=0}\left( V_{go} + \frac{V_{ao}}{\mu_{es}} \right) \qquad (5.1\text{-}5)
$$

Thus the off-cathode field in the absence of space charge is proportional to the voltage $(V_{go} + V_{ao}/\mu_{es})$. We shall find in the next section that the electrical behavior of a triode in the presence of space charge is dependent upon a voltage which is very nearly equal to $(V_{go} + V_{ao}/\mu_{es})$.

## 5.2   The Triode with Space Charge

Grid-controlled tubes are almost always operated with the current drawn from the cathode space-charge-limited since only then is it possible for the grid to act effectively as a control electrode. If the cathode emission were ideally temperature-limited, the current drawn from the cathode would be independent of the grid voltage for all grid voltages at which temperature-limited emission prevailed.

Figure 5.2-1 shows the grid and anode "characteristic curves" for the 417A triode. These curves give the relationship between the current reaching the anode and the voltages applied to the grid and anode. The grid of the 417A is generally "biased" negatively with respect to the potential minimum to prevent it from intercepting the electron beam.

The circuit designer is frequently concerned with the small-signal behavior of the active devices in his circuits and consequently with the slopes of the curves relating the currents reaching the terminals of a device to the voltages applied between the terminals. In the case of the triode, the small-signal behavior of the tube can be described in terms of the slopes of the
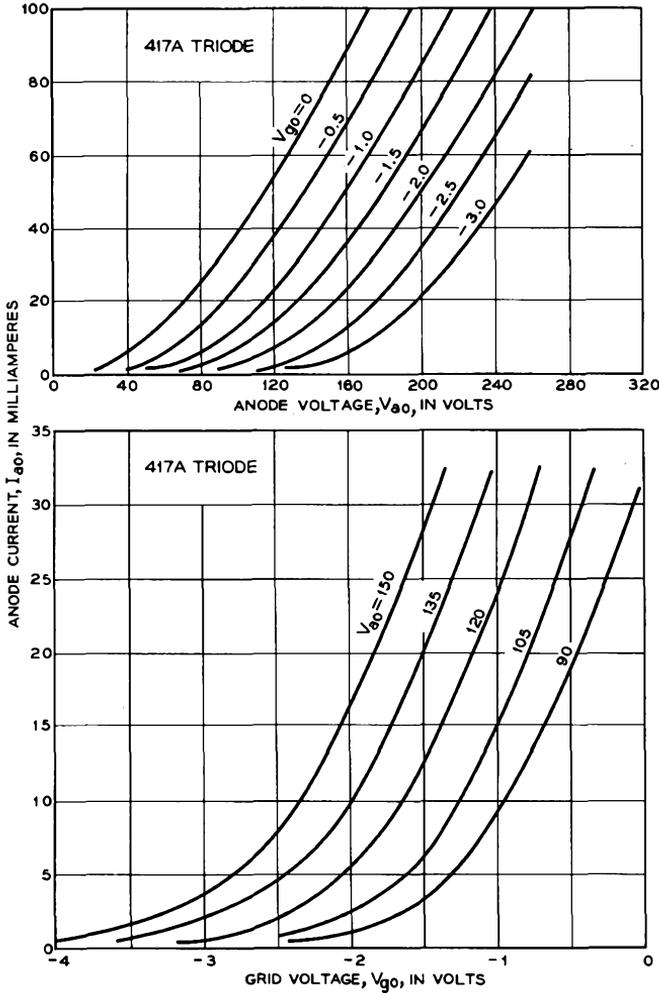
Fig. 5.2-1 The grid and anode characteristic curves for the 417A triode.

characteristic curves. Two parameters which are derived from these slopes are the transconductance and the dynamic anode resistance, or dynamic plate resistance. The transconductance is denoted by[5] $g_m$ and is defined as

$$g_m = \frac{dI_{ao}}{dV_{go}}\bigg|_{V_{ao}} = \frac{\partial I_{ao}}{\partial V_{go}} \tag{5.2-1}$$

_____

[5]Sometimes by $S_m$.

where $I_{ao}$ is the dc anode current. The dynamic anode resistance, denoted by $r_a$, is defined as

$$r_a = \frac{dV_{ao}}{dI_{ao}}\bigg|_{v_{go}} = \frac{\partial V_{ao}}{\partial I_{ao}} \qquad (5.2\text{-}2)$$
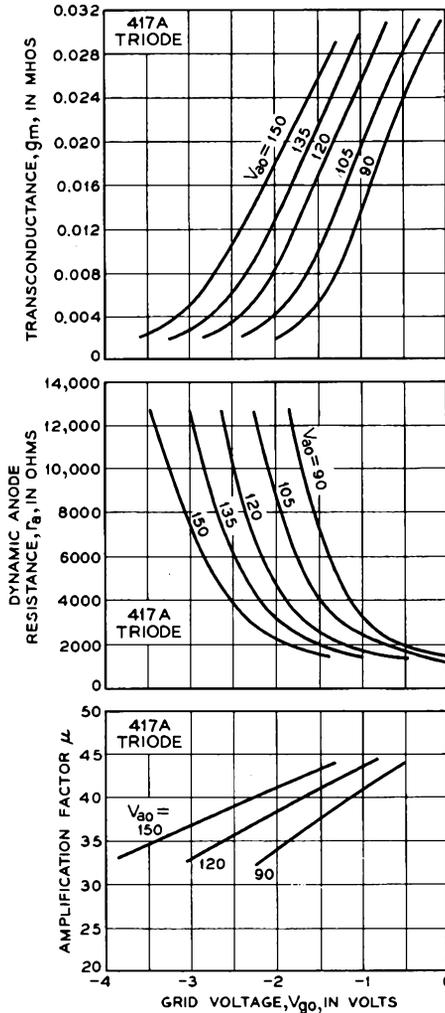


FIG. 5.2-2 The transconductance, the dynamic anode resistance, and the amplification factor of the 417A triode as a function of the grid voltage and anode voltage.

The product of $g_m$ and $r_a$ is called the *amplification factor* and is denoted by $\mu$. Thus

$$\mu = g_m r_a \tag{5.2-3}$$

The amplification factor also can be expressed in differential form, similar to the expressions given above for $g_m$ and $r_a$. Suppose that $V_{go}$ and $V_{ao}$ undergo differential changes $dV_{go}$ and $dV_{ao}$ at a time when the tube is drawing anode current $I_{ao}$. The resulting change in anode current is given by

$$dI_{ao} = \frac{\partial I_{ao}}{\partial V_{go}} dV_{go} + \frac{\partial I_{ao}}{\partial V_{ao}} dV_{ao}$$

$$= g_m dV_{go} + \frac{1}{r_a} dV_{ao} \tag{5.2-4}$$

Now if $dV_{go}$ and $dV_{ao}$ are such that $dI_{ao} = 0$, then

$$g_m r_a = -\frac{dV_{ao}}{dV_{go}}\bigg|_{I_{ao}} \tag{5.2-5}$$

and hence

$$\mu = -\frac{dV_{ao}}{dV_{go}}\bigg|_{I_{ao}} \tag{5.2-6}$$

The amplification factor $\mu$ is usually approximately equal to the electrostatic amplification factor $\mu_{es}$. Comparison of Equations (5.1-2) and (5.2-6) makes this approximate equality seem reasonable.

Figure 5.2-2 shows plots of the transconductance, the dynamic anode resistance, and the amplification factor of the 417A for various values of grid voltage and anode voltage. Evidently the transconductance increases with increasing anode current, the amplification factor is nearly independent of anode current, and the dynamic anode resistance decreases with increasing anode current. Typical operating conditions for the 417A are given in Table 5.2-1.

Consider the dependence of the anode current on the electrode potentials. It is found experimentally that an approximate expression for the current drawn to the anode of a triode is

$$I_{ao} = C\left(V_{go} + \frac{V_{ao}}{\mu}\right)^n \tag{5.2-7}$$

where $C$ and $n$ are constants. Values of $n$ generally lie between $3/2$ and $2$, but in some cases may be as high as $5/2$. The expression is found to hold even for small positive grid voltages, provided the anode voltage is much greater than the grid voltage. The dependence of the anode current upon the voltage $V_{go} + V_{ao}/\mu$ is perhaps not surprising, since we found in the last section that the off-cathode field in the absence of space charge is pro-

## TABLE 5.2–1   SOME CHARACTERISTICS OF TUBES DESCRIBED IN CHAPTERS 5 AND 7

| | | 417A Triode | 416B Triode | 448A Tetrode | 1983 Tetrode | 350B Beam Power Tetrode | 403A/6AK5 Pentode |
|---|---|---|---|---|---|---|---|
| (a) | Anode Voltage, volts | 130 | 200 | 125 | 300 | 250 | 120 |
| | Anode Current, milliamperes | 27 | 30 | 25.5 | 40 | 62 | 7.5 |
| | Anode Power, watts | 3.5 | 6.0 | 3.2 | 12 | 15.5 | 0.9 |
| (b) | Screen Grid Voltage, volts | — | — | 125 | 200 | 250 | 120 |
| | Screen Grid Current, milliamperes | — | — | 9.0 | 5 | 7 | 2.5 |
| | Screen Grid Power, watts | — | — | 1.1 | 1.0 | 1.7 | 0.3 |
| (c) | Control Grid Bias, volts | −1.2 | −0.1 | −1.1 | −0.8 | −18 | −2.0 |
| (d) | Transconductance $g_m$, micromhos | 25,000 | 60,000 | 34,000 | 45,000 | 6,900 | 5,000 |
| | Dynamic Anode Resistance $r_a$, ohms | 1,700 | 5,000 | 32,000 | — | 18,000 | 300,000 |
| | Amplification Factor $\mu$ | 43 | 300 | 1100 | — | 124 | 1500 |
| (e) | Cathode Area A, cm$^2$ | 0.38 | 0.164 | 0.86 | 0.164 | 3.6 | 0.18 |
| | Cathode Current Density, milliamperes/cm$^2$ | 71 | 180 | 40 | 270 | 19 | 56 |
| | Transconductance Per Unit Area of Cathode $g_m/A$, micromhos/cm$^2$ | 66,000 | 370,000 | 40,000 | 280,000 | 1,900 | 28,000 |
| (f) | Cathode-Control Grid Spacing, millimeters | 0.045 | 0.018 | 0.060 | 0.030 | 0.56 | 0.077 |
| | Input Capacitance (Operating) $C_i$, pf | — | — | 23 | 13.6 | — | 6.4 |
| | Output Capacitance $C_o$, pf * | 1.8 | 1.4 | 2.1 | 1.3 | 8.0 | 2.0 |
| | Gain-Bandwidth Product With Single Tuned Circuit Between Stages $\dfrac{g_m}{2\pi(C_o+C_i)}$, megacycles | — | — | 215 | 480 | — | 95 |
| (g) | Cathode Lead Inductance, millimicrohenries | — | — | 4 | 0.7 | — | — |

* I picofarad = $10^{-12}$ farad.

160

PRINCIPLES OF ELECTRON TUBES

portional to $V_{go} + V_{ao}/\mu_{es}$, and we have noted in the present section that $\mu \approx \mu_{es}$.

In a triode in which the grid is well beyond the region of the potential minimum, the anode and grid combine to create a field on the grid side of the potential minimum which is approximately proportional to $V_{go} + V_{ao}/\mu$. For such a triode we would expect from the discussion given in Section 4.2 that a value of $n$ equal to 3/2 would be applicable in Equation (5.2-7), and the anode current would vary as $(V_{go} + V_{ao}/\mu)^{3/2}$.

The constant $C$ in Equation (5.2-7) can be evaluated for a planar triode in which $n = 3/2$ in the manner outlined below. Let us first assume that the grid is removed and the tube is operated as a space-charge-limited diode. From Equation (4.1-9) the current drawn to the anode would be

$$I_{ao} = 2.33 \times 10^{-6} \frac{V_{ao}^{3/2}}{d_{ca}^{2}} A \qquad (5.2\text{-}8)$$

where $d_{ca}$ is the distance from the potential minimum to the anode, and $A$ is the cathode area. Solving for $V_{ao}$, we obtain

$$V_{ao} = \left[ \frac{I_{ao}}{2.33 \times 10^{-6} A} \right]^{2/3} d_{ca}^{4/3} \qquad (5.2\text{-}9)$$

Next let the grid be inserted at a distance $d_{cg}$ from the potential minimum, and let the applied grid voltage be that which was present in the beam at the same location before the grid was inserted. A plot of the potential distribution between the potential minimum and the anode for these conditions is
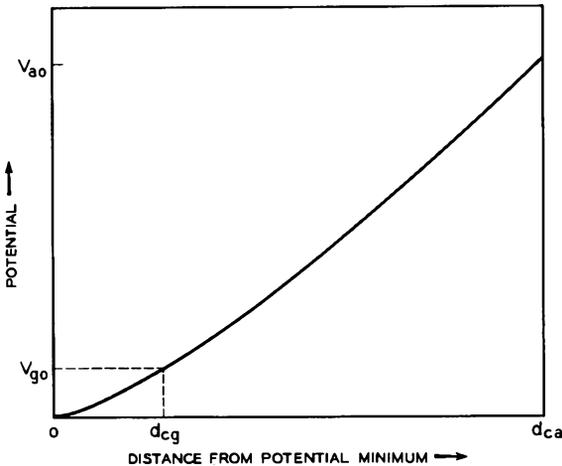


FIG. 5.2-3 The potential distribution in a space-charge-limited triode with the grid at beam potential.

shown in Figure 5.2-3. The applied grid voltage under these circumstances is given by

$$V_{go} = \left[\frac{I_{ao}}{2.33 \times 10^{-6}A}\right]^{2/3} d_{cg}^{4/3} \qquad (5.2\text{-}10)$$

With the grid at the potential of the surrounding beam, some of the beam current is intercepted by the grid wires. However, if the grid-wire diameter is much smaller than the grid-wire spacing, this interception will be a small part of the total anode current.

Substituting the above expressions for $V_{ao}$ and $V_{go}$ into Equation (5.2-7) and setting $n = 3/2$, we obtain

$$I_{ao} = C\frac{I_{ao}}{2.33 \times 10^{-6}A}\left[d_{cg}^{4/3} + \frac{d_{ca}^{4/3}}{\mu}\right]^{3/2} \qquad (5.2\text{-}11)$$

from which

$$C = \frac{2.33 \times 10^{-6}A}{\left[d_{cg}^{4/3} + \frac{d_{ca}^{4/3}}{\mu}\right]^{3/2}} \qquad (5.2\text{-}12)$$

Finally, the anode current can be expressed as

$$I_{ao} = \frac{2.33 \times 10^{-6}A(V_{go} + V_{ao}/\mu)^{3/2}}{d_{cg}^{2}\left[1 + \frac{1}{\mu}\left(\frac{d_{ca}}{d_{cg}}\right)^{4/3}\right]^{3/2}} \qquad (5.2\text{-}13)$$

This last equation states that the current density drawn from the cathode of a planar triode is the same as would be drawn by a planar diode having a cathode-to-anode distance of

$$d_{cg}\left[1 + \frac{1}{\mu}\left(\frac{d_{ca}}{d_{cg}}\right)^{4/3}\right]^{3/4}$$

and an applied anode voltage of $V_{go} + V_{ao}/\mu$. Replacing

$$d_{cg}\left[1 + \frac{1}{\mu}\left(\frac{d_{ca}}{d_{cg}}\right)^{4/3}\right]^{3/4}$$

by $d_e$, where $d_e$ is called the equivalent diode spacing of the triode, Equation (5.2-13) becomes $I_{ao} = 2.33 \times 10^{-6}A \ (V_{go} + V_{ao}/\mu)^{3/2}/d_e^{2}$. The distance $d_e$ is a function of the tube dimensions only; it is always greater than $d_{cg}$ and frequently less than $2d_{cg}$.

As the grid is moved closer to the cathode, the potential $V_{go} + V_{ao}/\mu$ affects not only the electric field on the anode side of the potential minimum, but *it affects the potential at the minimum*. The dependence of the anode current upon the voltage $V_{go} + V_{ao}/\mu$ then increases, and accordingly the

exponent $n$ in Equation (5.2-7) increases.  For many "close-spaced" triodes, that is, triodes with small $d_{cg}$, $n$ is more nearly equal to 2 than 3/2. Figure 5.2-4(a) shows a plot of $I_{ao}$ vs. $(V_{go} + V_{ao}/\mu)$ for the 417A triode. From the figure it can be seen that a value of $n$ equal to 2 is appropriate for
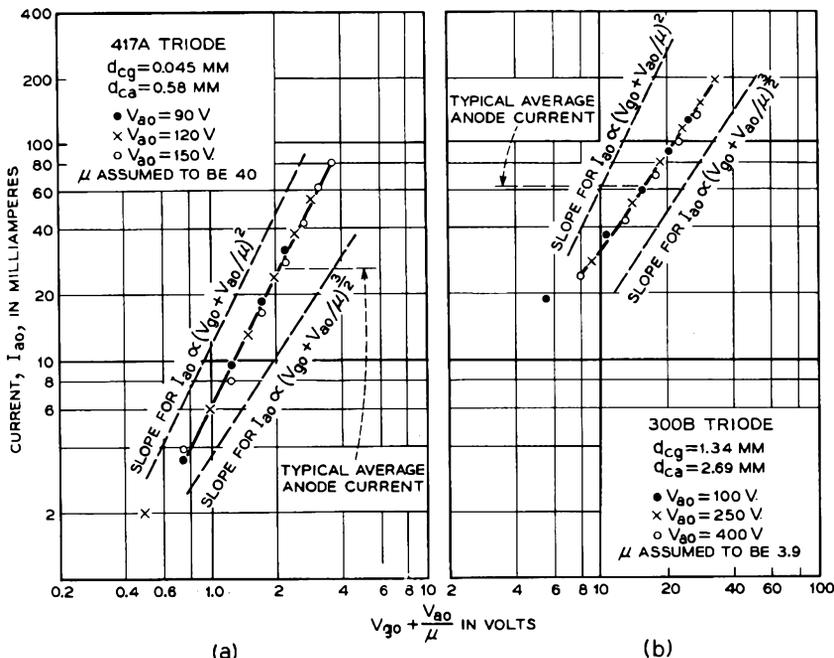


FIG. 5.2-4  Plots of $I_{ao}$ vs. $(V_{go} + V_{ao}/\mu)$ for the 417A triode and the 300B triode.

this tube.  Figure 5.2-4(b) shows a plot of $I_{ao}$ vs. $(V_{go} + V_{ao}/\mu)$ for the Western Electric 300B triode in which the cathode-to-grid spacing is approximately half the cathode-to-anode spacing.  Clearly, $n = 3/2$ is appropriate for this tube.

To determine the relative positions of the grid plane and the plane of the potential minimum in the 417A, we have plotted in Figure 5.2-5 the potential in a planar diode in which a current density of 0.071 amp/cm² passes the potential minimum.[6]  Such a current density might be typical of that passing the potential minimum in the 417A.  A cathode emission current density of 0.5 amp/cm² and a cathode temperature of 1025°K are assumed. The position of the grid wires in the 417A is shown in the figure.  The plane

---

[6]The method for obtaining the potential plot is described in References 5.2 and 5.3.

of the grid wires is about three times as far from the cathode as the plane of
the potential minimum. Of course, the potential on the grid side of the
potential minimum in the 417A would be much different from that shown
by the dashed curve in Figure 5.2-5, because the grid bias voltage depresses
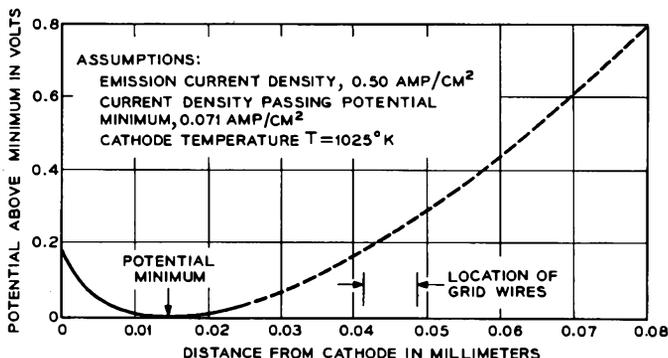


FIG. 5.2-5 The potential in a planar diode in which the current density passing
the potential minimum is typical of that used in the 417A.

the potential in the region of the grid wires, and the potential midway
between grid wires is much higher.

From Figure 5.1-3 it is evident that in the absence of space charge the
potential over a plane which lies one third of the way out from the cathode
to the grid wires is not uniform. Consequently, when space-charge-limited
conditions prevail, we would expect the potential at the potential minimum
would be slightly higher at points opposite the center of the opening be-
tween grid wires than directly under the grid wires. This means that the
current passing the potential minimum is lower directly under the grid
wires, and, if the grid voltage is made increasingly negative, the current
passing the potential minimum will first cut off directly under the grid
wires. This phenomenon is called *Inselbildung* or "island building." Some
consequences of Inselbildung are described in Reference 5.1. Inselbildung
effects become particularly important when small grid-to-cathode spacings
are used and when the ratio of grid pitch to grid-to-cathode spacing is large
(of the order of 1 or greater).

In Chapter 6 it is shown that high transconductance is needed for high
gain in an amplifier stage. Let us therefore proceed to examine what
factors affect the transconductance of a tube. Clearly, the transconduct-
ance is directly proportional to the cathode area $A$. An expression for the
transconductance can be obtained by differentiating Equation (5.2-7) with
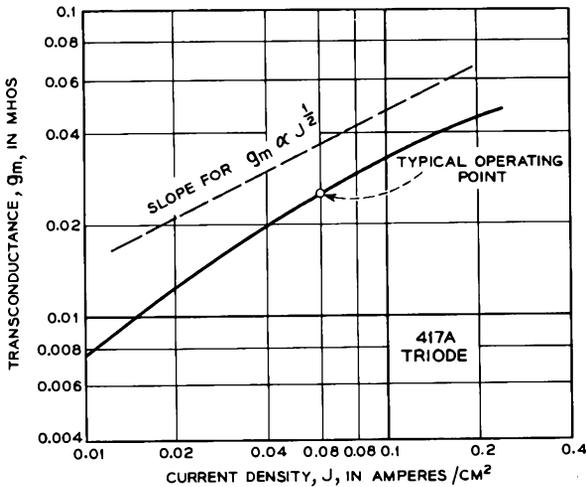
FIG. 5.2-6 Transconductance vs. cathode current density for the 417A triode.

respect to $V_{go}$. Thus

$$g_m = \frac{\partial I_{ao}}{\partial V_{go}} = Cn\left(V_{go} + \frac{V_{ao}}{\mu}\right)^{n-1} = Cn\left(\frac{I_{ao}}{C}\right)^{(n-1)/n} \qquad (5.2\text{-}14)$$

Since both $I_{ao}$ and $C$ are proportional to the cathode area $A$, $g_m$ is proportional to $A$, as we would expect. Furthermore $I_{ao} = JA$, where $J$ is the current density drawn from the cathode. Consequently

$$g_m \propto J^{(n-1)/n} \qquad (5.2\text{-}15)$$

For $n = 3/2$, $g_m \propto J^{1/3}$, and for $n = 2$, $g_m \propto J^{1/2}$. Figure 5.2-6 shows a plot of transconductance for the 417A triode vs. cathode current density. The normal operating point is marked with an $X$. In the neighborhood of this point it can be seen that $g_m$ is approximately proportional to $J^{1/2}$, whereas at lower current densities $g_m$ is proportional to a higher power of $J$.

The transconductance of a grid-controlled tube also increases as the distance between the grid and cathode is decreased, except at very small grid-to-cathode spacings, where Inselbildung effects become important. Figure 5.2-7 shows values of $g_m/J^{1/2}A$ vs. $d_{cg}$ for several Western Electric tubes with small grid-to-cathode spacings and for a current density of 0.02 amp/cm$^2$. For most of these tubes, $g_m$ is approximately proportional to $J^{1/2}$ at cathode current densities in the neighborhood of 0.02 amp/cm$^2$. The points appear to be distributed about a line with slope $-1$, indicating that $g_m \propto 1/d_{cg}$.
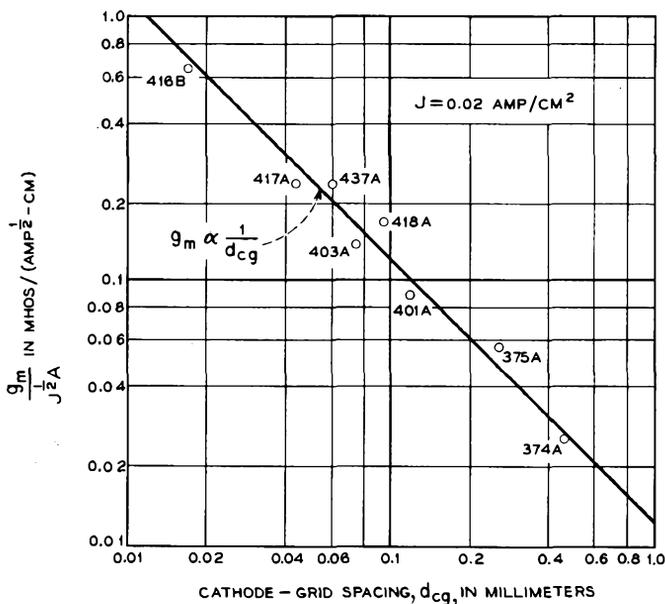
FIG. 5.2-7 Values of $G_m/J^{1/2}A$ plotted vs. $d_{cg}$ for several Western Electric tubes.

A theoretical upper limit to the transconductance that can be obtained from unit area of the cathode is reached when the grid is located at the potential minimum and when the grid wires and pitch are sufficiently small that the potential over the plane of the minimum is equal to that of the grid. From the discussion given in Section 2.4, it follows that the anode current under these circumstances is

$$I_{ao} = J_o A \epsilon^{eV_{go}/kT} \qquad (5.2\text{-}16)$$

where $J_o$ is the cathode emission current density, $A$ is the cathode area, and $V_{go}$ is the grid bias voltage, *a negative number*. By differentiating this with respect to $V_{go}$, we obtain

$$g_m = \frac{\partial I_{ao}}{\partial V_{go}} = J_o A \frac{e}{kT} \epsilon^{eV_{go}/kT} = \frac{eI_{ao}}{kT} \qquad (5.2\text{-}17)$$

Thus, in this theoretical upper limit of the transconductance of a grid-controlled tube, the transconductance is directly proportional to the current density $J = I_{ao}/A$ drawn from the cathode. (Note that $J_o$ is the cathode emission current density, and $J$ is the current density passing the potential minimum.) The ratio $g_m/I_{ao}$ is equal to $11,600/T$ mhos/amp.

Figure 5.2-8 shows a plot of transconductance vs. grid-to-cathode spacing for the Western Electric 416B triode, described in Chapter 7 (Section 7.4). The transconductance reaches a maximum of 0.075 mho when the grid-to-cathode spacing is about 0.012 mm, and at still smaller spacings the trans-
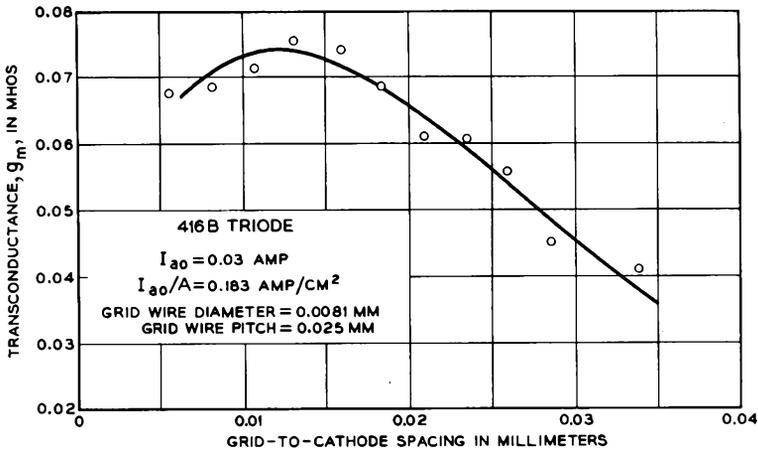


FIG. 5.2-8 Transconductance vs. grid-to-cathode spacing for the 416B triode described in Chapter 7.

conductance falls because of Inselbildung effects. The anode current for each of the experimental points was 0.03 amp, and the cathode temperature was close to 1025°K. Using these values in Equation (5.2-17), we find that the theoretical maximum transconductance is 0.34 mho. This is about 4.5 times the maximum observed transconductance, although the plane of the grid at maximum transconductance is approximately coincident with the plane of the potential minimum. It is probable that if a finer grid wire and smaller grid pitch were used, the theoretical maximum transconductance would be more nearly attained. However, the grid wire used in this tube is about as small as present technology permits.

In the 417A the grid is about three times as far from the cathode as the plane of the potential minimum. In consequence of this, the transconductance of the 417A is a still smaller fraction of the theoretical maximum. Typical values for the anode current and cathode temperature in the 417A are 0.027 amp and 1025°K. Substituting these values into Equation (5.2-17), we obtain a theoretical maximum transconductance of 0.31 mho, as compared with an actual transconductance of about 0.025 mho.

Since the plane of the potential minimum is always extremely close to the cathode, the control grid in most grid-controlled tubes is located well be-

yond the potential minimum. However in high-performance tubes, high transconductance is achieved by moving the grid closer to the potential minimum, generally at the expense of more difficult assembly procedures and higher cost.

Two more parameters that affect the transconductance are the grid-wire diameter and the grid pitch. At small grid-to-cathode spacings, the transconductance increases if both these quantities are reduced in such a manner that the screening fraction, or the ratio of the wire diameter to the pitch, is kept constant. An "open" grid structure, or small screening fraction, is usually desirable, since otherwise the beam current would be reduced, and the transconductance per unit area would be less. Often a pitch of between 4 and 10 times the grid-wire diameter is used. In many close-spaced, high-performance amplifier tubes the grid-wire diameter is chosen to be as small as is practicable from the standpoint of mechanical fabrication of the grid structure, and the grid pitch is then set to obtain maximum transconductance per unit area of the cathode, having due regard for limits imposed by the available cathode current density, the available electrode voltages, and the permissible electrode heat dissipation.

The following points will summarize our discussion about the transconductance of a triode:

1. The transconductance is directly proportional to the cathode area $A$.

2. The transconductance is proportional to a power of $J$ which is of the order of 1/3 to 1/2 in practical cases.

3. The transconductance for a given $J$ increases as the grid is moved closer to the cathode, until Inselbildung effects become important at very small grid-to-cathode spacings. When the grid is well beyond the potential minimum, the transconductance often varies approximately as $1/d_{cg}$.

4. At small grid-to-cathode spacings, the transconductance increases if the grid-wire diameter and pitch are reduced in such a manner that the screening fraction is kept constant.

The amplification factor is independent of the cathode area $A$, and it is almost independent of the cathode current density $J$ (see Figure 5.2-2). Like the transconductance, the amplification factor increases if the grid-wire diameter and pitch are reduced in such a manner that the screening fraction remains constant. The amplification factor also increases as the grid-to-anode distance is increased.

Figure 5.2-9 shows the values of $\mu$, $g_m$, and $r_a$ for a number of commercial and Western Electric grid-controlled tubes used in low-power amplifier applications. The code numbers of the Western Electric tubes begin with the numbers 3 and 4. Values of $\mu$ for triodes typically vary from 5 to 100, values of $g_m$ vary from 0.002 to 0.05 mho, and values of $r_a$ vary from $10^2$ to $10^4$ ohms.
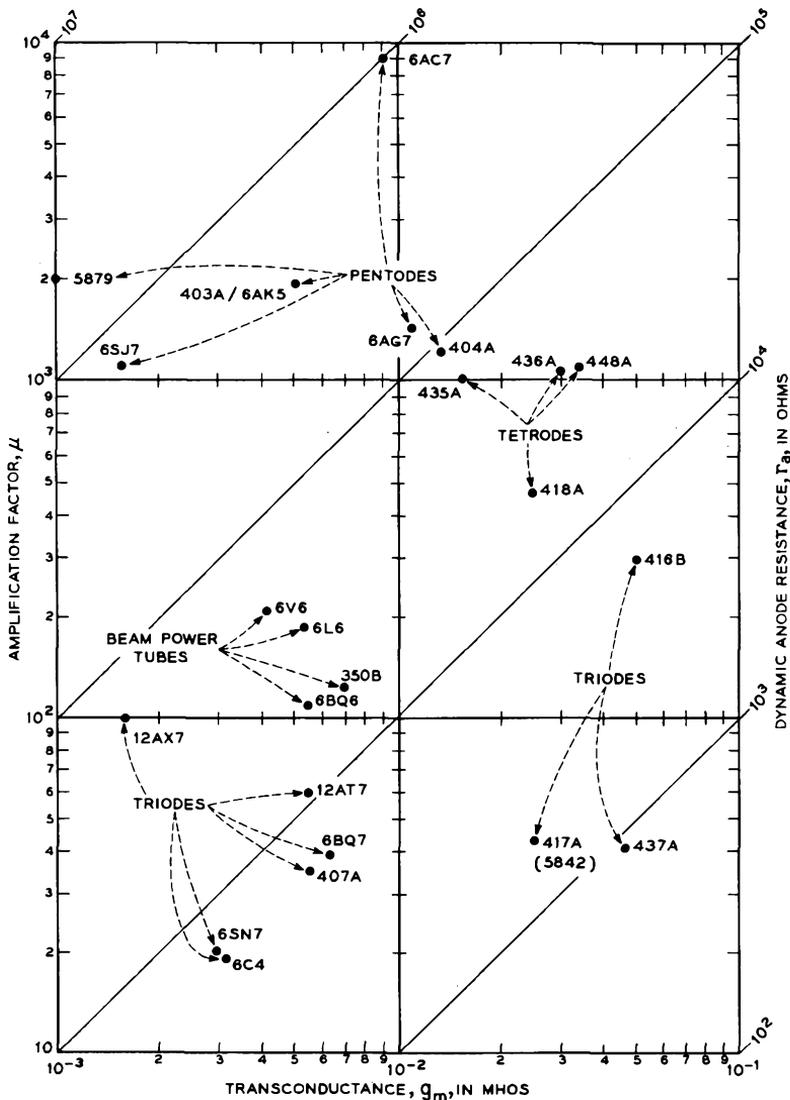
FIG. 5.2-9 The $\mu$, $g_m$, and $r_a$ of several commercial and Western Electric grid-controlled tubes used in low-power amplifier applications.

## 5.3   Tetrodes and Beam-Power Tubes

### (a)   *Tetrodes*

In the tetrode, a second grid, known as the screen grid, is inserted be-
tween the control grid and anode. The screen grid is usually of a coarser
mesh than the control grid and is operated at a fixed positive voltage with
respect to the cathode. It serves three principal functions:

1. It reduces the capacitance between the control grid and anode and
hence the coupling between the input circuit and the output circuit.

2. In Chapter 6 we show that the input capacitance of a grounded-
cathode triode amplifier stage is given by $C_{cg} + (1 + K)C_{ga}$, where $C_{cg}$ is
the cathode-to-grid capacitance, $C_{ga}$ is the grid-to-anode capacitance, and
$K$ is the voltage gain of the stage. Since $K$ may be a fairly large number, a
small capacitance between the grid and anode may cause considerable
shunting of the input signal at higher frequencies.[7] When a tetrode or
pentode tube is used in a grounded-cathode amplifier stage, the shunting
effect of the grid-to-anode capacitance is small, because $C_{ga}$ is small, and
consequently much higher input impedances are possible.

3. Over a range of positive anode voltages the current reaching the anode
of a tetrode is determined almost entirely by the voltages applied to the
control grid and screen grid and is nearly independent of the anode voltage.
This means that the dynamic anode resistance of the tetrode is high, and
this is an advantage in obtaining high gain per stage. In Chapter 6, the
gain of a simple amplifier stage without feedback is shown to be $g_m R_L/
(1 + R_L/r_a)$, where $R_L$ is the load resistance and $r_a$ is the dynamic anode
resistance. (See Equation (6.3-7).) Clearly, high $r_a$ is desirable where high
gain is needed.

Figure 5.3-1 shows a cross-sectional view of the Western Electric 448A
tetrode. Figure 5.3-1(c) shows a few of the grid wires and the relative
spacings of the electrodes. Notice that the distance from the cathode to
the screen grid in this tube is comparable with the distance from the cathode
to the anode in the 417A. (See Figure 5.1-1 for comparison.) The 448A
is used in a multistage amplifier which amplifies signals ranging in fre-
quency from 58 to 90 Mc.

---

[7]Certain grounded-cathode triode circuits provide for "neutralization" of the grid-
to-anode capacitance (Reference 5.4, p. 468), and much higher input impedances can
be obtained. However, these circuits usually require careful adjustment of the circuit
components in each stage if nearly complete neutralization is to be obtained. The
grounded-grid circuit shown in Figure 6.3-5 has a much smaller capacitance between
the anode and input circuit because the grid acts as an electrostatic shield. However,
this circuit has a relatively low input impedance, of the order of $1/g_m$ in parallel with
$R_K$. See Section 6.3.
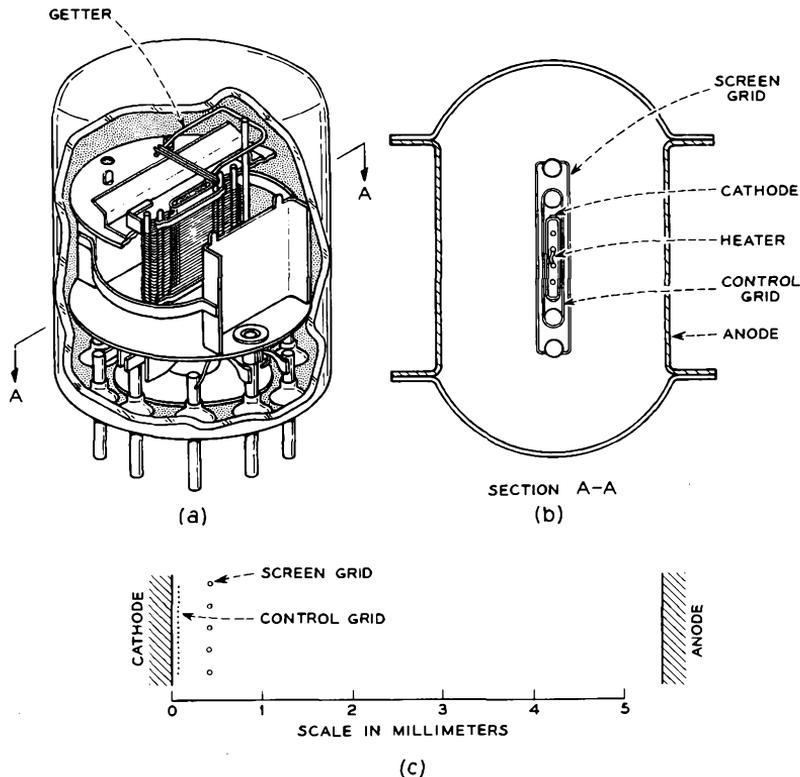
FIG. 5.3-1 The construction of the Western Electric 448A tetrode. The overall
height of the tube is 4.8 cm.

   The cathode of the 448A is a flattened nickel sleeve with a "double-
carbonate" oxide coating. The control-grid wires are made of tungsten
which is gold plated to raise its work function and reduce grid emission.
The screen grid is also made of tungsten, but the wires are coated with fine
carbon particles to increase heat radiation and reduce the operating tem-
perature of the wires. The anode is made of carbonized nickel. The carbon-
izing increases the heat radiation from the outer surface and reduces the
anode operating temperature. It also reduces secondary-electron emission
from the anode.
   Figure 5.3-2 shows the characteristic curves for the 448A. The screen-
grid voltage is indicated by $V_{sgo}$. Notice that the grid characteristics with
$V_{ao} = V_{sgo}$ are quite similar to the grid characteristics for a triode. The
anode characteristics show that $I_{ao}$ is nearly independent of $V_{ao}$ over a
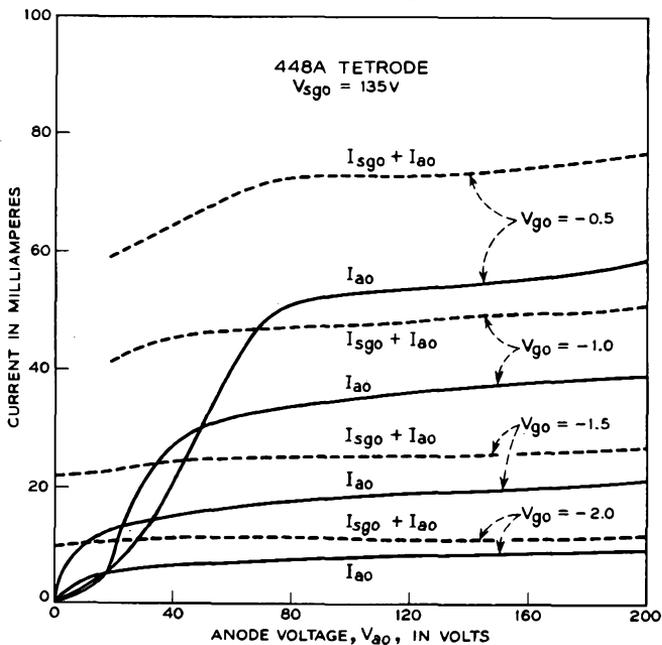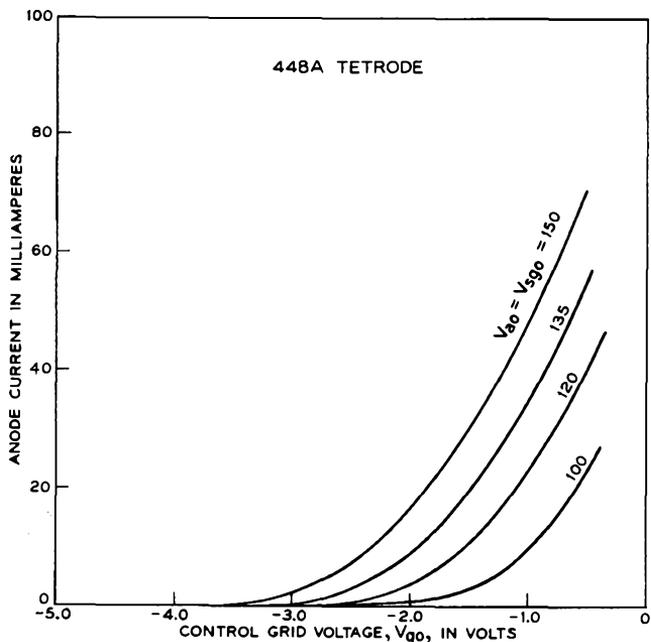
FIG. 5.3-2 The characteristic curves for the 448A tetrode.

range of positive anode voltages extending from well below screen-grid voltage to $+200$ volts or higher. Typical operating conditions for the 448A are given in Table 5.2-1.

Figure 5.3-1 shows that there is a large spacing between the screen grid and anode in the 448A. This serves two purposes:

1. With a large spacing, space charge between the screen grid and anode depresses the potential in the interelectrode space so that secondary electrons emitted from the anode are prevented from reaching the screen grid, and secondary electrons from the screen grid are prevented from reaching the anode. Without the potential depression, the current flowing in the anode circuit would be highly dependent on the relative voltages applied to the screen grid and anode, and the performance of the tube in an amplifier circuit would be seriously limited.

2. The large screen-anode spacing also reduces the output capacitance of the tube. The output capacitance is the capacitance between the anode and all other electrodes except the control grid. In Chapter 6 we shall find that a small output capacitance is desirable for tubes used in high-gain, broadband amplifiers.

Let us now look more closely at the potential between the screen grid and anode in the 448A. If we assume that the electron motion is normal to the plane of the electrodes and if edge effects are neglected, the potential between the screen grid and anode satisfies the one-dimensional form of Poisson's Equation,

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\varepsilon_o} = -\frac{J}{\varepsilon_o\sqrt{2\eta V}} \qquad (5.3\text{-}1)$$

where $J$ is the beam current density passing through the screen grid, and $x$ is the coordinate of the point at which $V$ is determined. The potential $V$ is measured relative to cathode potential, and the coordinate $x$ is measured in the direction normal to the electrodes. Solutions of this equation covering four ranges of anode voltages $V_{ao}$ are given in a paper by Fay, Samuel, and Shockley.[8] Figure 5.3-3 shows plots of these solutions for conditions which apply to the screen-anode space of the 448A. The four solutions of Fay, Samuel, and Shockley are described under separate headings below:

*Solution A.* $V_{ao} < 0$. In this case the potential decreases monotonically from the screen grid to the anode, but the shape of the potential is modified by the presence of space charge in the interelectrode space. At the point where the potential becomes negative, essentially all the electrons reverse
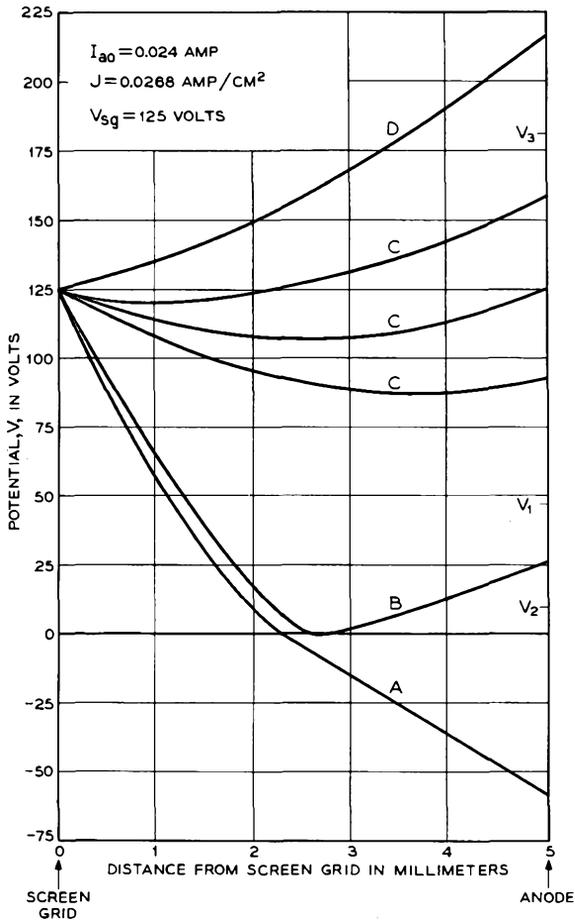
Fig. 5.3-3 Solutions of Equation (5.3-1) for a screen-grid voltage of 125 volts, a screen grid-anode spacing of 5 mm, and a beam current density of 0.0268 amp/cm². These are approximately the conditions that apply in the screen grid-anode space of the 448A tetrode under normal operating conditions.

the direction of their velocity and return to the screen grid. There is no space charge beyond this point. The anode current for this solution is, of course, zero.

*Solution B.* $0 \leq V_{ao} \leq V_1$. Here the anode voltage is positive, but a "virtual cathode" with potential $V = 0$ exists between the screen grid and anode. Since the electrons are emitted from the true cathode with a range

of velocities, the faster electrons pass the virtual cathode and reach the anode, while the remaining electrons reverse the direction of their velocity at the virtual cathode and return to the screen grid. The fraction of electrons passing the virtual cathode is determined by the applied anode voltage. When $V_{ao} = 0$, essentially all the electrons are returned to the screen grid. As $V_{ao}$ is increased from zero, the fraction of electrons passing the virtual cathode increases until at some voltage $V_1$, all the electrons pass the virtual cathode, and the potential distribution suddenly changes to that of Solution C described below. (Note that $0 \leq V_{ao} \leq V_1$ is a necessary but not sufficient condition for Solution B to prevail, as discussed below.)

*Solution C.* $V_2 \leq V_{ao} \leq V_3$. Here the anode voltage is positive, and space charge between the electrodes depresses the potential in the interelectrode space so that there is a plane of minimum potential at some point between the electrodes. The potential at the minimum is greater than zero and less than either the screen or anode potential. All the electrons passing through the screen grid reach the anode when this solution prevails.

If the anode voltage is increased from zero through the voltage $V_1$, the potential distribution changes abruptly from the "virtual-cathode" distribution of Solution B to the "potential-minimum" distribution of Solution C. If the anode voltage is then lowered through the voltage $V_1$, Solution C prevails until some lower voltage $V_2$ is reached. Lowering the anode voltage still further causes the potential distribution to change abruptly to the virtual-cathode solution. This "hysteresis effect" in which there are two possible solutions to the potential distribution in the interelectrode space for anode voltages between $V_1$ and $V_2$ can be observed in a number of tetrode vacuum tubes. Of course, the anode current is less than the full beam current when Solution B prevails.

The position of the plane of minimum potential of Solution C moves closer to the screen grid as the anode voltage is increased. When $V_{ao} = V_3$, the plane of the potential minimum coincides with that of the screen grid.

*Solution D.* $V_{ao} > V_3$. Here the potential increases monotonically from the screen grid to the anode, and all the electrons reach the anode, as in Solution C.

In normal operation of the 448A, Solution C prevails. From Figure 5.3-3 it can be seen that, when $V_{sgo} = V_{ao} = 125$ volts, the potential minimum is about 18 volts below the screen and anode voltage. This potential depression greatly reduces the exchange of secondary electrons between the anode and the screen grid.

There are several sources of error involved in the use of Equation (5.3-1) and the solutions of Fay, Samuel, and Shockley to obtain the potential in the interelectrode space of a tube such as the 448A. First, edge effects are

neglected. Second, many of the electrons that pass close to the screen-grid wires as they enter the screen-anode space are deflected by strong local fields close to the screen-grid wires, so that the motion of these electrons is certainly not entirely in the $x$ direction. Finally, from Figure 5.3-1(b) it is evident that the spacing between the screen grid and anode of the 448A is comparable with the linear dimensions of the cathode of the 448A, and consequently we would expect that the beam would spread as it travels between the screen grid and anode.

### (b)  Beam-Power Tubes

A second class of screen-grid tubes, known as beam-power tubes, also
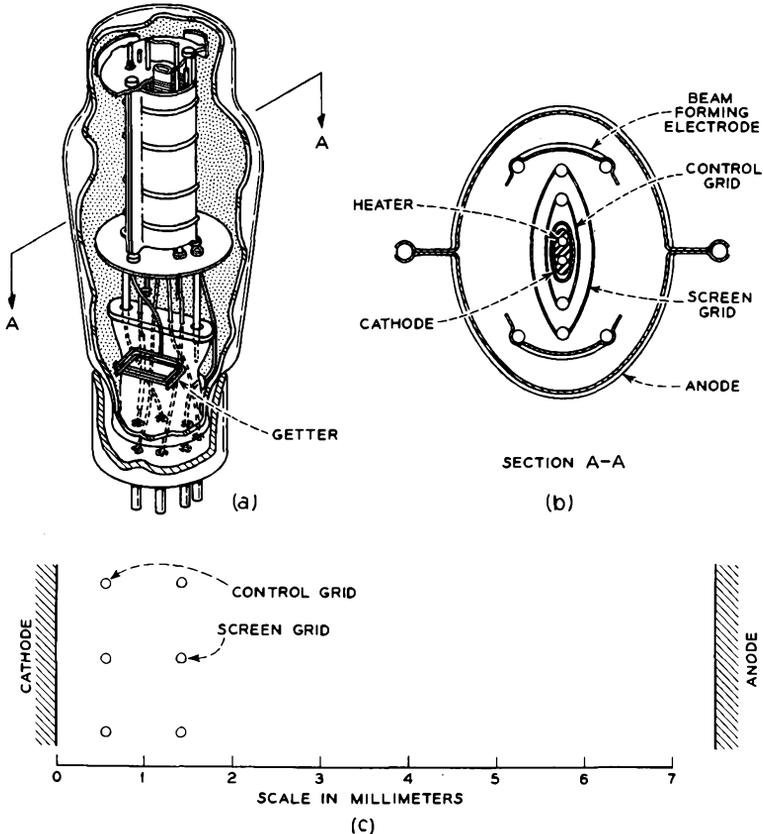


FIG. 5.3-4 The construction of the Western Electric 350B beam-power tube. The overall height of the tube is 14 cm.

makes use of a large screen-anode spacing so that the space charge of the electron beam depresses the potential between the screen grid and anode, and secondary electrons from one electrode are prevented from reaching the other. Figure 5.3-4 shows the construction of the Western Electric 350B beam-power tube. Notice that each screen-grid wire is directly opposite a control-grid wire. This construction greatly reduces the interception of the electron beam by the screen grid, since the negative bias on the control grid causes a "shadowing" of the screen-grid wires from the beam. In beam-power tubes an additional electrode, called a "beam-forming electrode," is located near the edge of the beam and held at cathode potential. This electrode also helps to depress the potential between the
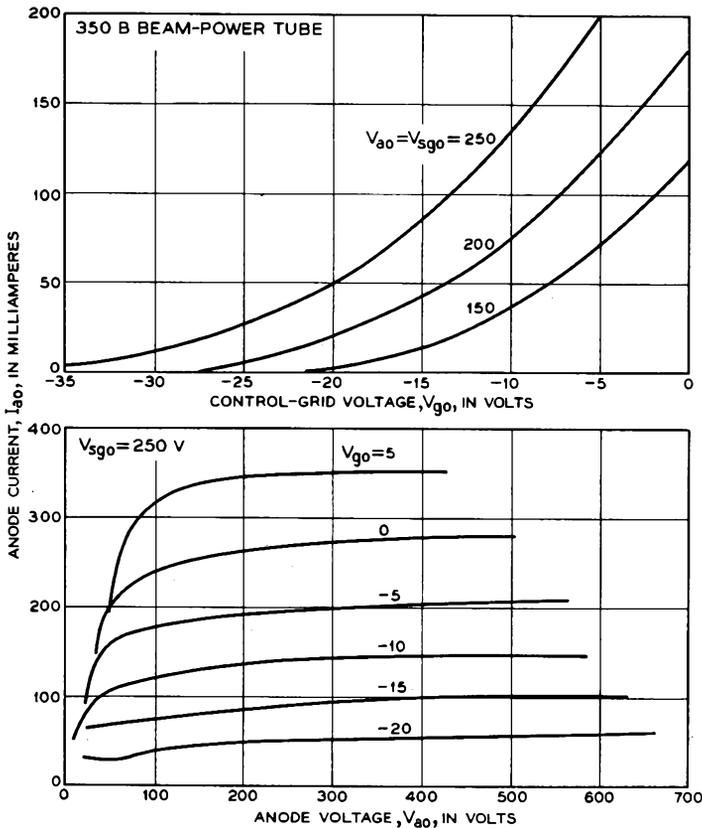


FIG. 5.3-5 The grid and anode characteristic curves for the Western Electric 350B beam-power tube.

screen grid and anode, but the principal cause of the potential depression in the region of the beam still arises from the space charge of the electron beam. The beam-forming electrode also prevents secondary electrons liberated from the anode from reaching the screen grid by means of paths outside the incident beam.

Typical operating conditions for the 350B are given in Table 5.2-1. The grid and anode characteristic curves for the 350B are shown in Figure 5.3-5. Note that the ratio of anode to screen-grid current is 8.9:1 for the 350B, whereas the same ratio for the 448A is 2.8:1.

The transconductance of a tetrode or beam-power tube is determined largely by the electrode geometry in the region between the cathode and the screen grid. However, it is reduced by the division of beam current between the screen grid and the anode. Often the interception of the beam current by the screen grid reduces the transconductance by 10 to 30 per cent over what it would be if the screen grid and anode were connected. The transconductance of the 350B is much lower than that of the 448A, principally because the 350B has a relatively large spacing between the cathode and control grid, and the distance between the control-grid wires is larger.

Figure 5.2-9 shows the $\mu$, $g_m$, and $r_a$ of a number of tetrodes and beam-power tubes. The transconductances of tetrodes and beam-power tubes fall in about the same range as those of triodes. However, the amplification factors of tetrodes are about an order of magnitude greater than those of triodes because of the shielding action of the screen grid.

## 5.4  Pentodes

Still another approach to the problem of eliminating the exchange of secondary electrons between the anode and the screen grid is that used in the pentode. Here a third grid, known as the suppressor grid, is inserted between the screen grid and anode. The suppressor grid is usually biased at cathode potential and therefore does not intercept any of the beam. Its pitch, or center-to-center wire spacing, is large, so that the potential at mid-point between grid wires is always well above cathode potential. In this way, most of the electrons that pass through the screen grid also pass through openings between suppressor-grid wires and travel on to strike the anode. However, the suppressor grid causes sufficient depression of the potential between the screen grid and the anode that it stops virtually all exchange of secondary electrons between these electrodes.

The pentode is by far the most widely used grid-controlled tube. Its advantages include high gain per stage and low grid-to-anode capacitance.

The construction of the Western Electric 403A pentode is illustrated in Figure 5.4-1. The electrode structure of this tube is similar to that of the 6AK5, a commercial code, and we shall refer to the tube as the 403A/6AK5 in subsequent discussion. The suppressor grid is located relatively close to
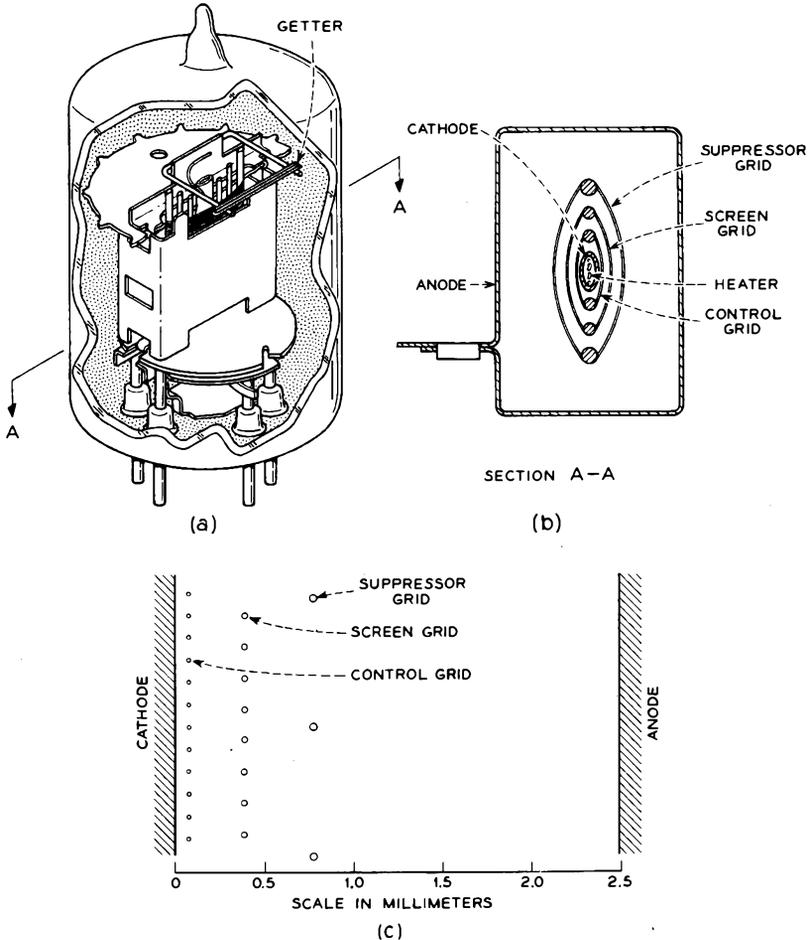


FIG. 5.4-1 The construction of the 403A/6AK5 pentode. The overall height of the tube is 4.4 cm.

the screen grid to reduce the output capacitance. Figure 5.4-2 shows the characteristic curves of the 403A/6AK5. Typical operating conditions are given in Table 5.2-1.
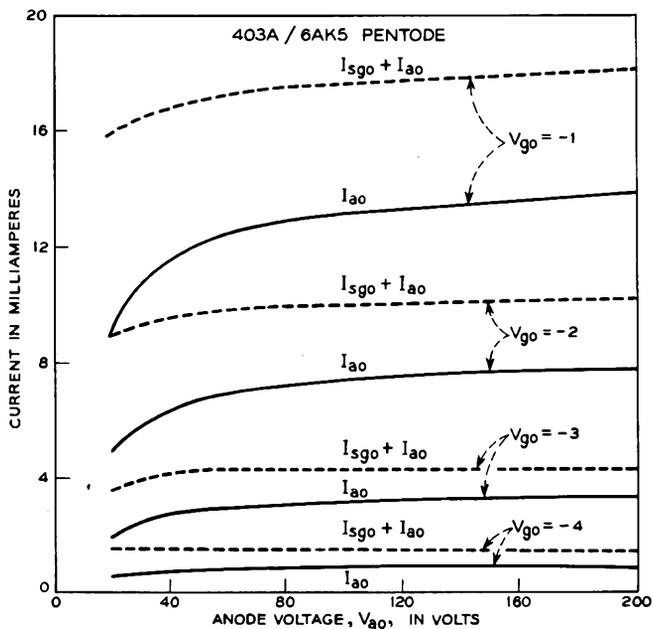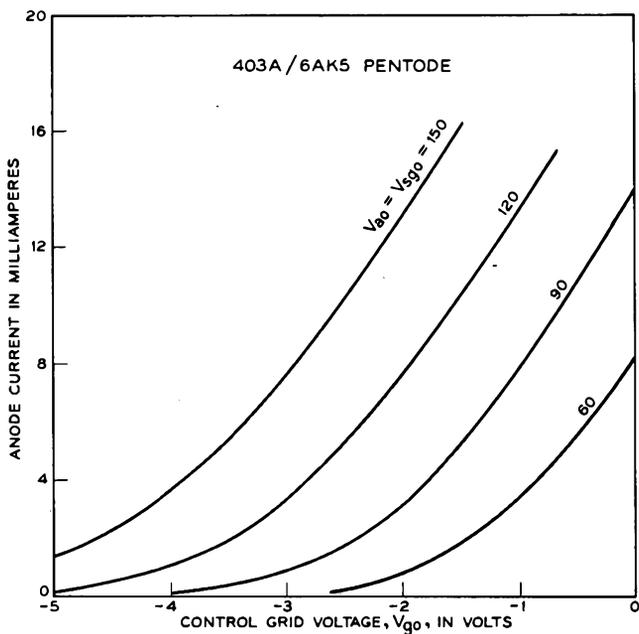
FIG. 5.4-2 The characteristic curves for the 403A/6AK5 pentode.

As in the tetrode, the transconductance of a pentode is determined largely by the electrode geometry in the region between the cathode and the screen grid and is reduced by the division of beam current between the screen grid and anode. Pentode transconductances are therefore of about the same magnitude as those of triodes and tetrodes. However, the amplification factor and dynamic anode resistance of a pentode tend to be even higher than for a tetrode because the suppressor grid provides added shielding between the anode and cathode. Pentode amplification factors usually lie between $10^3$ and $10^4$, whereas the dynamic anode resistance usually falls between 0.1 and a few megohms.

## PROBLEMS

5.1 Show that the electrostatic amplification factor $\mu_{es}$ of a planar triode is given by

$$\mu_{es} = \frac{C_{cg}}{C_{ca}}$$

where $C_{cg}$ is the capacitance between the grid wires and the cathode, but does not include the capacitance between the leads and electrode supports, and $C_{ca}$ is the capacitance between the cathode and anode electrodes, but does not include the capacitance between the leads and electrode supports.

5.2 Equation (5.1-3) indicates that the electrostatic amplification factor $\mu_{es}$ depends upon the grid-wire diameter, the grid pitch, and the grid-to-anode spacing. However, it is nearly independent of the cathode-to-grid spacing $d_{cg}$, provided $d_{cg}$ is large enough that the electric field at the surface of the cathode is uniform over the cathode surface. Can you explain qualitatively why $\mu_{es}$ should be nearly independent of $d_{cg}$?

5.3 An amplifier is constructed with two triodes in parallel. For the particular set of voltages applied to the electrodes, the dynamic characteristics of one tube are $g_{m1}$, $r_{a1}$, and $\mu_1$, and those of the other tube are $g_{m2}$, $r_{a2}$, and $\mu_2$. What are the $g_m$, $r_a$, and $\mu$ of the parallel combination?

5.4 A particular beam-power tube has a dynamic anode resistance of 200,000 ohms at a control-grid voltage of $V_{go} = V_1$, a screen-grid voltage $V_{sgo} = V_2$, and an anode voltage $V_{ao} = V_2$. If the screen grid is connected to the anode and the tube is operated as a triode, the dynamic anode resistance is 10,000 ohms with $V_{go} = V_1$ and $V_{ao} = V_{sgo} = V_2$. For a particular application the screen grid is used as the control electrode and the control grid is maintained at a constant bias voltage $V_1$. The screen grid and anode operating voltages are $V_2$. What is the transconductance of the tube using the screen grid as control electrode under these conditions? Assume that the screen-grid wires are ideally shielded behind the control-grid wires so that there is essentially zero current intercepted by the screen grid. Note that, for incremental variations in the applied potentials, the incremental change in anode current can be expressed as

$$dI_{ao} = \frac{\partial I_{ao}}{\partial V_{go}}dV_{go} + \frac{\partial I_{ao}}{\partial V_{sgo}}dV_{sgo} + \frac{\partial I_{ao}}{\partial V_{ao}}dV_{ao}$$

## REFERENCES

Three general references covering the principles and design of grid-controlled tubes are:

5a. K. R. Spangenberg, *Vacuum Tubes*, Chapters 7 to 11, McGraw-Hill Book Co., Inc., New York, 1948.

5b. W. G. Dow, *Fundamentals of Engineering Electronics*, 2nd Ed., Chapters 4 to 6, John Wiley and Son, Inc., New York, 1952.

5c. A. H. W. Beck, *Thermionic Valves*, Chapters 9 to 11, Cambridge University Press, Cambridge, England, 1953.

Other references covering specific topics referred to in this chapter are:

5.1 W. R. Bennett and L. C. Peterson, *Bell System Tech. J.* **28**, 303, 1949.

5.2 I. Langmuir and K. T. Compton, *Rev. Modern Phys.* **3**, 241, 1931.

5.3 P.H.J.A. Kleynen, *Philips Res. Reports* **1**, 81, 1946.

5.4 F. E. Terman, *Radio Engineer's Handbook*, McGraw-Hill Book Co., Inc., New York, 1943.

5.5 C. E. Fay, A. L. Samuel, W. Shockley, *Bell System Tech. J.* **17**, 49, 1938.

# Chapter 6

## INDUCED CURRENTS, EQUIVALENT NETWORKS, AND GAIN-BANDWIDTH PRODUCT

In the application of electron tubes as circuit elements we are interested in the currents that flow in the circuits external to the tube as a result of electron motion in the space between the electrodes. In this chapter we show how the external currents are related to the electron currents in the interelectrode space of a tube for cases in which the period of the ac voltages applied to the electrodes is long compared with the time taken by the electrons to travel between the electrodes.

When an electron is gaining kinetic energy under the influence of an electric field, the source that provides the field loses an equal amount of energy. Similarly, when an electron is slowed down by an electric field, an amount of energy equal to the kinetic energy lost by the electron appears elsewhere in the system. Several problems that illustrate these effects are discussed in Sections 6.1 and 6.2.

In Section 6.3 we consider the small-signal analysis of simple circuits using grid-controlled tubes. We shall find that a grid-controlled tube operated with negative bias on the control grid and driven by a small ac signal can be simulated by either of two networks. One network contains a constant current generator and passive elements, such as resistances, capacitances, and inductances, whereas the second contains a constant voltage generator and passive elements. In analyzing the small-signal behavior of an amplifier stage, these networks can be substituted in place of the tube, and the currents that flow in the various circuit elements can be determined by simple application of Kirchhoff's Laws.

By applying the networks to the analysis of an amplifier stage which is part of a multistage amplifier, we find that the product of gain and bandwidth that can be obtained from the stage is a constant that depends only on parameters of the tube itself and on the capacitances that shunt the

input and output circuit. Maximum possible gain-bandwidth product would be obtained if all external capacitances shunting the input and output circuits were reduced to zero. The expression for this maximum gain-bandwidth product provides a useful figure of merit for comparing tubes to be used in high-gain, broadband amplifiers. High figure-of-merit tubes have a high transconductance and low input and output capacitances.

## 6.1  Induced Currents Resulting from the Motion of Charge Between Electrodes

When an electron travels near a conductor, surface charges are induced on the conductor so that the electric field within the conductor is zero at all times. As the electron moves, the surface charges rearrange themselves to maintain zero field within the conductor. If several conductors are present in the region, and if the conductors are insulated from each other, their potentials vary with the motion of the electron. However, if two of the conductors are joined by a wire, the potential difference between them remains zero for all motions of the electron, and in general this condition can be satisfied only by a flow of charge along the wire joining the conductors.

Suppose an electron is very near one of two conductors that are joined by a wire. Practically all the lines of electric field arriving at the electron originate on positive charges on the surface of the nearby conductor. A surface charge distribution of total charge $+e$ is therefore induced on the nearby conductor. A similar situation exists when the electron is very near the other conductor. It follows, therefore, that motion of the electron from a point very near the first conductor to a point very near the second must be accompanied by a flow of charge $+e$ from the first conductor to the second through the wire joining them.

Figure 6.1-1 shows an electron moving in the region between two conductors $A$ and $B$. A battery maintains conductor $B$ at a potential $V_o$ volts above that of conductor $A$. If the electron travels a distance $\Delta x$ in the direction of conductor $B$ and the potential rise over the distance $\Delta x$ is $\Delta V$ volts, the kinetic energy of the electron is increased by $e\Delta V$
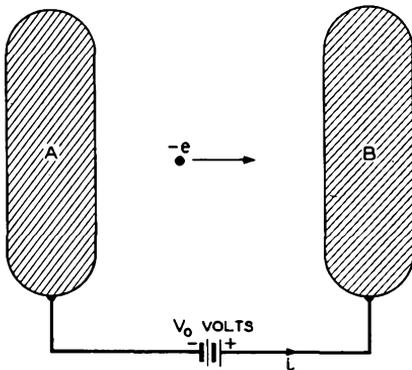


FIG. 6.1-1 An electron moving between two conductors.

joules, where $-e$ is the charge on the electron. This increase in kinetic energy must have come from the battery. We conclude that an amount of positive charge has flowed from conductor $A$ to conductor $B$ through the battery, and that the work done by the battery on the charge is $e\Delta V$. Let the amount of charge that flowed from $A$ to $B$ be $\Delta q$. Then

$$\Delta q V_o = e\Delta V \tag{6.1-1}$$

or

$$\Delta q = \frac{e}{V_o}\Delta V \tag{6.1-2}$$

Let $\Delta t$ be the time taken by the electron in traveling the distance $\Delta x$. Dividing both sides of Equation (6.1-2) by $\Delta t$, we obtain

$$\frac{\Delta q}{\Delta t} = \frac{e}{V_o}\frac{\Delta V}{\Delta t} \tag{6.1-3}$$

This can be expressed in differential form as

$$i = \frac{dq}{dt} = \frac{e}{V_o}\left[\frac{\partial V}{\partial x}\frac{dx}{dt} + \frac{\partial V}{\partial y}\frac{dy}{dt} + \frac{\partial V}{\partial z}\frac{dz}{dt}\right] = -\frac{e}{V_o}\mathbf{E}\cdot\mathbf{u} \tag{6.1-4}$$

where $i$ is the current that flows in the external circuit joining $A$ and $B$, $-e$ is the charge on the electron, $\mathbf{E}$ is the electric field acting on the electron, and $\mathbf{u}$ is the electron velocity. We notice that $\mathbf{E}$ is proportional to $V_o$, so that, *for a given electron velocity* $\mathbf{u}$, *the current $i$ is independent of the voltage applied between the electrodes.* Let us set $-\mathbf{E}(x,y,z)/V_o = \mathbf{E}_1(x,y,z)$, where $\mathbf{E}_1$ is a vector function of position having the dimensions of meters$^{-1}$ and equal in magnitude and direction to the electric field obtained when conductor $B$ is held 1 volt *negative* with respect to conductor $A$. Equation (6.1-4) can then be expressed as

$$i = e\mathbf{E}_1\cdot\mathbf{u} \tag{6.1-5}$$
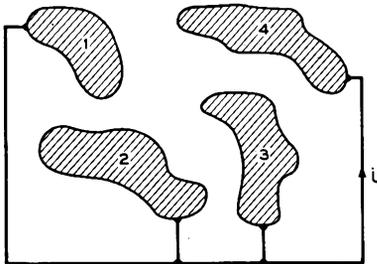


FIG. 6.1-2 Four arbitrarily shaped electrodes joined by a wire.
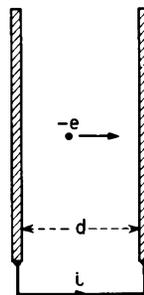
FIG. 6.1-3 An electron moving between two planar electrodes.

Since $i$ is independent of $V_o$, *this must be the current that flows when the battery is omitted and the conductors are joined only by a wire or an impedance.*

Equation (6.1-5) can be applied to a system of several conductors, such as that illustrated in Figure 6.1-2. If $i$ is the current flowing toward conductor 4, then $\mathbf{E}_1$ is a vector field having the same magnitude and direction as the electric field obtained when conductor 4 is held at a potential of $-1$ volt, and the remaining conductors are at ground potential.

Of particular interest is the problem of an electron moving between two planar electrodes whose linear dimensions are large compared with their spacing. Two such electrodes are illustrated in Figure 6.1-3. The electrodes are connected by a wire and are spaced by a distance of $d$ meters. Neglecting edge effects, the vector function $\mathbf{E}_1$ is equal in magnitude to $1/d$ at all points between the electrodes, and its direction is normal to the plane of the electrodes. If the velocity of the electron is also normal to the electrodes, the current flowing in the wire is given by

$$i = \frac{eu}{d} \tag{6.1-6}$$

This result also can be obtained directly from Equation (6.1-2). For the case of the planar electrodes Equation (6.1-2) can be written as

$$\Delta q = e\frac{\Delta V}{V_o} = e\frac{\Delta x}{d} \tag{6.1-7}$$

Dividing both sides by $\Delta t$ and taking the limit as $\Delta t \to 0$, Equation (6.1-6) is obtained.

If many electrons are present between the electrodes shown in Figure 6.1-1 and if they produce a charge density $\rho$, it follows from Equation (6.1-5) that the current in the external circuit is given by[1]

$$i = -\int_{\text{volume}} \rho\mathbf{E}_1 \cdot \mathbf{u}\, dx dy dz$$

$$= -\int_{\text{volume}} \mathbf{J} \cdot \mathbf{E}_1 dx dy dz \tag{6.1-8}$$

where $\mathbf{J} = \rho\mathbf{u}$ is the current density at the volume element $dx dy dz$, and the integral is taken over the region occupied by the space charge. This result also applies if the electrodes are joined by a wire or an impedance instead of the battery.

Next let us consider the currents that flow in the wires joining the elec-

---

[1]If $\rho$ is a positive charge and moving toward electrode $B$, the incremental induced current $\rho\mathbf{E}_1 \cdot \mathbf{u}\ dx dy dz$ flows *away* from electrode $B$.

trodes shown in Figure 6.1-4(a) when a single electron travels from electrode $A$ to electrode $B$. We shall assume that the electron starts from rest at electrode $A$ at time $t_o$. The field between the grid and electrode $A$ accelerates the electron toward the grid, and at time $t_1$ it passes through an



(a)                                (b)
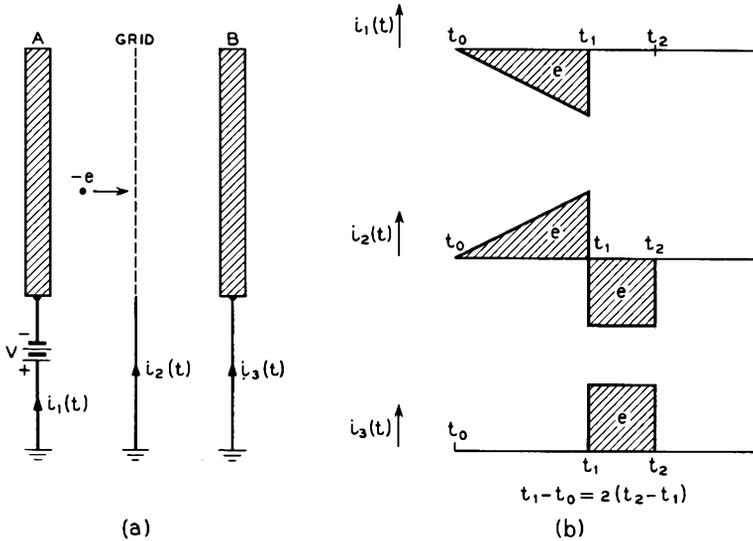
FIG. 6.1-4 Induced currents which flow in the external circuit when an electron travels from electrode $A$ to electrode $B$. It is assumed that the electron starts at rest from electrode $A$.

opening in the grid. The electron subsequently moves with constant velocity through the second region and strikes electrode $B$ at time $t_2$. In Figure 6.1-4(b) the induced currents that flow from ground toward the electrodes are plotted as functions of time. When the electron is between electrode $A$ and the grid, it experiences a uniform accelerating field, so that its velocity and the induced current, $eu/d$, increase uniformly with time. To the right of the grid, where the electron travels at a steady velocity, the induced current is constant with time. The area under each of the shaded regions in Figure 6.1-4(b) is equal to the electronic charge.

While the electron is traversing the distance between electrode $A$ and the grid, a total charge of $+e$ flows through the battery from the negative terminal to the positive terminal. This means that the battery expends $eV$ joules of work. The work is imparted to the electron in the form of kinetic energy by the field between electrode $A$ and the grid. When the electron strikes electrode $B$, its kinetic energy is dissipated in the electrode

as heat. Thus the energy expended by the battery is turned into heat energy in electrode $B$. The positive charge that flows through the external circuit to electrode $B$ is cancelled when the electron strikes the electrode.

## 6.2   Currents Induced in External Impedances

Let us consider the induced currents that flow in an external resistance connected between two electrodes. Figure 6.2-1 shows a beam of electrons which passes through a grid and strikes electrode $B$. Each electron is assumed to pass through the grid with the same kinetic energy. If $n$ electrons pass through the grid per second, the same number strike electrode $B$ per second, and a current $I_o = ne$ amperes flows from ground through the resistance to meet the arriving electrons. This current causes a voltage drop of $I_oR$ volts in the resistance. (We assume that the voltage drop is less than the voltage through which the electrons have been accelerated, so that the electrons are not stopped before reaching electrode $B$.) The power dissipated in the resistance by the flow of charge is $I_o^2R$ watts.



BEAM OF ELECTRONS

GRID

CATHODE

B

ELECTRONS PASS THROUGH GRID WITH ENERGY OF eV JOULES

V

R

FIG. 6.2-1 Two electrodes connected by a resistance.
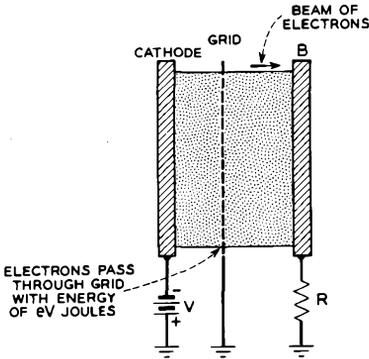
Because there is a voltage drop in the resistance, each electron faces a decelerating field and loses kinetic energy $eI_oR$ joules in traveling from the grid to electrode $B$. Since $n$ electrons lose this amount of kinetic energy per second, the total power lost by the electrons in traversing the region is $neI_oR = I_o^2R$ watts. But this is the power dissipated in the resistance by the positive charges flowing to meet the electrons. *Hence the kinetic energy lost by the electrons while traveling from the grid to electrode $B$ is transformed into heat energy which is dissipated in the resistance $R$.* When the electrons strike electrode $B$, the remaining part of their kinetic energy is transformed into heat energy in electrode $B$.

Next let us suppose that the number of electrons passing the grid per unit time can be varied without changing their velocity. Let the current of electrons passing through the grid be given by $i = I_o + I_1\cos \omega t$. An ac voltage $I_1R \cos \omega t$ appears across the resistance, and the total instantaneous power developed in the resistance is

$$i^2R = (I_o + I_1 \cos \omega t)^2R \text{ watts} \qquad (6.2\text{-}1)$$

When this is averaged over time, we obtain

$$P_{av} = I_o{}^2 R + \frac{I_1{}^2 R}{2} \text{ watts} \qquad (6.2\text{-}2)$$

Evidently this power is greater than the dc power developed in the resistance in the absence of modulation. However, the average number of electrons passing through the grid per second is the same as in the dc case, and each has the same kinetic energy as it passes through the grid. Hence the average energy of the electrons when they strike electrode $B$ must be less when the beam is modulated than in the dc case. To explain this, we may note that during the half cycle in which the total beam current is greater than $I_o$, the number of electrons passing the grid is greater than average. The voltage drop in the resistance and the retarding field are also greater than average during this half cycle. Consequently, more than half the electrons lose more kinetic energy than under dc conditions. During the other half of the cycle, less than half the electrons lose less kinetic energy than under dc conditions. Thus the ac power represented by the term $I_1{}^2 R / 2$ in Equation (6.2-2) is also obtained at the expense of the kinetic energy of the electrons.

It is easy to extend our considerations to include impedances in the external circuit. If the resistance $R$ in the above example were replaced by an impedance $Z$, the instantaneous voltage across the impedance and hence the instantaneous voltage across the interelectrode space would be $\text{Re}[(I_o + I_1 \epsilon^{j\omega t}) Z]$, where $i = \text{Re}(I_o + I_1 \epsilon^{j\omega t})$ is the instantaneous current of electrons crossing between the electrodes.

It is important to emphasize at this point that the voltage developed across the impedance is not caused by the electrons that strike electrode $B$ flowing through the impedance. The results of Section 6.1 showed that *the induced currents flow in the external circuit only while the charge is crossing between the electrodes.* When the individual electrons strike electrode $B$, they cancel positive charges that have flowed to meet them.

That the induced current flowing through the impedance in the external circuit is independent of the size of the impedance is indeed a very important result. The induced current is determined only by the current of electrons crossing between the electrodes and is equal to that current. If the beam is modulated, the ac power developed in the resistive part of the load increases linearly with this resistance. Consequently, if we can modulate the beam in a manner that consumes very little power, we have a means for amplifying ac power. The ac power output, of course, is obtained at the expense of the supplies that provide the dc voltage to accelerate the electrons.

In the remainder of this chapter we shall describe the ac operation of grid-controlled tubes, and in later chapters we shall describe klystron

amplifiers and traveling-wave amplifiers. We shall find that, although these tubes differ in the means used to modulate the beam and the type of load in which the ac power is developed, each provides ac power amplification by modulating an electron beam and causing ac induced currents to flow in external impedances. In klystron amplifiers and traveling-wave amplifiers the energy provided by the dc supply is first converted into kinetic energy of the electrons. This energy in turn is partly converted into ac power which is dissipated in the load and losses of the system and partly into heat energy of the electrode struck by the electrons. In grid-controlled tubes employing screen grids, a similar energy transfer occurs when the anode and screen grid are connected to the same dc supply voltage. (The screen grid and anode connections in this case would be similar to those illustrated in Figure 6.2-1.)
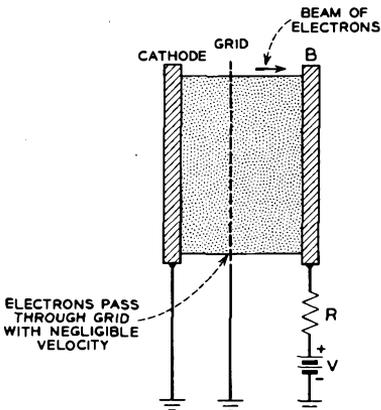


FIG. 6.2-2 Two electrodes connected by a dc supply and a resistance in series.

However, if the dc supply providing the field that accelerates the electrons is connected in series with the external impedance, as in the case of the triode tube, the transfer of power is somewhat different. In Figure 6.2-2 we show the electrodes of the previous example with a battery of $V$ volts connected in series with the external resistance. The electrons in this case are assumed to pass through the grid with negligible velocity and are accelerated toward electrode $B$ by the field provided by the battery. In this case the induced currents flowing through the battery and the load resistance cause some of the power expended by the battery to be transferred directly to the load resistance. The remaining power expended by the battery is converted into kinetic energy of the electrons, which in turn becomes heat energy of electrode $B$. If the instantaneous beam current is given by $i = I_o + I_1 \cos \omega t$, the average power developed in the resistance is $I_o^2R + I_1^2R/2$, just as in the previous example. The average power expended by the battery is $I_oV$, and the average power dissipated in electrode $B$ is $I_oV - \left( I_o^2R + \dfrac{I_1^2R}{2} \right).$

## 6.3  Equivalent Networks

Consider a grid-controlled tube in which the control-grid voltage and anode voltage are varied, while the potentials applied to the remaining

electrodes are held constant. The current reaching the anode can be expressed as

$$I_{ao} = I_{ao}(V_{go}, V_{ao}) \tag{6.3-1}$$

where $V_{go}$ and $V_{ao}$ are the control-grid and anode voltages. Generally this function is single-valued and continuous. If $I_{ao}$ undergoes a differential change because of differential changes in $V_{go}$ and $V_{ao}$, then

$$dI_{ao} = \frac{\partial I_{ao}}{\partial V_{go}} dV_{go} + \frac{\partial I_{ao}}{\partial V_{ao}} dV_{ao} \tag{6.3-2}$$

It follows that if very small ac voltages $v_g$ and $v_a$ are applied to the control grid and anode, the induced ac current flowing in the anode circuit will be

$$i_a = \frac{\partial I_{ao}}{\partial V_{go}} v_g + \frac{\partial I_{ao}}{\partial V_{ao}} v_a \tag{6.3-3}$$

In terms of the tube parameters $g_m$ and $r_a$ discussed in Chapter 5, Equation (6.3-3) can be rewritten as

$$i_a = g_m v_g + \frac{1}{r_a} v_a \tag{6.3-4}$$

This important equation gives the induced ac current $i_a$ which flows in the anode circuit when small ac voltages $v_g$ and $v_a$ are applied to the grid and anode.

Let us now consider two simple networks which we shall show to be described by Equation (6.3-4) and which can be used to simulate the tube
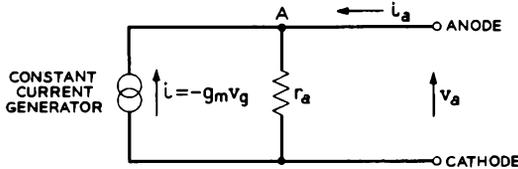


FIG. 6.3-1 The constant-current-generator small-signal equivalent network for a grid-controlled tube.

for network analysis. The first of these is illustrated in Figure 6.3-1. It involves a constant current generator,[2] which generates a current $-g_m v_g$, in parallel with a resistance equal to the dynamic anode resistance $r_a$. Referring to the figure, it is evident that the current flowing through the resistance $r_a$ away from the point $A$ is $-g_m v_g + i_a$. This current, multi-

---

[2] The symbol used in Figure 6.3-1 for a constant current generator will be used in subsequent illustrations in this and later chapters. Likewise, the symbol used in Figure 6.3-3 for a constant voltage generator will be used in subsequent illustrations in this and later chapters.

plied by the resistance $r_a$, must equal the voltage $v_a$ applied between the terminals. Hence

$$v_a = (-g_m v_g + i_a)r_a \qquad (6.3\text{-}5)$$

A simple rearrangement of this equation shows that it is just another form of Equation (6.3-4). Thus, when a voltage $v_a$ is applied between the terminals of the network shown in Figure 6.3-1, a current $i_a$, given by Equation (6.3-4), flows through the network from one terminal to the other. The same current flows in the anode lead of a grid-controlled tube when ac voltages $v_g$ and $v_a$ are applied to the control grid and anode. Consequently, the network shown in Figure 6.3-1 can be used in place of the tube for purposes of network analysis. Figure 6.3-2(b) shows the construction of an
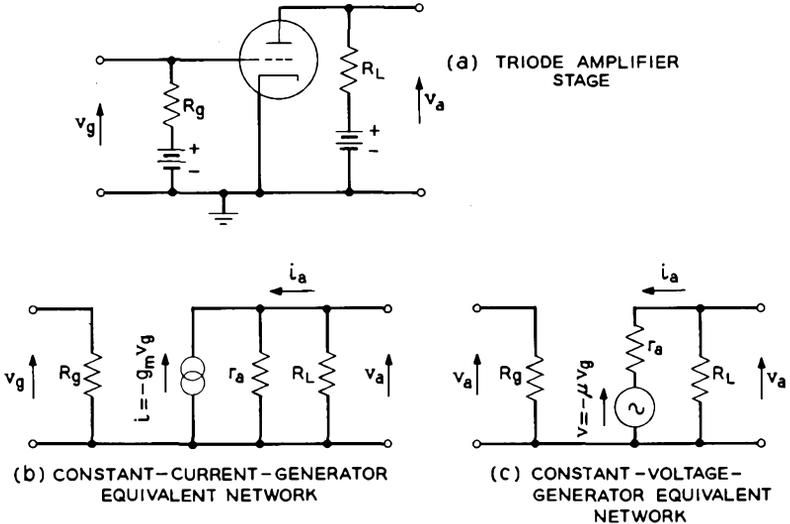


FIG. 6.3-2 A grounded-cathode triode amplifier stage and two low-frequency, small-signal equivalent networks.

equivalent network for analysis of the low-frequency response of the simple grounded-cathode triode amplifier stage shown in Figure 6.3-2(a). Constant current generators always have infinite internal impedance. This means that the voltages applied across the terminals of the generator do not affect the current generated.

In constructing the equivalent network shown in Figure 6.3-2(b), we have assumed that the control grid of the triode has a sufficiently negative bias that it does not draw any current. Secondly, since Equation (6.3-4) is valid only when the amplitudes of the ac signals are small, the equivalent

network is suitable only for analysis of the small-signal operation of the stage. Finally, since we have not accounted for interelectrode capacitances, interwiring capacitances, and lead inductances, the equivalent network can be used only at sufficiently low frequencies that these capacitances and inductances can be neglected.

The second network described by Equation (6.3-4) is shown in Figure 6.3-3. It involves a constant voltage generator, which generates a voltage $-\mu v_g$, in series with a resistance equal to the dynamic anode resistance $r_a$. By equating the sum of the voltages around the loop in this network to zero, we obtain.

FIG. 6.3-3 The constant-voltage-generator small-signal equivalent network for a grid-controlled tube.

$$v_a - i_a r_a + \mu v_g = 0 \qquad (6.3\text{-}6)$$

Substituting $\mu = g_m r_a$ into this, we see that this equation is also equivalent to Equation (6.3-4). Thus the network shown in Figure 6.3-3 also can be used to simulate the ac response of a grid-controlled tube with ac voltages $v_g$ and $v_a$ applied to the control grid and anode. Figure 6.3-2(c) shows the construction of a constant-voltage-generator equivalent network for the triode stage in Figure 6.3-2(a). Constant voltage generators always have zero internal impedance.

If the triode in Figure 6.3-2(a) were replaced by a tetrode or pentode and the additional electrodes were maintained at constant potentials, the equivalent networks shown in Figures 6.3-2(b) and 6.3-2(c) would still be applicable at low frequencies. However, since $\mu$ and $r_a$ tend to be extremely high for tetrodes and pentodes, the constant-current-generator equivalent network is usually found more satisfactory for the analysis of stages using these tubes.

The voltage gain of the amplifier stage shown in Figure 6.3-2 is given by the magnitude of the ratio of the ac voltage developed across the load resistance $R_L$ to the input voltage $v_g$. From the figure it can be seen that this ratio is given by

$$\text{voltage gain} = \left| \frac{v_a}{v_g} \right| = \frac{\mu R_L}{r_a + R_L} = g_m \frac{r_a R_L}{r_a + R_L} \qquad (6.3\text{-}7)$$

In stages employing tetrodes and pentodes, $r_a$ may be very large compared with $R_L$, in which case the gain is very nearly given by

$$\text{voltage gain} = g_m R_L \qquad (6.3\text{-}8)$$

Next let us extend the equivalent network shown in Figure 6.3-2(b) to include the effects of interelectrode capacitances and stray wiring capacitances. In the 417A triode described in Sections 5.1 and 5.2, the interelectrode capacitances are:

<div align="right">pf*</div>

Grid to-anode capacitance........................1.5
Grid-to-cathode capacitance......................5.4
Anode-to-cathode capacitance ....................0.2

*One picofarad $= 10^{-12}$ farad.

These capacitances include those between the internal leads to the electrodes. In addition, the stray capacitances associated with external wiring and other circuit components may easily amount to 4 or 5 pf both between the grid circuit and ground and between the anode circuit and



(a) TRIODE AMPLIFIER STAGE

(b) TETRODE AMPLIFIER STAGE

$C_{cg}$ = CATHODE-CONTROL GRID CAPACITANCE PLUS STRAYS
$C_{ca}$ = CATHODE-ANODE CAPACITANCE PLUS STRAYS
$C_{sa}$ = SCREEN-ANODE CAPACITANCE
$C_{ga}$ = CONTROL GRID-ANODE CAPACITANCE
$C_{gs}$ = CONTROL GRID-SCREEN CAPACITANCE
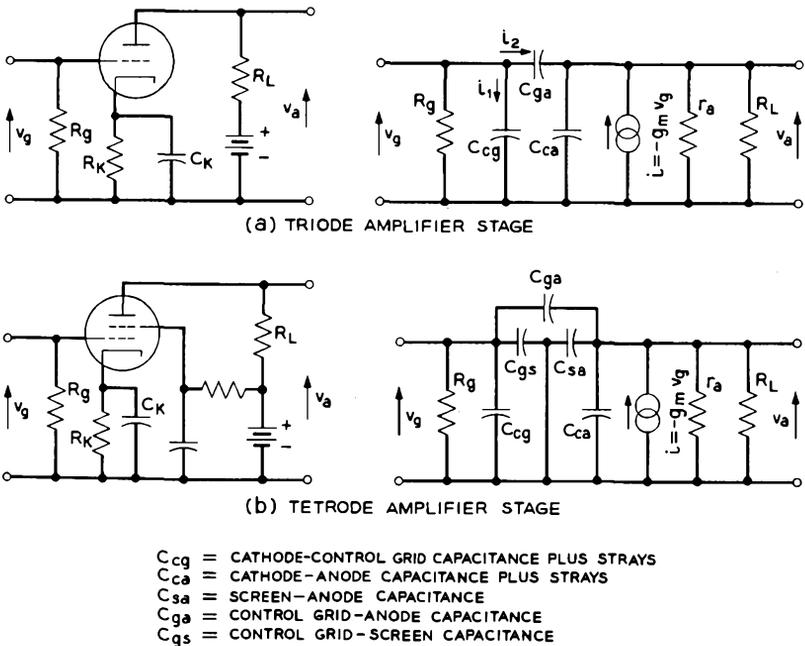
Fig. 6.3-4 Small-signal equivalent networks for a triode amplifier stage and a tetrode amplifier stage. The networks include the effects of interelectrode capacitance and stray capacitance.

ground.  At frequencies above 100 kc the reactances associated with
these capacitances may be comparable with other circuit impedances and
consequently they must be considered in the circuit analysis.  Figure
6.3-4 shows equivalent networks for simple triode and tetrode amplifier
stages where the interelectrode and stray capacitances are taken into
account.  In both networks we have assumed that the reactance of the
capacitance $C_K$ is small compared with $R_K$ and therefore that both quanti-
ties can be neglected.

Let us use the equivalent network shown in Figure 6.3-4(a) to determine
the input admittance of the triode stage shown in the figure.  This ad-
mittance is given by

$$Y_i = \frac{1}{R_g} + \frac{I_1 + I_2}{V_g} \tag{6.3-9}$$

where $I_1$ is the phasor corresponding to the current $i_1$ which flows through
the cathode-to-grid capacitance $C_{cg}$ (that is, $i_1 = \text{Re } I_1 \epsilon^{j\omega t}$, where $\omega$ is the
angular frequency of the signal), $I_2$ is the phasor corresponding to the
current $i_2$ which flows through the grid-to-anode capacitance $C_{ga}$, and $V_g$ is
the phasor corresponding to the input voltage $v_g$.  Now $I_1 = V_g \, j\omega C_{cg}$,
and $I_2 = (V_g - V_a) \, j\omega C_{ga}$, where $V_a$ is the phasor corresponding to the
output voltage $v_a$.  Substituting these expressions into Equation (6.3-9),
we obtain

$$Y_i = \frac{1}{R_g} + j\omega \left[ C_{cg} + \left( 1 - \frac{V_a}{V_g} \right) C_{ga} \right] \tag{6.3-10}$$

The ratio $- V_a/V_g$ is the complex gain of the stage.  Since this is likely
to have a large, positive real part, the input signal may be shunted by a
large apparent capacitance.  Because $C_{ga}$ is much smaller in tetrodes and
pentodes, the shunting capacitance given by Equation (6.3-10) is greatly
reduced in stages employing these tubes.  In Section 6.4 we shall see that
a low input capacitance is needed for tubes used in high-gain, broadband
amplifiers.  For this reason, most multistage high-gain, broadband ampli-
fiers employ tetrode or pentode tubes.  However, triodes are sometimes used
in the input stages of these amplifiers because of their better noise per-
formance.  (See Chapter 13.)

Finally, let us determine the low-frequency input admittance of the
grounded-grid amplifier stage shown in Figure 6.3-5(a).  A low-frequency
equivalent network for the stage is shown in part (b) of the figure.  When a
small ac voltage $v_c$ is applied to the cathode, an ac current $i_a$ flows in the
cathode and anode leads of the tube, and an ac current $i_1$ flows in the re-
sistance $r_a$ of the equivalent network.  We neglect the effects of the anode-
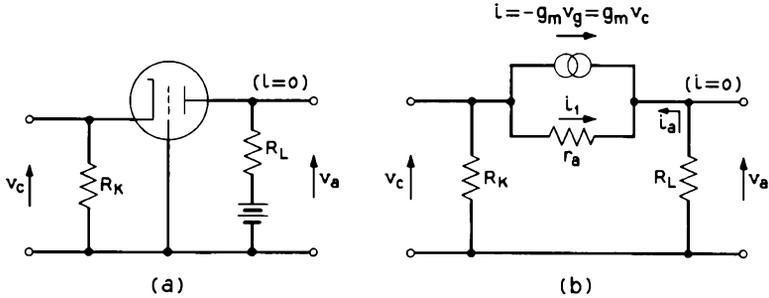
FIG. 6.3-5 A grounded-grid amplifier stage and its low-frequency equivalent network.

to-cathode capacitance in this analysis. Applying Kirchhoff's Laws to the network, we obtain

$$I_a = -(g_m V_c + I_1) \tag{6.3-11}$$

and

$$V_c = I_1 r_a - I_a R_L \tag{6.3-12}$$

where $I_a$, $I_1$, and $V_c$ are the phasors corresponding to $i_a$, $i_1$, and $v_c$. These equations can be solved for $I_a/V_c$, and the input admittance can be expressed as

$$Y_i = \frac{1}{R_K} - \frac{I_a}{V_c} = \frac{1}{R_K} + \frac{g_m + 1/r_a}{1 + R_L/r_a} \tag{6.3-13}$$

If $r_a \gg R_L$ and $\mu \gg 1$, this reduces to

$$Y_i = \frac{1}{R_K} + g_m \tag{6.3-14}$$

If the tube has a transconductance of 0.025 mho, the input impedance is 40 ohms in parallel with $R_K$. Grounded-grid amplifier stages are of limited usefulness because of their low input impedance.

## 6.4  Gain-Bandwidth Product

Resonant circuits are used in high-frequency amplifier stages to establish the frequency at which maximum gain is obtained. Because the gain falls off on either side of this frequency, each stage can be characterized by a bandwidth, or a range of frequencies about the resonant frequency over which the gain of the stage is within certain limits. Usually these limits are expressed as a number of db below the maximum gain. Thus the "3-db bandwidth" of an amplifier stage is the range of frequencies over which the

power gain is within 3 db of the maximum gain. Since 3 db very nearly corresponds to a power ratio of 2 and a voltage ratio of $\sqrt{2}$, the voltage gain at the extreme frequencies of the 3-db bandwidth is $1/\sqrt{2}$ times that at maximum gain.

In this section we show that the product of gain and bandwidth for an amplifier stage employing resonant circuits is a constant which depends only upon parameters of the tube itself and upon the external capacitances shunting the tube. By changing the load resistance in the anode circuit, both the gain of the stage and the bandwidth change in such a manner that the product of gain and bandwidth remain unchanged. First, it will be helpful to examine a few properties of resonant circuits.

If a parallel-resonant circuit is excited by an external source and then allowed to oscillate freely, the excitation energy is stored alternately in the electric field of the capacitance and in the magnetic field of the in- ductance. As the oscillation continues, the losses of the circuit cause the amplitude of the oscillation to de- crease. Although there is always some capacitance and resistance be- tween the terminals of the induct- ance, and some inductance and re- sistance between the terminals of the capacitance, for most purposes the circuit can be assumed to consist of a pure inductance, a pure capacitance, and a pure resistance, all in parallel. Such a circuit is shown in Figure 6.4-1. The resistance $R$ is assumed
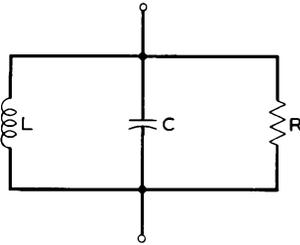


FIG. 6.4-1 A parallel resonant circuit.

to be of a magnitude which accounts for losses in the inductance and capaci- tance, as well as any additional resistance which is connected in parallel with the circuit. The magnitude of the admittance of the parallel combina- tion is given by

$$|Y| = \sqrt{(1/R)^2 + (\omega C - 1/\omega L)^2} \qquad (6.4\text{-}1)$$

where $\omega$ is the angular frequency of the exciting signal. If the circuit is excited by a constant current source which generates a current $I_o \sin \omega t$, a sinusoidal voltage with amplitude $I_o/|Y|$ is developed across the circuit. Resonance occurs when $\omega C = 1/\omega L$, or $\omega = 1/\sqrt{LC}$. At this frequency $|Y|$ is a minimum and equal to $1/R$.

A measure of the quality of a parallel resonant circuit is given by the $Q$ of the circuit, which is defined as

$$Q = 2\pi \frac{\text{energy stored at resonance}}{\text{energy lost per cycle}} \qquad (6.4\text{-}2)$$

Since the energy stored in a capacitance $C$ which is charged to a voltage $V$ is $\frac{1}{2}CV^2$, it is easily shown that, for a parallel resonant circuit, $Q$ is given by

$$Q = 2\pi \frac{\frac{1}{2}CV_{max}^2}{\frac{V_{max}^2}{2R} \cdot \frac{1}{f_o}} = \omega_o CR = \frac{R}{\omega_o L} \qquad (6.4\text{-}3)$$

where $V_{max}$ is the peak voltage appearing across the circuit, and $\omega_o = 2\pi f_o$ is the angular frequency of resonance.

From Equation (6.4-1) it follows that, if the circuit is excited by an ac current of constant amplitude but variable frequency, the voltage developed across the circuit falls to $1/\sqrt{2}$ of its maximum value when the angular frequency $\omega$ is such that

$$\left| \omega C - \frac{1}{\omega L} \right| = \frac{1}{R} \qquad (6.4\text{-}4)$$

Multiplying both sides of this equation by $R/Q$ and substituting for $R/Q$ from Equation (6.4-3), we obtain

$$\left| \frac{\omega}{\omega_o} - \frac{\omega_o}{\omega} \right| = \frac{1}{Q} \qquad (6.4\text{-}5)$$

Rearranging this gives

$$\left| \frac{(\omega - \omega_o)(\omega + \omega_o)}{\omega \omega_o} \right| = \frac{1}{Q} \approx \frac{2\Delta f}{f_o} \qquad (6.4\text{-}6)$$

or

$$Q \approx \frac{f_o}{2\Delta f} \qquad (6.4\text{-}7)$$

where $\Delta f$ is the number of cycles away from resonance at which the voltage across the circuit falls to $1/\sqrt{2}$ of its maximum value. The frequency $2\Delta f$ gives a measure of the width of the resonance response of the tuned circuit when excited by a constant current generator.

Let us now use these parallel-resonant-circuit concepts to determine the gain-bandwidth product for a pentode amplifier stage. The amplifier stage is shown in Figure 6.4-2, together with its equivalent network. The capacitance $C_i$ in the equivalent network is the input capacitance of the tube, or the sum of the capacitances between the control grid and all other electrodes except the anode. $C_o$ is the output capacitance of the tube, or the sum of the capacitances between the anode and all other electrodes except the control grid. $C_s$ is the sum of the stray capacitances shunting the output circuit plus any lumped capacitances that are connected across the output circuit. We shall assume that the capacitance $C_{ga}$ between the

(a) PENTODE AMPLIFIER STAGE



(b) EQUIVALENT NETWORK
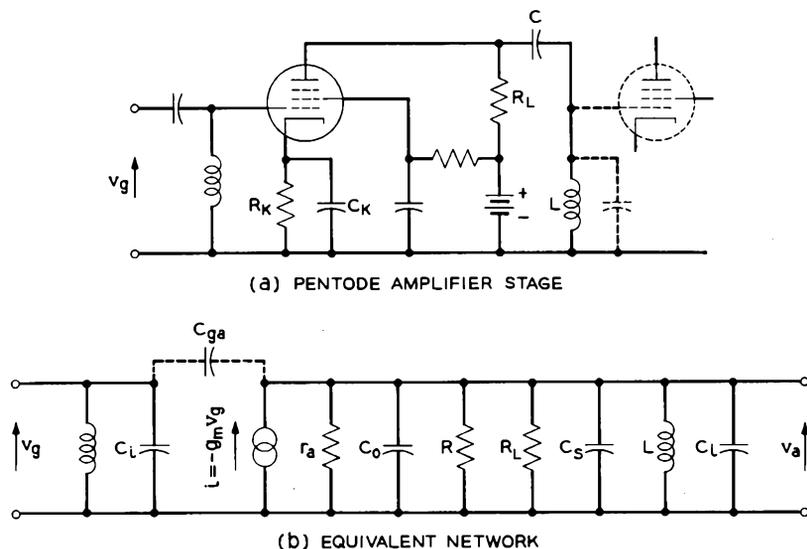
FIG. 6.4-2 A pentode amplifier stage with a parallel resonant circuit connected to the anode.

control grid and anode is sufficiently small that it can be neglected. Also, the coupling capacitance $C$ is assumed to be large enough that it has negligible reactance at the frequencies under consideration. The resistance $R$ accounts for the losses in the output circuit, including those in the inductance and capacitances. Usually it will be much larger than the load resistance $R_L$.

Maximum gain for the stage occurs at the frequency at which the inductance $L$ resonates with the capacitances in parallel with it. If we let $R_e$ be the equivalent resistance of $r_a$, $R$, and $R_L$ in parallel, the maximum gain of the amplifier stage is given by

$$\text{gain} = g_m R_e \qquad (6.4\text{-}8)$$

Since $r_a$ and $R$ are likely to be large compared with $R_L$, this is very nearly equal to $g_m R_L$. The $Q$ of the tuned circuit shunted by the resistance $R_e$ is given by

$$Q = \omega_o (C_o + C_i + C_s) R_e \qquad (6.4\text{-}9)$$

where $C_i$ is the input capacitance of the following stage. From Equation (6.4-7), the bandwidth over which this stage gives at least $1/\sqrt{2}$ of the maximum voltage gain is given by

$$2\Delta f = \frac{f_o}{Q} = \frac{1}{2\pi(C_o + C_i + C_s)R_e} \qquad (6.4\text{-}10)$$

Finally, the product of gain and bandwidth for the stage is obtained by multiplying Equations (6.4-8) and (6.4-10). Thus,

$$\text{gain} \times \text{bandwidth} = \frac{g_m}{2\pi(C_o + C_i + C_s)} \qquad (6.4\text{-}11)$$

If the stray capacitances were zero and there were no lumped capacitances shunting the output circuit, the expression would reduce to

$$\text{gain} \times \text{bandwidth} = \frac{g_m}{2\pi(C_o + C_i)} \qquad (6.4\text{-}12)$$

The product given by Equation (6.4-12) represents a theoretical limit that is not attained in practice because of the stray capacitances that are always present. If we assume that the input capacitance $C_i$ of the tube in the next stage is the same as that of the tube under consideration, the gain-
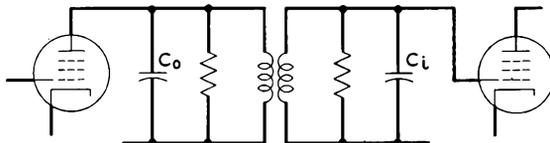


FIG. 6.4-3 A double-tuned resonant circuit between amplifier stages.

bandwidth product given by Equation (6.4-12) depends entirely on the parameters of the tube itself, and not upon $R_L$ or $L$. The equation shows that we can design the stage to have high gain at the expense of narrow bandwidth, or lower gain and a larger bandwidth. High gain, of course, is associated with large $R_L$. It should be noted that $R_L$ includes the effect of the input loading of the following stage (see Section 7.3), and in practice this loading places an upper limit on the maximum gain per stage.

Sometimes double-tuned resonant circuits are employed between amplifier stages, as illustrated in Figure 6.4-3. For such a circuit the gain-bandwidth product depends upon the degree of coupling between the two circuits and upon the $Q$'s of the circuits. If the circuits are adjusted for "critical coupling" and if the $Q$'s of the primary and secondary circuits are equal, it can be shown[3] that the gain-bandwidth product of the stage is given by

$$\text{gain} \times \text{bandwidth} = \frac{g_m}{2\pi\sqrt{2}\sqrt{C_i C_o}} \qquad (6.4\text{-}13)$$

where $C_i$ is the total input capacitance of the following stage, including strays, and $C_o$ is the total output capacitance, including strays.

---

[3] Reference 6b.

Equations (6.4-12) and (6.4-13) tell us that tubes to be used in high-gain, broadband amplifiers should have high transconductance and low input and output capacitances. Unfortunately, increasing the transconductance by increasing the cathode area also increases both the input and output capacitances, and if the transconductance is increased by decreasing the cathode-to-grid spacing, this also increases the input capacitance. However, in practice the stray capacitances which shunt the input circuit represent an appreciable part of the input capacitance, so that, if the tube design is changed to double the cathode area and hence double the transconductance, it is likely that the total input capacitance including strays will not be doubled, the stray capacitance being nearly constant. Furthermore, when double-tuned circuits are employed, the gain-bandwidth product increases directly with the transconductance but depends only on the reciprocal of the square root of the input capacitance. For these reasons, tubes for use in high-gain, broadband amplifiers are designed to have high transconductance. A large spacing between the anode and screen grid or suppressor grid is also desirable, since the output capacitance $C_o$ is then reduced.

Equation (6.4-12) is frequently used to give a figure of merit for comparing tubes for use in high-gain, broadband amplifier applications.[4] By using a tube with a high gain-bandwidth product in a multistage amplifier, less stages are needed to achieve a total over-all gain and bandwidth. Table 5.2-1 lists the maximum-possible-gain-bandwidth product $g_m/2\pi(C_o + C_i)$ for the 448A tetrode and 403A/6AK5 pentode as 215 and 95 Mc, respectively. In practice, stray capacitance amounting to a total of 9 pf might shunt the input and output circuits, and in this case the *actual* gain-bandwidth products would be reduced to 160 Mc and 46 Mc for stages using these tubes.

<div align="center">

*PROBLEMS*

</div>

6.1 Show that for a space-charge-limited planar diode the current pulse induced in the external circuit by the passage of a single electron from the cathode to the anode is given by

$$i = \frac{3et^2}{\tau^3}$$

where it is assumed that the potential minimum coincides with the cathode, $\tau$ is the length of the current pulse, and $t$ is the time measured from the instant the electron passes the potential minimum. Hint: Show that $i$ is proportional to $t^2$ without

---

[4] Sometimes Equation (6.4-13) is used.

calculating the exact value of the constant of proportionality, then use the fact that $\int_0^\tau idt = e$ to find the value of this constant.
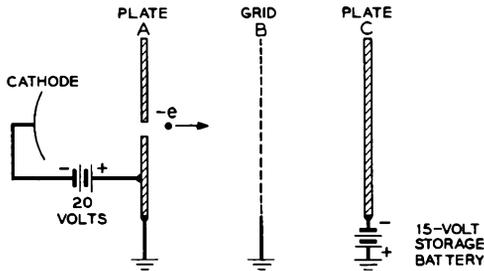


Problem 6.2

6.2 In the apparatus shown, a single electron leaves the cathode and is accelerated toward plate $A$. It passes through the hole in plate $A$ with 20 electron volts of kinetic energy and travels on through grid B to strike plate $C$.

(a) Sketch the current that flows from ground toward grid $B$ as the electron travels from plate $A$ to plate $C$. Indicate the relative values of the induced current at the times when the electron is at electrodes $A$, $B$, and $C$.

(b) When the electron has struck plate $C$, where can the 20 electron volts of kinetic energy that the electron had when it passed through plate $A$ be found?
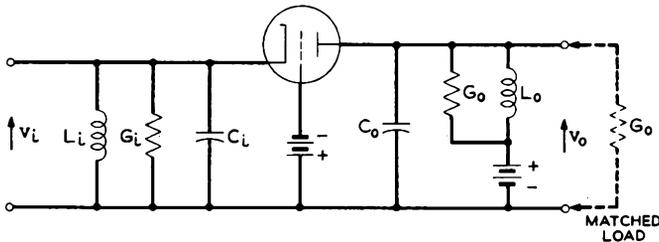


Problem 6.3

6.3 The figure shows a grounded-grid triode amplifier stage. $C_i$ and $C_o$ are, respectively, the input and output capacitances of the tube plus strays. (Notice that in this case the input capacitance is the cathode-to-grid capacitance, and the output capacitance is the grid-to-anode capacitance.) The input and output inductance $L_i$ and $L_o$ resonate with $C_i$ and $C_o$, respectively, at the same frequency. The conductance $G_i$ and $G_o$ take into account the effects of losses in the input and output resonant circuits.

(a) Sketch the constant current generator equivalent network for the stage. (Neglect the cathode-to-anode capacitance.)

(b) The power gain of the stage can be defined as

$$\frac{\text{power delivered to a matched load with conductance } G = G_o}{\text{power dissipated in input circuit}}$$

Show that the power gain at resonance is given by $g_m/4G_o$. Assume that the dynamic anode resistance $r_a$ is sufficiently large that very little of the current from the constant current generator flows through $r_a$.

(c) Show that the product of the power gain at resonance and the 3-db bandwidth for the stage is $g_m/4\pi C_o$, where the 3-db bandwidth is the bandwidth between frequencies at which the power gain has dropped 3 db below maximum gain.

6.4 In an amplifier stage such as that shown in Figure 6.3-4(a) the gain falls off with increasing frequency because of the capacitance shunting the output circuit. Show that the product of the zero-frequency voltage gain and bandwidth over which the gain is within 3 db of the zero-frequency gain is given by $g_m/2\pi C$, where $C$ is the total capacitance shunting the output. Neglect the effect of the control grid-to-anode capacitance.

## REFERENCES

The induced currents that flow in the circuit joining two electrodes when charge moves in the space between the electrodes is discussed in the following reference:

6a. W. Shockley, *J. Appl. Phys.* 9, 635, 1938.

A thorough treatment of amplifiers employing grid-controlled tubes is given in the following reference:

6b. *Vacuum Tube Amplifiers, MIT Radiation Laboratory Series*, Vol. 18, McGraw-Hill Book Co., Inc., New York, 1948.

# Chapter 7

## HIGH-FREQUENCY EFFECTS AND
## BEAM ADMITTANCE

In Chapter 6 we found that a modulated beam crossing between two electrodes induces an ac current in the external circuit joining the electrodes. If the electron transit time across the gap between the electrodes is small compared with the period of the beam modulation frequency, the induced ac current equals the ac component of the beam current. However, at high frequencies and moderate or low electron velocities, the electron transit time may be an appreciable part of the period of the ac signal. Suppose, for example, that the electron transit time is one half the period of the modulation frequency. In this case one half a cycle of the modulated beam lies between the electrodes at any instant. Since the induced current is a sum of contributions from the moving electrons in each volume element between the electrodes and since the phase of the beam modulation varies over the distance between the electrodes, the contributions to the induced current are not all in phase, and consequently the total induced current is less than the ac beam current.

Consider a modulated beam crossing between the screen grid and anode of a tetrode or between the grid and anode of a grounded-grid triode. If the modulation frequency is sufficiently high that the electron transit time is comparable with the period of the signal, the ac induced current flowing in the external circuit is less than the ac beam current, and the tube transconductance is correspondingly reduced. In general, at frequencies sufficiently high that electron transit-time effects are important, the transconductance and other tube parameters become complex numbers instead of real numbers, and their magnitudes are functions of the signal frequency.

Suppose an unmodulated beam passes through two grids. If an ac signal is applied between the grids, the electrons are acted on by the ac field, and their velocity acquires an ac component. Each electron in transit between

204

the grids induces a current proportional to its velocity in the external circuit joining the grids and hence in the signal generator. Since the total induced current is a sum of contributions by the individual electrons in transit between the grids and since the electron velocities have an ac component, the total induced current has an ac component. Furthermore, because the electrons have inertia, the ac component of their velocity is generally not in phase with the applied signal, and the ac induced current is likewise not in phase with the applied signal. The ratio of the ac induced current flowing through the signal generator to the ac voltage applied to the grids is called the *beam admittance* or beam loading. Beam admittance effects become important at high frequencies where the electron transit time is comparable with the period of the ac signal. The small-signal beam admittance is directly proportional to the dc beam current. Generally it has both a conductive and a susceptive part.

In Chapter 6 we noted that the capacitance between the control grid and anode in a grounded-cathode amplifier stage causes a capacitive loading of the input circuit. At high frequencies a conductive loading of the input circuit also occurs in grounded-cathode stages. This is caused partly by the inductance in the cathode lead and partly by beam admittance. At signal frequencies where input conductance first becomes important, the total input conductance is approximately proportional to the product of the low-frequency transconductance of the tube and the square of the signal frequency.

High input conductance is a principal limitation of the performance of high-frequency grounded-cathode amplifiers and oscillators. In some tubes, input conductance limitations become important at frequencies of the order of a few tens of megacycles. However, by using very short cathode leads, and sometimes multiple cathode leads, the cathode lead inductance can be reduced, and the useful operating frequency raised. Similarly, short electrode spacings and high electrode voltages help to increase the operating range by reducing the electron transit times. The electron transit time across the cathode-control grid region can be reduced by locating the plane of the grid wires close to the potential minimum and by using a higher voltage at the grid plane and hence a higher cathode current density.

Still another increase in the input admittance takes place when series resonance occurs between the inductance in the cathode and control grid leads and the capacitance between the cathode and control grid. In designing high-frequency grid-controlled tubes, care must be taken that the frequency of this resonance is well above the operating range.

Generally tubes in which the electrode connections are brought out through the base are useful only at frequencies below 100 Mc, or perhaps a few hundred megacycles in exceptional cases. If operation at still higher

frequencies is needed, planar electrodes are used, and the electrode connections are brought out radially by means of disc leads which pass through the envelope. In this way the lead inductance and rf losses in the leads can be reduced to a minimum.

The chapter concludes with a description of two tubes designed for operation at frequencies well above 100 Mc.

## 7.1 Electron Transit Time and Beam Admittance

When the electron transit time between the electrodes of a tube is comparable with the period of the ac signal applied to the electrodes, the tube behavior can be modified in several important ways. In this section we consider under separate headings three examples which illustrate these effects.

### (a) A Modulated Beam Passing Between Two Electrodes

As a starting point, it will be helpful to consider the dimensions, beam voltages, and frequencies that are likely to be involved in problems in which transit-time effects are important. Suppose an electron beam is accelerated from zero velocity through 100 volts and then passes at constant velocity through two grids separated by 1 mm. The electron velocity is $5.93 \times 10^6$ meters/sec, and the transit time from one grid to the other is $(1/5.93) \times 10^{-9}$ sec. This time is equal to the period of a 5930-Mc signal. If the beam current were modulated at this frequency, one whole cycle of the modulated beam would be between the grids at any instant. If the grids were spaced by 1 cm, the electron transit time would equal the period of a 593-Mc signal.

Let us proceed to determine the current induced in the external circuit
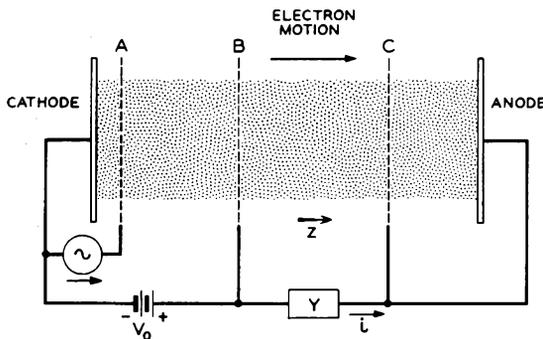


FIG. 7.1-1 The beam passing through grids $B$ and $C$ is modulated by an ac signal applied to grid $A$.

joining two grids when a modulated beam passes between the grids. A suitable electrode arrangement is shown schematically in Figure 7.1-1. The current of electrons drawn from a planar cathode is modulated by an ac voltage applied to grid $A$. The electrons are then accelerated through a potential rise $V_o$ and pass at nearly uniform velocity between grids $B$ and $C$. Finally, they are collected by a planar anode. An admittance $Y$ is connected between grids $B$ and $C$. It is assumed that the ac voltage developed across this admittance is small compared with the voltage $V_o$. It is further assumed that grids $B$ and $C$ are ideal electrostatic shields, so that electric fields on one side of the grids do not penetrate through the grids to the other side.

The results of Chapter 6 tell us that *the current flowing in the admittance Y is determined only by the electron current in the space between grids B and C. The electrons striking the anode cancel positive charges which have flowed to the anode to meet them, but they do not give rise to an additional component of current through the admittance Y.* In fact, the induced current flowing in the admittance $Y$ would be no different if grid $C$ were replaced by the anode.

Equation (6.1-8) indicates that the ac current flowing in the admittance $Y$ is given by

$$i(t) = -\int_{\text{volume}} \mathbf{J}(x,y,z,t) \cdot \mathbf{E}_1 dxdydz \qquad (7.1\text{-}1)$$

where $\mathbf{J}(x,y,z,t)$ is the instantaneous ac current density at the volume element $dxdydz$, and $\mathbf{E}_1$ is a vector function of position discussed in Section 6.1. The integral is taken over the volume of the beam between grids $B$ and $C$. If the grids extend well beyond the edge of the beam on all sides, $\mathbf{E}_1$ within the beam is normal to the plane of the grids, it is directed from grid $B$ toward grid $C$, and it is equal in magnitude to $1/d$, where $d$ is the spacing between the grids. Let the electron velocity corresponding to the beam voltage $V_o$ be $u_o$, and let the direction normal to the plane of the electrodes be the $z$ direction. We shall assume that the magnitude of $\mathbf{J}$ can be expressed as $J_1(x,y) \sin \omega(t - z/u_o)$, and that $\int J_1(x,y)dxdy = I_1$, where the integral is taken over the beam cross section. Substituting for $\mathbf{E}_1$ and $\int J_1(x,y)dxdy$ in Equation (7.1-1), and assuming that $\mathbf{J} \cdot \mathbf{E}_1$ is a *negative* quantity, we obtain

$$i(t) = \frac{I_1}{d} \int_0^d \sin \omega\left(t - \frac{z}{u_o}\right)dz$$

$$= \frac{I_1}{d} \frac{u_o}{\omega}\left[\cos\left(\omega t - \frac{\omega d}{u_o}\right) - \cos \omega t\right] \qquad (7.1\text{-}2)$$

where distance $z$ is assumed to be measured from grid $B$ toward grid $C$. If we set $(\omega t - \omega d/2u_o) = A$, and $\omega d/2u_o = B$, the part of Equation (7.1-2)

in brackets can be written in the form $\cos(A - B) - \cos(A + B)$. Then using the identity that $\cos(A - B) - \cos(A + B) = 2\sin A \sin B$, we can write Equation (7.1-2) as

$$i(t) = I_1 \frac{\sin\dfrac{\omega d}{2u_o}}{\dfrac{\omega d}{2u_o}} \sin \omega\left(t - \frac{d}{2u_o}\right) \qquad (7.1\text{-}3)$$

Finally, the electron transit time $T_o$ between grids $B$ and $C$ is equal to $d/u_o$, so that

$$i(t) = I_1 \frac{\sin\dfrac{\omega T_o}{2}}{\dfrac{\omega T_o}{2}} \sin \omega\left(t - \frac{T_o}{2}\right) \qquad (7.1\text{-}4)$$

The factor $[\sin(\omega T_o/2)]/(\omega T_o/2)$ is known as the beam-coupling coefficient, and the angle $\omega T_o$ is called the transit angle. The beam-coupling coefficient measures the ratio of the ac current induced in the external circuit to the ac component of the beam current. In subsequent discussion we shall designate this ratio with the letter $M$. Figure 7.1-2 shows a plot of $M$ as a function of the transit angle $\omega T_o$.

From Equation 7.1-4 we see that the phase of the induced current flowing in the external circuit is the same as that of the beam current at a point midway between the grids.
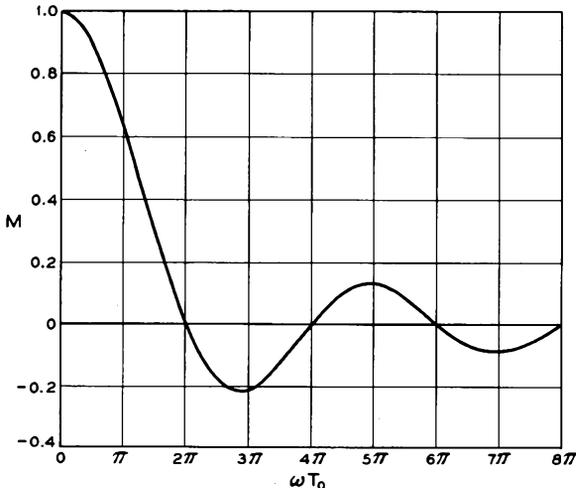


FIG. 7.1-2 The beam-coupling coefficient $M$ plotted as a function of $\omega T_o$.

Consider a tetrode amplifier stage in which the screen grid and anode are operated at the same voltage, and there is little potential depression between the screen grid and anode. The effect of a non-zero transit angle for the screen-grid to anode space is that the magnitude of the current generated by the constant current generator of the equivalent network is multiplied by the factor $M$ and hence is given by $|Mg_m V_g|$.

In the 448A tetrode, described in Section 5.3, the screen grid and anode are operated at 125 volts, and the spacing between these electrodes is 0.5 cm. If the beam current is reduced so that there is little potential depression between the electrodes, the transit angle $\omega T_o$ is equal to 4.74 $\times 10^{-9} f$ radians, where $f$ is the signal frequency in cycles per second. At a frequency of 1320 Mc, $\omega T_o = 2\pi$, and from Figure 7.1-2 we see that the beam-coupling coefficient $M$ is zero. Furthermore at integral multiples of this frequency, $\omega T_o$ is an integral multiple of $2\pi$, and the beam-coupling coefficient is again zero. The frequency 1320 Mc corresponds to the reciprocal of the electron transit time for this particular electrode spacing and beam voltage, so that at this frequency there is one whole cycle of the modulated beam between the electrodes at any instant. When the frequency is an integral multiple of 1320 Mc, there is an integral number of cycles of the modulated beam between the electrodes. Since the induced current flowing in the external circuit is a sum of contributions from each volume element between the electrodes, the *total* induced current must be zero when there is an integral number of cycles of the modulated beam between the electrodes.

The 448A is normally operated at frequencies below 100 Mc. Since $\omega T_o$ for the screen-grid to anode space of the 448A is equal to $0.15\pi$ at a frequency of 100 Mc, it is evident from Figure 7.1-2 that $M$ for this interaction region is nearly equal to unity for frequencies below 100 Mc. However, in designing tubes to operate at frequencies approaching 1000 Mc, transit-time effects in the output region of the tube may be a serious limitation. Notice that for a given ac beam current passing between the screen grid and anode of a tetrode, the electron transit time from the cathode to the screen grid does not affect the *magnitude* of the ac current induced in the output circuit provided the screen grid acts as a good electrostatic shield. However, the *phase* of the output signal is delayed by the finite transit times involved, and consequently the transconductance as defined by Equation (5.2-1) becomes a complex number.

(b)  *Small-Signal Admittance of an Unmodulated Beam Passing Between Two Grids at Equal Potential*

Suppose an ac signal is applied between two parallel planar grids, and an initially unmodulated electron beam passes through the grids. The motion

of the electrons in the space between the grids varies in response to the applied ac field, and this causes an induced ac current to flow through the signal generator. The ratio of the induced ac current flowing in the external circuit to the ac voltage applied by the signal generator is called the beam admittance, or beam loading.

Let us derive an expression for the small-signal admittance of an unmodulated beam passing between two grids whose dc potentials are equal. A suitable arrangement of electrodes is shown schematically in Figure 7.1-3. The grids are assumed to be ideal electrostatic shields. An electron
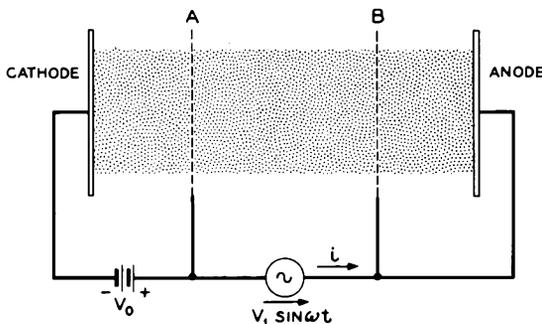


FIG. 7.1-3 The ac voltage applied between grids $A$ and $B$ causes the electron motion to vary in response to the applied field. An induced ac current therefore flows through the signal generator, and the electron beam acts as a load on the signal generator.

beam of current $I_o$ is accelerated through a potential rise $V_o$ before passing through grid $A$. The space-charge density between the grids is sufficiently low that there is negligible potential depression between the grids. An ac voltage $V_1 \sin \omega t$ is applied between the grids by means of a signal generator. We assume that the signal generator has zero internal impedance for both ac and dc signals and that $V_1$ is small compared with $V_o$.

We shall first determine the transit time of an electron which passes from grid $A$ to grid $B$ and is acted on by the ac field $(V_1/d) \sin \omega t$ as it crosses the interelectrode space, where $d$ is the grid spacing. The acceleration of the electron in the region between the grids is given by

$$\frac{d^2z}{dt^2} = \frac{\eta V_1}{d}\sin \omega t = \frac{u_o}{2T_o}\frac{V_1}{V_o}\sin \omega t \qquad (7.1\text{-}5)$$

where $u_o = \sqrt{2\eta V_o}$ is the velocity of the electrons passing through grid $A$, and $T_o = d/u_o$ is the electron transit time in the absence of an applied signal. Let the time the electron passed through grid $A$ be $t_o$. The velocity

of the electron at time $t$ is found by integrating Equation (7.1-5) with respect to time from $t_o$ to $t$ and adding $u_o$. Thus

$$\frac{dz}{dt} = u_o - \frac{u_o}{2\omega T_o}\frac{V_1}{V_o}(\cos \omega t - \cos \omega t_o) \qquad (7.1\text{-}6)$$

Integrating once more from $t_o$ to $t$, we find that the $z$ coordinate of the electron at time $t$ is given by

$$z = u_o(t - t_o) - \frac{u_o}{2\omega^2 T_o}\frac{V_1}{V_o}[\sin \omega t - \sin \omega t_o - \omega(t - t_o)\cos \omega t_o]$$
$$(7.1\text{-}7)$$

where distance $z$ is measured from grid $A$ toward grid $B$. If we set $z = d$ in this equation, the time $t$ corresponds to the time of arrival of the electron at grid $B$, and the time $t - t_o$ for $z = d$ is the electron transit time. We shall designate this electron transit time by $T$. Then

$$d = u_o T - \frac{u_o}{2\omega^2 T_o}\frac{V_1}{V_o}[\sin \omega t - \sin \omega t_o - \omega T \cos \omega t_o] \qquad (7.1\text{-}8)$$

or

$$T = T_o + \frac{1}{2\omega^2 T_o}\frac{V_1}{V_o}[\sin \omega t - \sin \omega t_o - \omega T \cos \omega t_o] \qquad (7.1\text{-}9)$$

where we have substituted $T_o$ for $d/u_o$. Now as $V_1 \to 0$, $T \to T_o$. Since we assume that $V_1$ is small, the second term on the right-hand side of Equation (7.1-9) is small compared with the first term. We shall use the approximations that $t_o = t - T_o$ and $T = T_o$ in the second term of the right-hand side of the equation. This is equivalent to neglecting terms which involve the product of two or more small quantities. The electron transit time then becomes

$$T = T_o + \frac{1}{2\omega^2 T_o}\frac{V_1}{V_o}[\sin \omega t - \sin \omega(t - T_o) - \omega T_o \cos \omega(t - T_o)]$$
$$(7.1\text{-}10)$$

We can now proceed to determine the current that flows in the external signal generator at time $t$. This current is a sum of contributions from all electrons in transit between the grids and hence includes all electrons that passed through grid $A$ from time $t$ back to time $t - T$. One electron induces a current $i = \frac{e}{d}\frac{dz}{dt}$ in the external circuit. In an increment of time $dt_o$ an amount of charge $I_o dt_o$ passes through grid $A$. At time $t$ this charge causes an induced current $\frac{I_o dt_o}{d}\frac{dz}{dt}$ to flow in the external circuit, where $dz/dt$ is the velocity of the electrons that comprise the charge $I_o dt_o$, and $dz/dt$ is

evaluated at time $t$. The total induced current flowing in the signal generator therefore is

$$i(t) = \int_{t_0=t-T}^{t_0=t} \frac{I_o}{d} \frac{dz}{dt} dt_o \qquad (7.1\text{-}11)$$

Substituting for $dz/dt$ from Equation (7.1-6) and carrying out the integration, we obtain

$$i(t) = I_o \frac{T}{T_o} - \frac{I_o}{2(\omega T_o)^2} \frac{V_1}{V_o} [\omega T \cos \omega t - \sin \omega t + \sin \omega(t - T)]$$
$$(7.1\text{-}12)$$

In the first term on the right we can substitute for $T$ from Equation (7.1-10), and in the second term we can use the approximations $t_o = t - T_o$ and $T = T_o$, since the second term is already a small quantity. Thus we obtain

$$i(t) = I_o + V_1 \frac{I_o}{2V_o} \left[ \frac{2(1 - \cos \omega T_o) - \omega T_o \sin \omega T_o}{(\omega T_o)^2} \sin \omega t \right.$$
$$\left. + \frac{2 \sin \omega T_o - \omega T_o(1 + \cos \omega T_o)}{(\omega T_o)^2} \cos \omega t \right] \quad (7.1\text{-}13)$$

This is of the form $i(t) = I_o + gV_1 \sin \omega t + bV_1 \cos \omega t$, where

$$g = g_o \frac{2(1 - \cos \omega T_o) - \omega T_o \sin \omega T_o}{2(\omega T_o)^2}$$

$$b = g_o \frac{2 \sin \omega T_o - \omega T_o(1 + \cos \omega T_o)}{2(\omega T_o)^2} \qquad (7.1\text{-}14)$$

and $g_o = I_o/V_o$. Figure 7.1-4 shows plots of $g/g_o$ and $b/g_o$.

Equation (7.1-13) shows that the current flowing in the signal generator due to the presence of the beam is made up of a dc term equal to the dc beam current and two ac terms, one in phase with the applied voltage and one in quadrature with the applied voltage. The capacitance between the grids also shunts the voltage generator, and an *additional* current is drawn from the signal generator to charge this capacitance. An equivalent network for the system is shown in Figure 7.1-5. The capacitance $C_o$ is the capacitance between the grids in the absence of the beam. Positive $b$ corresponds to a capacitative susceptance, and negative $b$ corresponds to an inductive susceptance. The admittance $g + jb$ is called the beam admittance or beam loading. It shunts any external circuit connected between the grids. Figure 7.1-4 shows that as $\omega T_o$ approaches zero, both $g$ and $b$ go to zero.

In further explanation of the beam admittance, let us return to Equation (7.1-6). Because the electrons are acted on by the applied ac field, their
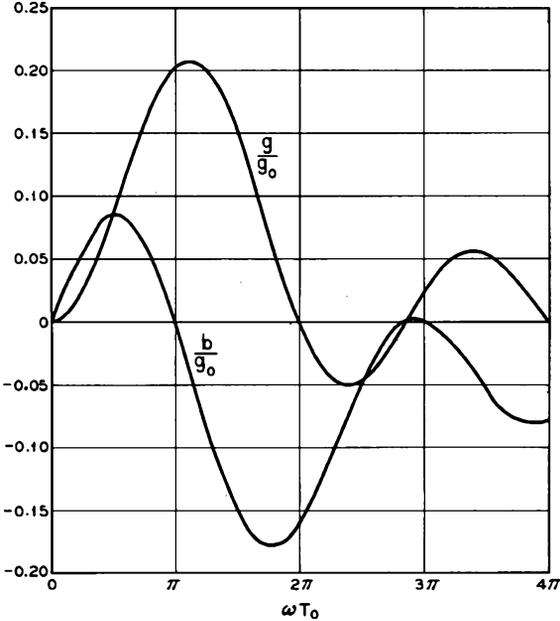
FIG. 7.1-4 The quantities $g/g_o$ and $b/g_o$ plotted as functions of the angle $\omega T_o$.

velocity has an ac component. However, each electron induces a current $i = \dfrac{e}{d}\dfrac{dz}{dt}$ in the external circuit, and since $dz/dt$ has an ac component, the induced current resulting from an individual electron has an ac component. When the induced currents from all the electrons in transit between the grids are added together, the resulting total induced current also has an ac component. Furthermore, because the electrons have inertia, the ac component of their velocity is not in phase with the applied field. This accounts for the susceptive part of the beam admittance.
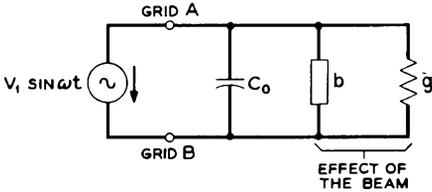


F IG. 7.1-5 An equivalent network for the region between grids $A$ and $B$ in Figure 7.1-3.

To illustrate the magnitudes that might arise from beam-loading effects, consider a 100-volt, 100-ma beam crossing between two grids spaced by 1 cm. We shall assume that the beam current density is sufficiently low that there is negligible potential depression between the grids. If a signal frequency of 200 Mc is applied to the grids, $\omega T_o = 2.11$ radians, and $g + jb = (1.38 + j\,0.78) \times 10^{-4}$ mho. This is equivalent to a resistance of 7200 ohms in parallel with a capacitive reactance of 13,000 ohms. If a parallel resonant circuit were connected between the grids, the beam admittance would change both the tuning and $Q$ of the circuit.

### (c)    Impedance of a Space-Charge-Limited Planar Diode

As a final example of electron transit-time effects, we derive in Appendix X the impedance of a space-charge-limited planar diode. It is assumed in the derivation that the potential minimum coincides with the cathode and that edge effects can be neglected. The impedance of the diode is found to be

$$z = \frac{r}{A} + j\frac{x}{A} \qquad (7.1\text{-}15)$$
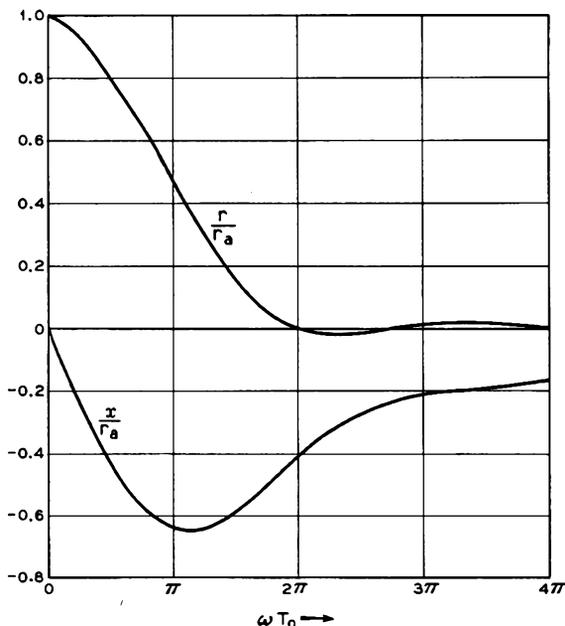


Fig. 7.1-6  The quantities $r/r_a$ and $x/r_a$ plotted as functions of the transit angle $\omega T_o$.

where $A$ is the area of the electrodes, and $r$ and $x$ are given by

$$r = 12r_a \left[ \frac{2(1 - \cos \omega T_o) - \omega T_o \sin \omega T_o}{(\omega T_o)^4} \right] \qquad (7.1\text{-}16)$$

and

$$x = -12r_a \left[ \frac{1}{6\omega T_o} + \frac{\omega T_o(1 + \cos \omega T_o) - 2 \sin \omega T_o}{(\omega T_o)^4} \right] \qquad (7.1\text{-}17)$$

Here $T_o$ is the time required for an electron to travel from the cathode to the anode under the influence of the applied dc field, and $\omega T_o$ is the corresponding transit angle. The resistance $r_a$ is the "dynamic anode re-
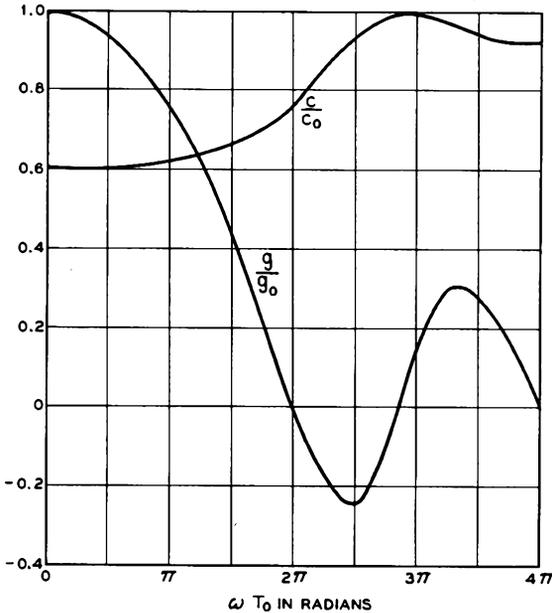


FIG. 7.1-7 The conductance and capacitance of a space-charge-limited planar diode. The conductance $g_o$ is equal to $3J_0A/2V_o$, and the capacitance $C_o$ is equal to the capacitance of the parallel plate capacitor formed by the electrodes in the absence of space charge.

sistance" for a unit area of the diode and is equal to $2V_{ao}/3J_o$, where $V_{ao}$ is the dc voltage applied to the anode, and $J_o$ is the dc current density drawn to the anode. Figure 7.1-6 shows plots of $r/r_a$ and $x/r_a$ as functions of the electron transit angle $\omega T_o$. Figure 7.1-7 shows the conductance and capacitance of the diode as functions of $\omega T_o$.

It is also shown in Appendix X that at *low* frequencies the diode can be represented by an admittance $Y$ such that

$$Y \approx g_o + j\omega \frac{3}{5} C_o \qquad (7.1\text{-}18)$$

where $g_o = A/r_a$, and $C_o$ is the capacitance of the parallel-plate capacitor formed by the electrodes in the absence of space charge. Thus at low frequencies the capacitance of the diode is *reduced* by the presence of space charge to 3/5 of the diode's capacitance in the absence of space charge. That the capacitance should be *changed* by the space charge is perhaps not surprising if we note that the field distribution and charges on the electrodes are quite different when space charge is present. In fact, when the potential minimum coincides with the cathode, there is no surface charge on the cathode at all. A low-frequency equivalent network for the diode is shown in Figure 7.1-8.



FIG. 7.1-8 A low-frequency equivalent network for a space-charge-limited planar diode.

## 7.2   The Llewellyn and Peterson Equations

An important contribution to present understanding of the high-frequency electronics of grid-controlled tubes with planar electrodes was made in some studies by Llewellyn and Peterson.[1] We shall not attempt to summarize their paper here but merely show the form of the equations from



(a)

(b)
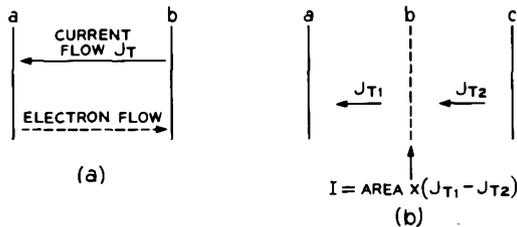
FIG. 7.2-1   (a) The interaction region applicable to Equation (7.2-1).   (b) The interaction region bounded by planes $a$ and $b$ is followed by a second interaction region bounded by planes $b$ and $c$. From $J_a$, $U_a$, and $J_{T1}$, the quantities $J_b$ and $U_b$ can be calculated. These can be used as entrance conditions for the second interaction region, and $J_c$ and $U_c$ can be calculated.

[1]Reference 7a.

which their studies developed. The reader who is interested in pursuing further the subject of high-frequency effects in grid-controlled tubes will find many interesting problems examined in the Llewellyn-Peterson paper.

The approach taken by Llewellyn and Peterson considers the parallel flow of electrons between two planes, $a$ and $b$, as illustrated in Figure 7.2-1 (a). Plane $a$ might be a cathode or an ideal grid, whereas plane $b$ might be an anode or an ideal grid. Suppose the electron beam passes through plane $a$ with an ac component of convection current density $J_a$ and an ac component of velocity $U_a$. An ac voltage $V_a - V_b$ is applied between the planes, where $V_a - V_b$ is small compared with the average dc beam voltage in the space between the planes. The electron velocity is assumed to be single-valued over any plane normal to the beam, and the ac component of the electron velocity is small compared with the dc component. The Llewellyn-Peterson equations then take the form:

$$V_b - V_a = A^*J_T + B^*J_a + C^*U_a$$
$$J_b = D^*J_T + E^*J_a + F^*U_a \qquad (7.2\text{-}1)$$
$$U_b = G^*J_T + H^*J_a + I^*U_a$$

where $J_b$ is the ac convection current density at plane $b$, $U_b$ is the ac component of the electron velocity at plane $b$, and $J_T$ is the *total* ac beam current density between the planes, that is, the convection current density plus displacement current density, as discussed in Appendix X. Between any two electrodes $J_T$ is a constant, independent of distance from an electrode. If plane $b$ were an anode, then $J_T$ times the area of the beam would be the ac current flowing in the anode lead, edge effects being neglected. The coefficients $A^*$ to $I^*$ are tabulated in Appendix XI. The coefficients are simple functions of the dc electron velocities at planes $a$ and $b$, the dc transit time for the electrons crossing from plane $a$ to plane $b$ and the corresponding transit angle, and a space charge parameter $\zeta$ which varies from 0 with no space charge to 1 with maximum possible space charge. It should be noted that all quantities in Equations (7.2-1) are phasors.

If plane $b$ is followed by a second "interaction region" bounded by planes $b$ and $c$, as shown in Figure 7.2-1(b), the quantities $J_b$ and $U_b$, together with the corresponding dc quantities, can be used as entrance conditions for the second region. The total current density in the second region, $J_{T2}$, is equal to $J_{T1}$ minus the current per unit area flowing into plane $b$ through an external lead. In terms of $J_b$, $U_b$, and $J_{T2}$, the ac convection current density and electron velocity at the third plane, $J_c$ and $U_c$, can be calculated by further application of Equations (7.2-1).

Now the total current density $J_T$ flowing toward any grid plane or electrode (from both sides) multiplied by the area of the beam equals the ac current flowing away from the electrode in the external lead. Llewellyn

and Peterson used this relationship and the expressions for $J_T$ in the various interaction regions of a multi-electrode tube to construct an equivalent network for such a tube. The network is of particular value in analyzing the high-frequency performance of planar grid-controlled tubes. The equivalent network of a triode operating with the grid negative with respect to the cathode is shown in Figure 7.2-2. Llewellyn and Peterson show that under usual operating conditions the effects of space charge in the region between the control grid and anode can be neglected, and in this case the admittance $y_{22}$ shown in the figure is simply the free-space capacitance $C_{22}$ between the anode and a conducting plane coincident with the grid plane. The capacitance $C_g$ is equal to $\mu C_{22}$, where $\mu$ is the amplification factor of the tube. (If the grid were an ideal electrostatic shield, $\mu$ would be infinitely large, and $C_g$ could be replaced by a direct connection from the grid terminal to the central node of the network.) The admittance $y_{11}$ is the admittance of a space-charge-limited planar diode of spacing equal to the cathode-grid spacing and applied dc voltage equal to the effective beam voltage at the grid. It is the reciprocal of the impedance $z$ given by Equation (7.1-15). At low frequencies the transadmittance $y_{12}$ is approximately equal to minus the low-frequency transconductance of the tube. However, at higher frequencies, it is modified to take into account the effects of the finite transit angles in the cathode-grid region and the grid-anode region. In Reference 7a, Figure 8 shows the effect of frequency on the phase and magnitude of $y_{12}$.
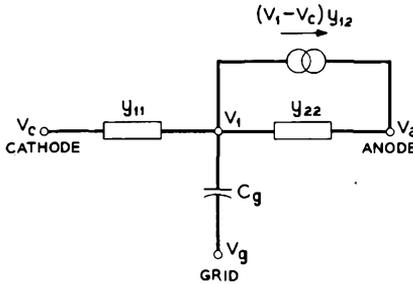


FIG. 7.2-2 The equivalent network for a triode with grid negative with respect to the cathode, as derived by Llewellyn and Peterson.

Some consequences of Llewellyn's and Peterson's work are as follows:

1. The three results presented in Section 7.1 can be obtained directly by the application of Equations (7.2-1) to the particular problems considered. However, the equations are of much more general applicability in the sense that they can be used to solve a variety of similar problems with different dc electrode voltages and different amounts of space charge in the beam.

2. At high frequencies the effects of non-zero transit angles in the various interaction regions of a grid-controlled tube can be evaluated by examination of the appropriate equivalent network.

3. When the transit angles in the input region of a grounded-cathode

amplifier are significantly greater than zero, the electron beam causes a loading of the input circuit. (See Section 7.3(a) for further discussion of this effect.) An expression for this beam loading was derived by Llewellyn and Peterson using the appropriate equivalent network. The expression indicates that at frequencies at which the beam loading first becomes important, the conductive part of the loading is approximately proportional to the product of the low-frequency transconductance and the square of the signal frequency. The derivation assumes that the electrons are emitted from the cathode with zero velocity and hence that the potential minimum coincides with the cathode.

## 7.3 Input Admittance

At high frequencies a principal limitation of grid-controlled tubes when operated as grounded-cathode amplifiers or oscillators arises from conductive loading of the input circuit. Two effects contribute to the loading: one results from inductance in the cathode lead, and the second results from beam loading. Both contribute a conductive loading which is approximately proportional to the product of the transconductance of the tube and the square of the signal frequency. We consider the effects under separate headings below.

### (a) Lead Inductance Effects

Although the distance from the socket to the cathode electrode of most grid-controlled tubes is small, often only 1 or 2 cm, the inductance associated with this length of lead can give rise to an important loading of the input circuit. Figure 7.3-1 shows schematically a grid-controlled tube in which the cathode lead inductance is represented by a lumped inductance within the tube. We shall assume that the impedance between the socket and ground is negligible, and we shall neglect the loading which results from the non-zero transit angle in the input region of the tube. The phasor corresponding to the input signal $v_i$ is then given by

$$V_i = V_g + j\omega L_c(I_a + I_1) \quad (7.3-1)$$



Fig. 7.3-1 Schematic representation of grid-controlled tube with cathode-lead inductance.

where $V_g$ is the phasor corresponding to the ac voltage between the cathode and grid electrodes, $I_a$ is the phasor corresponding to the current $i_a$ flowing

in the anode circuit, and $I_1$ is the phasor corresponding to the current $i_1$ flowing in the input circuit. If the load resistance in the anode circuit is small compared with the dynamic anode resistance, the current flowing in the anode circuit can be expressed as

$$I_a = g_m V_g \qquad (7.3\text{-}2)$$

where $g_m$ is the transconductance of the tube. If we assume that $I_1 \ll I_a$, Equation (7.3-1) can be written as

$$V_i = V_g(1 + j\omega L_c g_m) \qquad (7.3\text{-}3)$$

If the input capacitance of the tube is principally that between the control grid and cathode, the voltage $V_g$ is related to the current flowing in the input circuit by

$$V_g = \frac{I_1}{j\omega C_{cg}} \qquad (7.3\text{-}4)$$

where $C_{cg}$ is the capacitance between the control grid and cathode. Combining Equations (7.3-3) and (7.3-4), we can express the input admittance as

$$Y_i = \frac{I_1}{V_i} = \frac{j\omega C_{cg}}{1 + j\omega L_c g_m} = \frac{j\omega C_{cg}(1 - j\omega L_c g_m)}{1 + \omega^2 L_c{}^2 g_m{}^2} \qquad (7.3\text{-}5)$$

At frequencies of interest the term $\omega^2 L_c{}^2 g_m{}^2$ in the denominator is small compared with unity and can be neglected. The input admittance then has a positive real part given by

$$G_i = \omega^2 L_c C_{cg} g_m \qquad (7.3\text{-}6)$$

This conductance shunts the input circuit, and since it is proportional to the product of $L_c$ and $C_{cg}$, it is desirable that both of these quantities be as small as possible in tubes used at very high frequencies.

In "miniature" tubes, such as the 403A/6AK5 pentode, described in Section 5.4, low cathode-lead inductance is achieved because of the short distance between the internal electrodes and the pins at the base of the tube. Furthermore, two cathode leads are used in the 403A/6AK5, each connected to separate pins at the tube base. This permits parallel connection at the tube socket and results in a further reduction in the cathode-lead inductance. In the 403A/6AK5, $L_c = 5$ millimicrohenries with parallel connection to the cathode, $C_{cg} = 3.7$ pf (i.e., $3.7 \times 10^{-12}$ farad), and $g_m = 5 \times 10^{-3}$ mho. Using these values in Equation (7.3-6), we find that at a frequency of 100 Mc, $G_i = 3.6 \times 10^{-5}$ mho. This is equivalent to a shunting resistance of 28,000 ohms at the input of the tube. This resistance is paralleled by the beam-loading conductance discussed in Part (b) below.

In the 448A tetrode, described in Section 5.3, higher transconductance is achieved by the use of a larger cathode area than in the 403A/6AK5 and

a closer cathode-control grid spacing. This increases $C_{cg}$, and consequently it is even more important to have a low cathode-lead inductance in this tube. For this reason, three separate cathode leads are brought out of the 448A for parallel connection at the base. Measurement of the cathode-lead inductance of the 448A with the three cathode leads in parallel indicates that it is about 4 millimicrohenries. The capacitance $C_{cg}$ is 18 pf, and $g_m$ is 0.034 mho. Substituting these values into Equation (7.3-6), we find that $G_i = 9.7 \times 10^{-4}$ mho at 100 Mc. This is equivalent to a shunting resistance of 1030 ohms across the input circuit.

A second important lead inductance effect in a tube such as the 448A is the fact that series resonance can occur between the inductance of the cathode and control-grid leads and the capacitance between the cathode and control grid. If we assume a total of 10 millimicrohenries inductance in the cathode and control-grid leads in the 448A, series resonance with the 18-pf capacitance between the control grid and cathode occurs at a frequency of 375 Mc. At this frequency the input impedance would be reduced to a very small value.

### (b)  Beam Loading in Grounded-Cathode Stages

The beam loading of the input circuit also has a conductive part which at low and moderate frequencies is proportional to the product of the transconductance and the square of the signal frequency. The discussion that follows explains why this is so. Consider a single electron that travels from the cathode to the anode of a grounded-cathode triode. An induced current somewhat like that shown in Figure 7.3-2(b) flows in the external circuit connected between ground and the grid electrode. At time $t_o$ the electron leaves the cathode. At time $t_1$ it passes the grid, and the direction of the induced current reverses because the direction of the electron velocity relative to the grid plane reverses. At time $t_2$ the electron strikes the anode. (If the tube were a tetrode or pentode, the time $t_2$ could be taken as the time at which the electron passes the plane of the screen grid.) The area under the part of the induced current curve from $t_o$ to $t_1$ equals the area under the part from $t_1$ to $t_2$, and if the grid were an ideal electrostatic shield, each area would equal the electronic charge $e$. At low frequencies, and with many electrons passing the grid per cycle, the current induced in the grid circuit by electrons crossing from the cathode to the grid plane is just balanced by the induced current caused by the electrons crossing from the grid plane to the anode, and there is no net current induced in the grid circuit. However, at higher frequencies, where the electron transit time $T_o = t_2 - t_o$ is a significant part of the period of the ac signal, the induced currents resulting from electrons crossing the two regions may not be exactly 180° out of

phase. As a result, current $i_g$ flows in the grid circuit when a voltage is applied by the signal generator to the grid terminal. The ratio of this ac induced current to the applied signal is called the beam loading admittance.
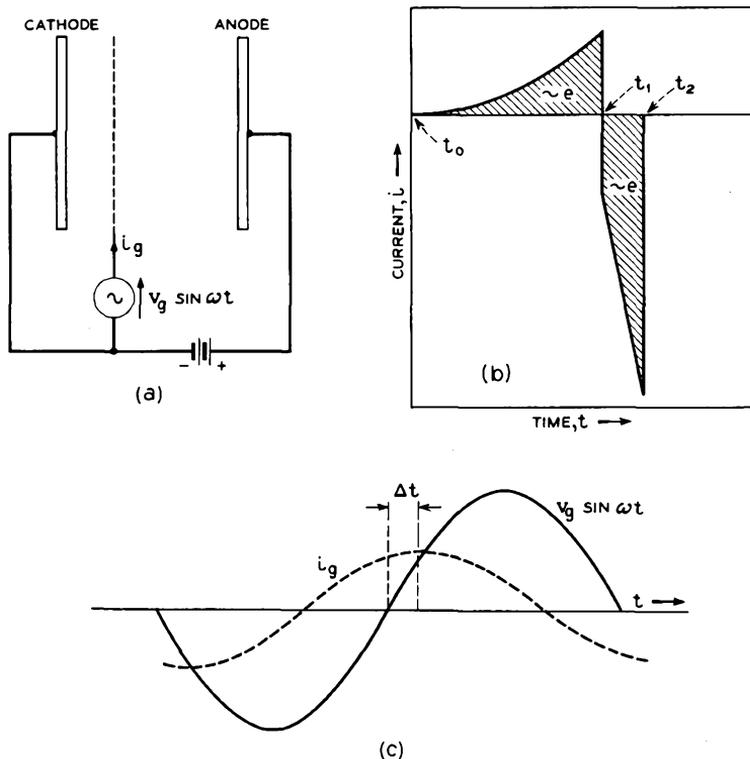


FIG. 7.3-2 The induced current $i$ shown in part b flows in the grid lead when a single electron travels from the cathode to the anode. $i_g$ is the total induced current from all the electrons.

Suppose the ac grid voltage is given by $V_g \sin \omega t$. The ac beam current is then $g_m V_g \sin \omega t$, where $g_m$ is the transconductance of the tube. During the half cycle when the grid voltage is increasing, the current of electrons leaving the cathode (or really the current leaving the potential minimum) is also increasing. We would expect that during most of this half cycle the current induced in the grid circuit by electrons crossing from the cathode to the grid plane would exceed that resulting from electrons crossing from the grid plane to the anode, whereas during the other half cycle the opposite would be the case. Consider the instant when $\sin \omega t = 0$,

and the ac part of the grid voltage is changing from negative to positive. The grid voltage is increasing most rapidly at this instant, and likewise the current of electrons leaving the cathode is increasing most rapidly. Because of the finite transit time $t_1 - t_o$ required for electrons to cross the cathode-grid region, the number of electrons in transit in this region is not maximum at the same instant as the current leaving the cathode is a maximum, but a short time later. Furthermore, the current induced in the grid circuit by an individual electron is essentially zero as the electron leaves the cathode (because the electron velocity is nearly zero), but it increases as the electron approaches the grid plane. For these reasons, *the net current induced in the grid circuit is not maximum when the rate of change of the current leaving the cathode is maximum, but a short time $\Delta t$ later, where $\Delta t$ is an appreciable fraction of $t_1 - t_o$.* Of course, similar reasoning applies to the grid-anode region of the tube and the current induced in the grid circuit by electrons crossing this region. Consequently, the time $\Delta t$ is actually a function both of $t_1 - t_o$ and $t_2 - t_1$.

From the foregoing discussion, we would expect that the magnitude of the induced grid current $i_g$ would be proportional to the maximum rate of change of the beam current and hence proportional to the product of $g_m V_g$ and the angular frequency $\omega$. Also, for transit times which are small compared with the period of the ac signal, doubling the transit time in the two regions of the tube also doubles the difference between the induced grid currents resulting from electrons crossing the two regions of the tube, and the net induced grid current $i_g$ is doubled. Thus $i_g$ is also proportional to the transit time $T_o$ and can be expressed as

$$i_g = K g_m V_g \omega T_o \cos \omega(t - \Delta t) \qquad (7.3\text{-}7)$$

where $K$ is a constant. This can be expanded to give

$$i_g = K g_m V_g \omega T_o [\cos \omega t \cos \omega \Delta t + \sin \omega t \sin \omega \Delta t] \qquad (7.3\text{-}8)$$

The second term in brackets is in phase with the applied grid signal, so that the ratio of this term to the grid voltage $V_g \sin \omega t$ is the conductive part of the input loading. Hence,

$$G_{in} = K g_m \omega T_o \sin \omega \Delta t \qquad (7.3\text{-}9)$$

If the electron transit time is small compared with the period of the ac signal, $\omega \Delta t$ is a small angle, and to a first approximation the sine of the angle can be replaced by the angle. Thus

$$G_{in} \propto g_m \omega^2 T_o \Delta t \qquad (7.3\text{-}10)$$

It is significant that both the input conductance resulting from cathode-lead inductance and the input conductance resulting from beam loading are proportional to $g_m \omega^2$. Some experimental measurements of the input

loading with grounded-cathode operation are described in Section (c) below.

The input conductance of a *grounded-grid* amplifier stage is high even at relatively low frequencies. From Figure 7.2-2 and the accompanying discussion it can be seen that the input admittance with grounded-grid operation is essentially $y_{11}$ in series with $C_g + C_{22} = C_{22} (\mu + 1)$. (We assume the anode is bypassed to ground.) At moderate frequencies the reactance of $C_{22} (\mu + 1)$ is negligible, and the input admittance of the amplifier is $y_{11}$. Now $y_{11}$ is the admittance of a space-charge-limited diode of the same electrode area as the triode, with current density equal to the average beam current density, electrode spacing equal to the cathode-grid spacing, and applied anode voltage equal to the effective beam voltage at the grid. The variation with frequency of the conductive and capacitive parts of this admittance are plotted in Figure 7.1-7. The input conductance of a grounded-grid amplifier stage, therefore, varies as the conductive term plotted in this figure. (We should qualify this statement by noting that the derivation which led to the conductive term plotted in Figure 7.1-7 assumed that the electrons leave the cathode with zero velocity and hence that the potential minimum coincides with the cathode. Actually, it is probable that the electrons which travel part way out to the potential minimum and return to the cathode contribute significantly to the input loading. Furthermore, the derivation does not take into account the Maxwellian distribution of emission velocities, and this also must have an important effect.)

(c)  *Some Measurements of the Input Admittance of Grounded-Cathode Amplifier Stages*

When the beam current of a tube is varied by changing the control-
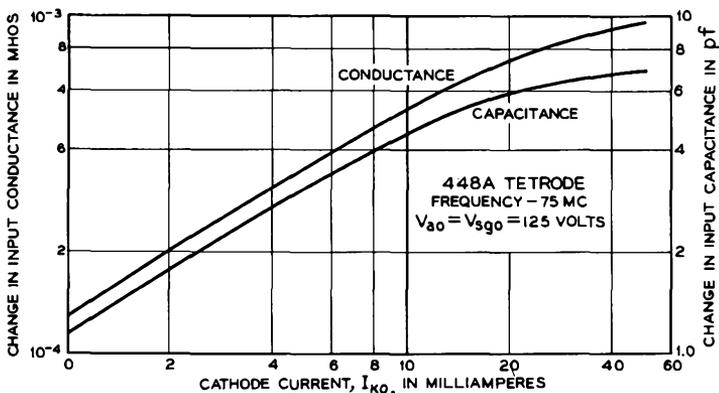


FIG. 7.3-3 Change of input conductance and capacitance with cathode current for the 448A tetrode.

grid voltage, both the input capacitance and input conductance vary. If the control grid is biased negatively to cut off the beam, the input capacitance results only from the interelectrode capacitance plus the capacitance between the leads to the electrodes, and the input conductance
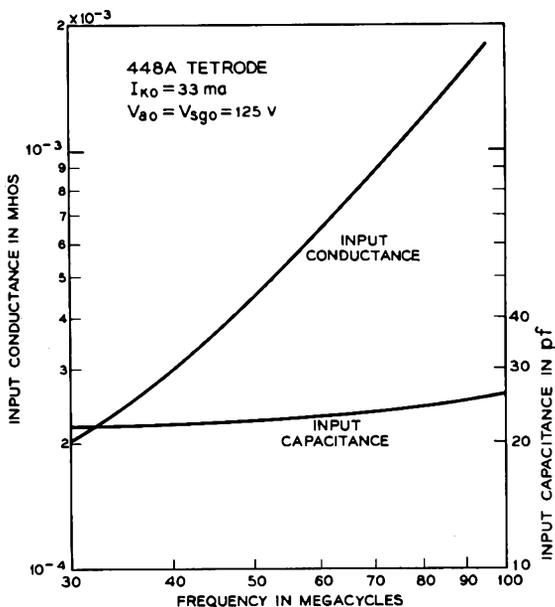


Fig. 7.3-4 Input conductance and input capacitance vs. frequency for the 448A tetrode.

results only from the effects of the series resistance in the leads to the electrodes. However, as the beam current is increased, both the input capacitance and input conductance increase. The increase in input capacitance results from beam loading, whereas the increase in input conductance results from both beam loading and cathode-lead inductance.

Figure 7.3-3 shows the change in input conductance and change in input capacitance vs. beam current for the 448A tetrode, described in Section 5.3. The measurements were made at a frequency of 75 Mc. When the beam is cut off, the input conductance is $1.5 \times 10^{-4}$ mho, and the input capacitance is 18 pf. Under normal operating conditions, the cathode current is about 35 ma. At this cathode current the input conductance is about $1.03 \times 10^{-3}$ mho. Figure 7.3-4 shows the variation of input conductance and capacitance with frequency for the 448A. At about 70 Mc the input conductance increases approximately as the square of the signal frequency,
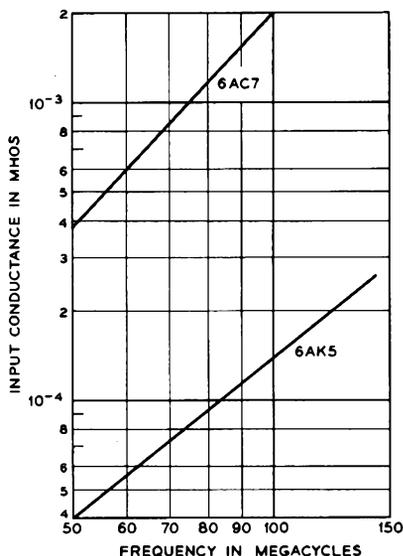
FIG. 7.3-5 Input conductance vs. frequency for the 6AC7 and 6AK5 (Courtesy Radio Corporation of America.)

whereas at lower frequencies it does not increase as rapidly. Figure 7.3-5 shows plots of the input conductance vs. frequency for the 6AC7 and 6AK5 codes. The 6AK5 is described in Section 5.4.

## 7.4 Two Grid-Controlled Tubes for High-Frequency Amplification

From the discussion given in the previous sections and in Chapter 6, we can list several important electrical and physical characteristics needed in a grid-controlled tube which provides high-gain, broadband amplification at frequencies above 100 Mc/sec:

1. The tube must have a high gain-bandwidth product and hence a high transconductance and low input and output capacitances. This means that the transconductance per unit area of the electrodes must be high, and a short cathode-control grid spacing must be used.

2. The transit angle in the cathode-control grid region must be small so that the input loading (for grounded-cathode operation) is small. The transit angle in the cathode-control grid region can be made small by using a short spacing between the control grid and cathode and a relatively high average voltage at the grid plane and hence a high cathode current density.[2]

3. High voltages and not-too-large electrode spacings must be used at the output interaction gap to reduce the transit angle and keep the beam-coupling coefficient near to unity.

4. The cathode-lead inductance must be small to reduce the input loading, and other lead inductances must be small to prevent the occurrence of series resonance with the interelectrode capacitances.

---

[2]Equation (2) of Appendix X shows that the electron transit time for a planar diode (with potential minimum coincident with the cathode) is proportional to $d^{1/3}/J_o^{1/3}$, where $d$ is the electrode spacing, $J_o$ is the cathode current density. Consequently, the use of a high cathode current density and short cathode-control grid spacing reduces the input transit angle.

Figure 7.4-1 shows the construction of the Bell Telephone Laboratories 1983 tetrode, a developmental tube designed for operation at frequencies of several hundred megacycles per second. The cathode, control grid,
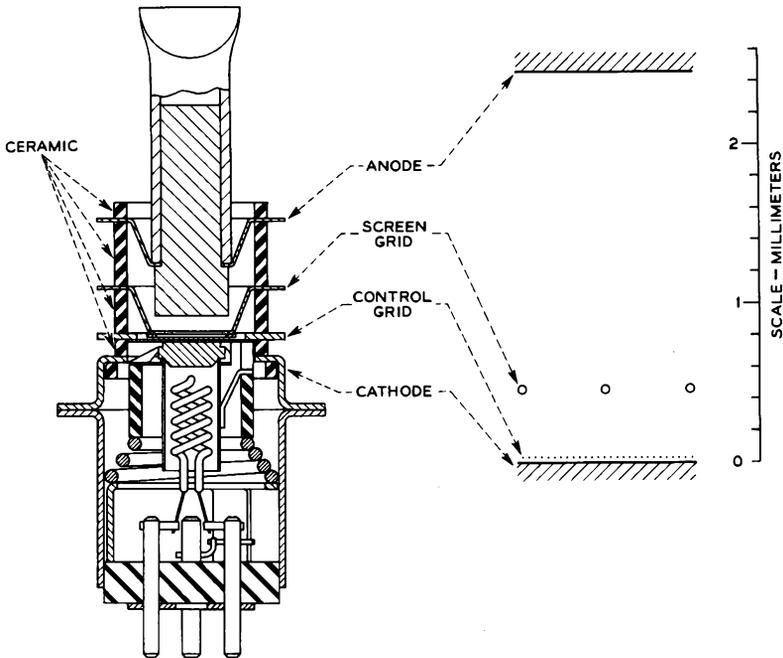


FIG. 7.4-1 The construction of the Bell Telephone Laboratories 1983 tetrode. The overall height of the tube is 5.1 centimeters.

screen grid, and anode connections are brought out radially from the electrodes by means of "disc" leads which are spaced by ceramic rings. The heavy slanted lines in the figure indicate the cross section of the ceramic rings. Vacuum-tight seals are formed between the ceramic rings and the disc leads. This type of envelope construction offers several important advantages:

1. The lead inductance can be made extremely small.

2. At frequencies above 100 Mc/sec, rf losses in wire leads to electrodes become important, and the effects of these losses increase with signal frequency. By using disc leads, the surface area of the leads is greatly increased, and the losses are correspondingly reduced.

3. Power dissipated as heat in the screen grid and anode can be effectively

conducted away, thus increasing the permissible power dissipation within the tube.

Table 5.2-1 lists the operating conditions and important performance characteristics of the 1983 tetrode. A comparison with the 448A tetrode data shows that the high gain-bandwidth product is achieved at the expense of higher cathode current density and reduced cathode-control grid spacing. The transconductance per unit area of the cathode in the 1983 is seven times that of the 448A.

Much more power can be dissipated in the anode of the 1983 than in the anode of the 448A, because the thermal conduction from the anode to the external connections is much better in the 1983. A higher anode voltage can therefore be used, and this helps to reduce the transit angle in the output gap. Also, a smaller spacing between the screen grid and anode is used — 2 mm for the 1983 compared with 5 mm for the 448A. By operating the anode of the 1983 at a voltage well above that of the screen grid, secondary electrons emitted from the anode are prevented from reaching the screen grid. However, part of the primary current to the anode results from secondary emission at the screen grid. To reduce this secondary emission, and in fact to reduce the primary current to the screen grid, a larger screen-grid pitch is used in the 1983, about twice that of the 448A, the screen-grid wire diameter being the same in the two tubes. The heat dissipated in the anode and screen grid of the 1983 is carried away partly by conduction through the external connections to the leads and partly by forced-air cooling.

At the time of writing, a one-stage, grounded-cathode amplifier has been assembled using the 1983. The amplifier has a 50-ohm resistance connected between the control grid and cathode, and a signal generator with a 50-ohm internal impedance is used to drive the amplifier. The amplifier provides a midband gain of 10 db with a 3-db bandwidth extending from 0.5 Mc/sec to 250 Mc/sec. (Note that because of the low input impedance of the stage, only the output capacitance limits the bandwidth in such a stage. Consequently, the gain-bandwidth product given by Equation (6.4-11) is not applicable here.)

The second tube which we shall describe is the Western Electric 416B triode, a tube designed as a grounded-grid amplifier for signal frequencies in the neighborhood of 4000 Mc/sec. The construction of the 416B is illustrated in Figure 7.4-2. The envelope is a metal-and-glass structure in which glass rings separate the anode, grid, and cathode terminals and form vacuum-tight seals with these terminals. The input and output connections to the tube are made by means of waveguides.[3] A cross-sectional view of

---

[3] See Chapter 8.

the associated waveguide components is shown in Figure 7.4-3. The lower part of the bulb is capacitively coupled to the cathode, and rf connection to the cathode is made through this capacitance. DC connection to the cathode is made through a pin in the base. The electrical characteristics of the 416B are summarized in Table 5.2-1.
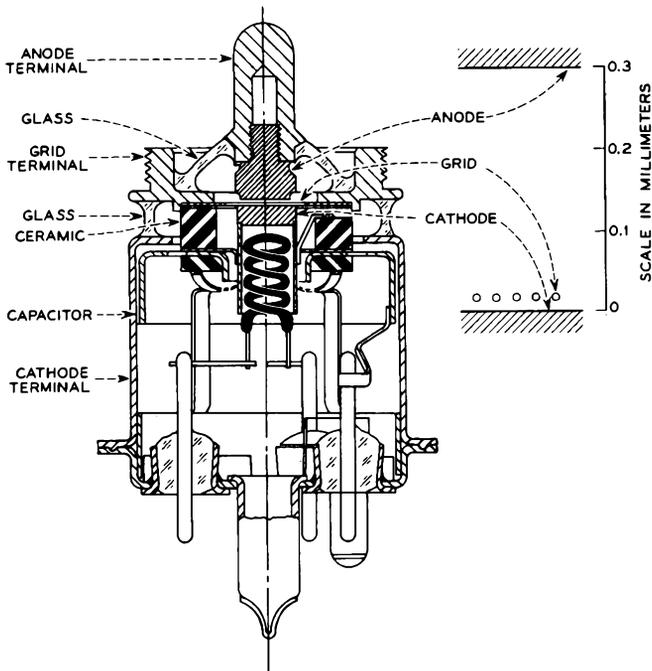


FIG. 7.4-2 The construction of the Western Electric 416B triode. The overall height of the tube is 4.8 cm.

The principal design considerations which led to the electrode structure of the 416B are described in Reference 7c. The design was chosen to obtain as large as possible a product of (midband power gain) × (bandwidth) consistent with practical values of cathode current density and anode and grid power dissipation. (At signal frequencies of 4000 Mc/sec it is more meaningful to use a gain-bandwidth product involving the power gain rather than the voltage gain, since the power gain can be measured directly, whereas the voltage gain must be calculated from measurements of power gain.) Using the equivalent network for a triode given in Figure 7.2-2, the midband power gain with grounded-grid operation can easily be shown to be $|y_{12}|^2/2G_{11}G_o$, where $G_{11}$ is the real part of $y_{11}$, and $G_o$ is the conductance
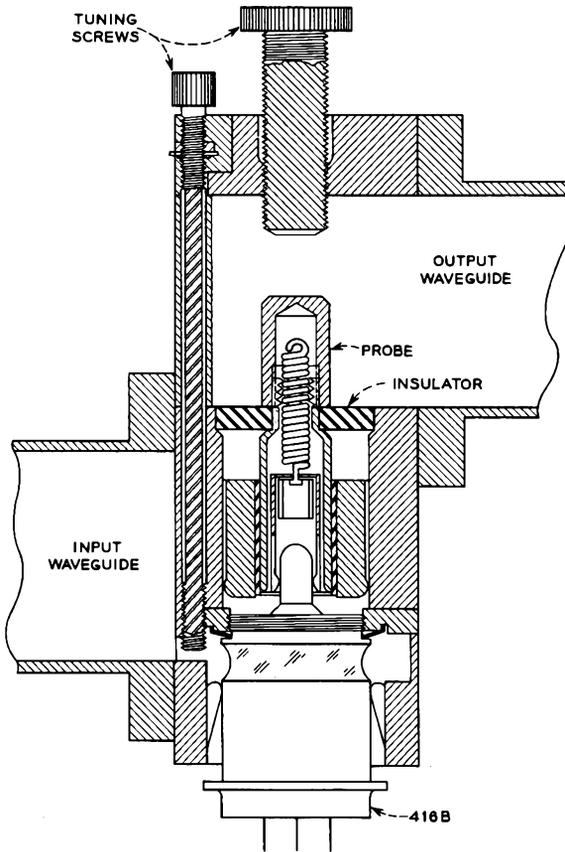
FIG. 7.4-3 The waveguide connections to the 416B. There is effectively a coaxial line from the anode terminal of the tube to a probe which extends into the output waveguide and which acts as a transducer between the coaxial line and the output waveguide.

of the output circuit. We assume here that the characteristic admittance of the output waveguide appears at the tube as an admittance $G_o/2$ and that the losses in the output circuit of the amplifier are adjusted to match this admittance. Hence the total conductance shunting the output circuit is $G_o$. If it is further assumed that the losses in the input circuit contribute a shunting admittance at the input which is small in comparison with $G_{11}$, and hence the effect of the input losses can be neglected. The capacitance $C_g$ in Figure 7.2-2 is assumed to have negligible reactance at 4000 Mc/sec.

Since the input conductance of a grounded-grid stage is extremely high, the $Q$ of the input circuit is correspondingly low, and the bandwidth of the stage is determined primarily by the output circuit. From Equations (6.4-3) and (6.4-7) the 3-*db bandwidth* of the output circuit is $G_o/2\pi C_o$, where $C_o$ is the capacitance shunting the output circuit. Thus

$$\text{(midband power gain)} \times \text{(3-db bandwidth)} = \frac{|y_{12}|^2}{4\pi G_{11}C_o} \qquad (7.4\text{-}1)$$

The power gain decreases with increasing transit angle in the cathode-grid region, and accordingly a high cathode current density and extremely small cathode-grid spacing are used to minimize the input transit angle.
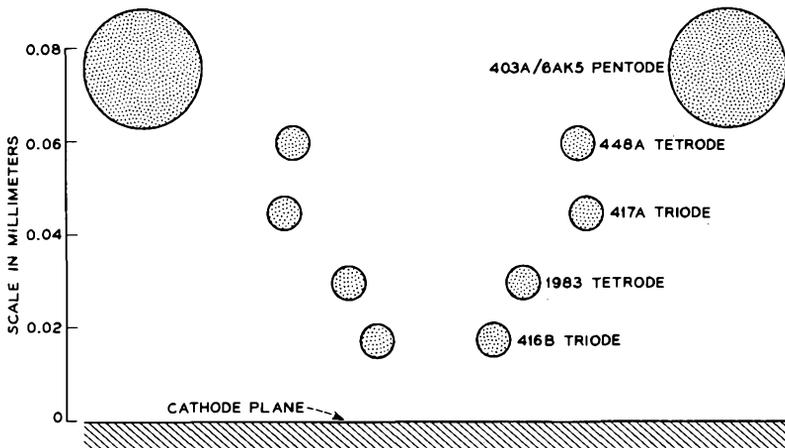


FIG. 7.4-4 The grid-wire diameter, the grid pitch, and the cathode-grid spacing for several tubes described in this chapter and in Chapter 5 are compared. Only two grid wires from each code are shown.

Figure 7.4-4 shows a comparison of the cathode-grid spacing used in this tube and in several tubes described earlier. The cross section of two grid wires from each tube are shown in relation to a common "cathode plane." The 416B has the same cathode area as the 1983, but a smaller cathode-grid spacing.

Several factors bearing upon the choice of grid-anode spacing in the 416B are:

1. The bandwidth $G_o/2\pi C_o$ can be increased by increasing the grid-anode spacing and hence reducing $C_o$.

2. Increasing the grid-anode spacing with a fixed anode voltage increases the transit angle for the grid-anode region and reduces the power gain.

3. Increasing the anode voltage to reduce the transit angle in the grid-anode region increases the power dissipation in the envelope structure. In practice, the permissible anode dissipation is limited to about 6 watts because of the proximity of the metal-glass seal which joins the anode terminal to the glass ring that surrounds it.

4. Increasing the grid-anode spacing with a fixed anode voltage requires a higher average voltage at the grid plane in order to draw the required cathode current density and total anode current. However, the peak grid voltage cannot be permitted to go very far positive, or the grid interception would become excessive.

In practice the highest possible anode voltage is used consistent with the power handling capabilities of the envelope structure. This permits a reasonably large grid-anode spacing with not too large a transit angle (92 degrees at 4000 Mc/sec) and without excessive grid interception. Forced-air cooling of the anode is used to help conduct away the 6 watts of power dissipated in the anode by the electron beam.

The 416B is used in a three-stage amplifier which provides a small-signal power gain of 9db per stage at a midband frequency close to 4000 Mc/sec. The 3-db bandwidth of a single stage is 100 Mc/sec. As the input signal is increased from zero, the power gain at 4000 Mc/sec remains nearly constant up to a power output of about 20 milliwatts. However, at higher power outputs, the power gain falls with increasing power output; and at a power output of 0.5 watt, the midband power gain of the output stage is reduced to 5db.

*PROBLEM*

1. The beam of a cathode-ray tube passes between two parallel deflection plates of length $d$ in the direction of the electron motion. Show that for small deflections of the beam by an $ac$ signal applied to the deflection plates, the amplitude of the deflection is proportional to the beam coupling coefficient $\dfrac{\sin(\omega T_o/2)}{\omega T_o/2}$, where $T_o$ is the time the electrons spend in traveling the distance $d$, and $\omega$ is the angular frequency of the $ac$ signal applied to the plates. Assume that edge effects at the deflection plates can be neglected and that the field between the plates is uniform at any instant.

*REFERENCES*

7a. F. B. Llewellyn and L. C. Peterson, *Proc. IRE* **32**, 144, 1944.
7b. F. B. Llewellyn, *Electron Inertia Effects*, Cambridge University Press, Cambridge, England, 1943.
7c. J. A. Morton and R. M. Ryder, *Bell System Tech. J.* **29**, 496, 1950.

# Chapter 8

## MICROWAVE COMPONENTS AND CIRCUITS

We have discussed in the previous chapter some of the high-frequency effects which occur in grid-controlled tubes. Lead inductances, inter-electrode capacitances, conductor resistances, beam loading, and electron transit-time effects contribute deleteriously to the performance of high-frequency, grid-controlled amplifiers. As will be noted later in this chapter, conductor resistance losses actually worsen with increasing frequency due to an effect known as "skin effect." Furthermore, as frequency increases, it is possible for lead wires to have lengths comparable with a wavelength, in which case they can act as antennas and radiate electromagnetic energy. These considerations lead one to abandon the wires and lumped components used at lower frequencies and to seek new and more appropriate components for microwave frequencies.

Let us first consider the evolution of the tuned circuit as the resonant frequency is increased into the microwave range. At low and moderate frequencies, lumped-constant resonant circuits, such as the one illustrated schematically in Figure 8-1(a), are frequently used in electronic circuitry. The resonant frequency is given by $f = 1/2\pi\sqrt{LC}$, where $L$ is the induct-ance, and $C$ is the capacitance. In a tetrode amplifier circuit, for example, an inductance $L$ may be used to resonate with the interelectrode and stray capacitances of the output circuit so as to give maximum gain at a partic-ular frequency.

As the operating frequency is increased, both the capacitance and in-ductance could be decreased in order to maintain resonance at the operating frequency. However, for the case of the tetrode amplifier, a limiting value of the capacitance is soon reached for two practical reasons. First, transit-time effects set a limit to how far the electrodes can be pulled apart. Second, the area to which the beam cross section can be reduced may be limited by (1) the maximum allowable cathode current density or (2) considerations
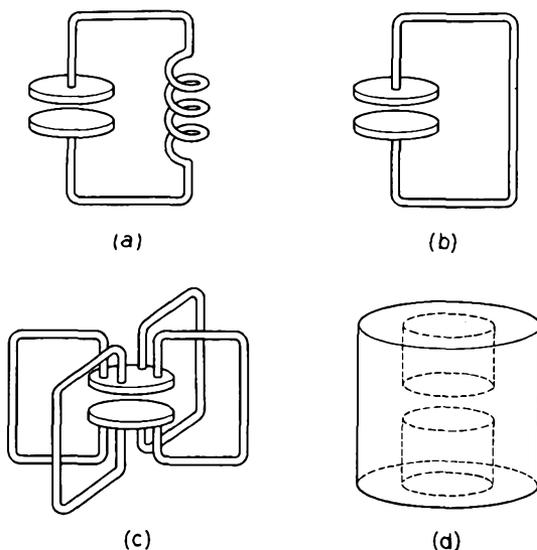
FIG. 8-1 Evolution of a cavity resonator from its low-frequency prototype. (a) Low-frequency prototype. (b) Inductance decreased to that of a single turn wire. (c) Single wires in parallel, reducing the inductance further. (d) Cavity resonator resulting from a continuation of the process of Figure 8-1(c).

relating to beam spreading and confining the beam with external fields, as discussed in Section 3.4. This determines a minimum area for the electrodes. As the frequency is further increased, therefore, one must resort to reducing the inductance. However, we soon reach the point where the inductance is a single short wire, as shown in Figure 8-1(b). Still higher resonant frequencies can be obtained by paralleling the single wire with additional single-wire inductances, as indicated in Figure 8-1(c). As this procedure is carried to the limit, one obtains the re-entrant cavity structure shown in Figure 8-1(d). A cross-sectional view of such a cavity is shown in Figure 8-2. Such resonant cavities are used in klystrons and microwave triodes and tetrodes. Not only has the inductance been decreased in the resonant cavity, but also the resistance losses are lessened, and the self-shielding configuration prevents radiation losses. The fact that all of the electromagnetic fields are confined to the interior of the cavity will become more obvious after a discussion of "skin effect."

To calculate the resonant frequency of a cavity such as that shown in Figure 8-2 is often a difficult process.[1] However, approximate calculations

[1]References 8.1, 8.4.

can often be made to obtain useful
information. For instance, we note
that in the cavity shown in Figure
8-2, the capacitive gap is short com-
pared with its diameter. This cavity
may be thought of as two shorted
coaxial lines joined by a capacitance.
It can be shown that the input im-
pedance to each shorted coaxial line
is given by the expression[2]

$$Z_{\text{in}} = \frac{j}{2\pi}\sqrt{\frac{\mu_o}{\varepsilon_o}} \ln \frac{a}{b} \tan \frac{\omega l}{c} \quad (8\text{-}1)$$

where $c = 1/\sqrt{\mu_o \varepsilon_o}$ is the velocity of
light. The capacitance of the gap is
given by the expression

$$C_g = \frac{\varepsilon_o \pi b^2}{h} \quad (8\text{-}2)$$



FIG. 8-2 Re-entrant cavity resonator
which can be analyzed by simple
transmission line theory. The cavity
is airfilled.

where fringing effects are neglected. At resonance, the inductive reactance
of the two shorted coaxial lines in series is equal in magnitude to the capac-
itive reactance of the gap, but of opposite sign. Hence,

$$\frac{j}{\pi}\sqrt{\frac{\mu_o}{\varepsilon_o}}\left(\ln \frac{a}{b}\right)\tan \frac{\omega l}{c} - j\frac{h}{\omega \varepsilon_o \pi b^2} = 0 \quad (8\text{-}3)$$

The solution to this equation gives the resonant frequency. Rearranging
the equation, we obtain

$$\frac{\omega l}{c} \tan \frac{\omega l}{c} = \frac{hl}{b^2 \ln \dfrac{a}{b}} \quad (8\text{-}4)$$

For the particular set of dimensions given by $a = l$, $h = 0.0318l$, and
$a = 4b$, Equation (8-4) is satisfied by $\omega l/c = 0.571$, and we can scale the
dimensions to suit any frequency. At 3000 Mc, $l$ is equal to 0.91 cm. This
sort of scaling operation is a general property of microwave components.
That is, if we multiply all dimensions by a factor $K$, all frequencies of
interest are divided by $K$.

In the above example, the solution $\omega l/c = 0.571$ is equivalent to saying
that $l$ is 0.0908 wavelength long. It can be shown that averaged over a
cycle, a shorted coaxial line of this length contains 8.84 times as much

---

[2]Reference 8.2.

magnetic stored energy as it contains electric energy. The balance of the electric stored energy appears in the gap, since at resonance the electric and magnetic stored energies are equal in magnitude (but 90 degrees out of time phase). The region outside of the gap is called the inductive region of the cavity, and to a good approximation it can be considered to have only magnetic fields. The cavity can therefore be represented by the equivalent circuit of Figure 8-1(a), where the capacitance $C$ is the gap capacitance, and the inductance is chosen to give the correct resonant frequency.

Since Equation (8-4) contains the tangent function, it has an infinite number of solutions with larger values of frequency. Physically, this corresponds approximately to additional half wavelengths on the coaxial line at higher frequencies. This behavior is typical of all microwave cavities; that is, there are an infinite number of resonant frequencies or modes of oscillation. However, resonant cavities are nearly always operated in the lowest frequency, or dominant, mode because the resistive losses are usually lower in that mode. Resistive losses in the cavity can be represented in the equivalent circuit of Figure 8-1(a) by a parallel resistance of such a value as to give the correct power loss per cycle for a given amount of stored energy.

Next let us consider the problem of transmitting microwave energy from one point to another with as little resistive and radiation losses as possible. Radiation losses can be kept to a minimum by using a suitable form of transmission line, such as a coaxial line, stripline, or a waveguide. Of these types of transmission line, the waveguide is capable of giving minimum attenuation per unit length at a given signal frequency, and it is the most commonly used form of transmission line at microwave frequencies. A study of wave propagation in a waveguide provides a suitable introduction to a discussion of wave propagation along other forms of transmission line such as are used in microwave tubes.

Our principal mathematical tool for studying the transport of electromagnetic energy from one point to another is a set of equations, known as *Maxwell's Equations*. These equations can be used to describe electromagnetic wave propagation in free space, and in principle they can be used to describe electromagnetic wave propagation along an arbitrarily shaped transmission line. We shall first consider the plane electromagnetic wave in free space and then show that electromagnetic wave propagation in a waveguide can be considered as a superposition of two plane electromagnetic waves.

Suppose the direction of propagation of a plane electromagnetic wave is taken to be the $z$ direction. With proper choice of the rectangular coordinate system, the wave consists of an electric field component $E_x$ and a magnetic field component $H_y$, both of which vary sinusoidally in the $z$

direction with wavelength $\lambda_o = c/f$, where $c$ is the velocity of propagation, and $f$ is the frequency of the signal. For a plane wave in free space $H_y$ is related to $E_x$ by $H_y = (\sqrt{\epsilon_o/\mu_o})E_x$. The power density flowing in the $z$ direction is equal to $|E_x||H_y|$ watts per square meter of wave front.

Figure 8-3 shows the field lines associated with two plane waves of equal amplitude. One wave is propagating upwards and to the right with velocity



— MAGNETIC FIELD LINES
⊙ ELECTRIC FIELD LINES COMING OUT OF PAGE
⊗ ELECTRIC FIELD LINES GOING INTO PAGE
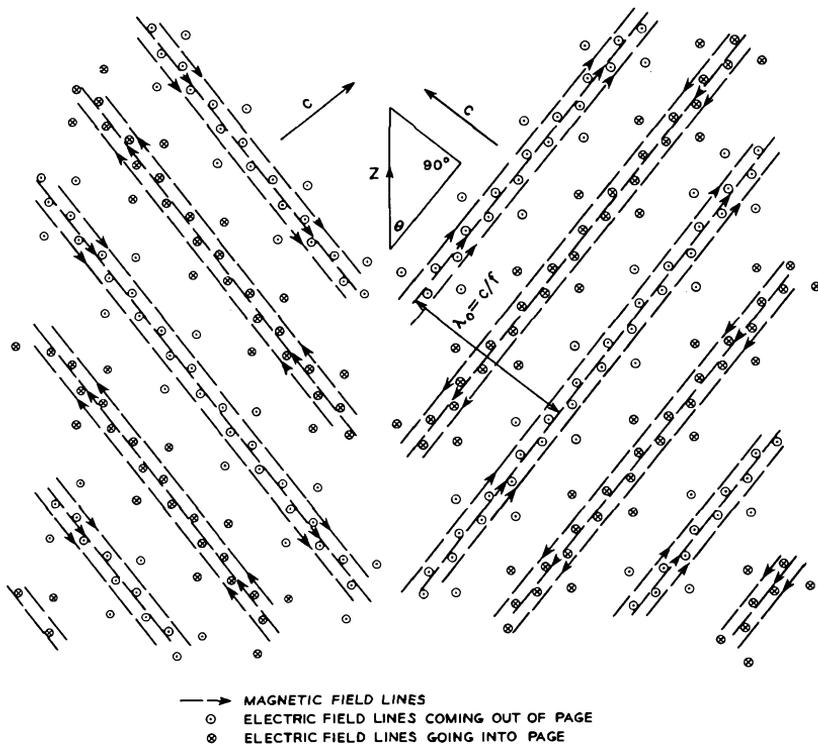
FIG. 8-3 Two plane waves. One is advancing upwards and to the right with velocity $c$, and one is advancing upwards and to the left with velocity $c$.

$c$, and one is propagating upwards and to the left with velocity $c$. Each wave front makes an angle $\theta$ with the vertical, or $z$, direction. Maxwell's Equations are linear, so that the field pattern resulting from a superposition of the two waves is obtained by a vector addition of the individual field components. Figure 8-4 shows the field pattern which results from this vector addition.

The field pattern of Figure 8-4 moves only in the $z$ direction. Examination of the vector diagram shown in the upper right-hand part of the

figure shows that, in the time taken for an individual plane wave to travel the distance $EF$, the field pattern of Figure 8-4 travels in the $z$ direction a distance $EG$. Thus the *phase velocity* of the field pattern in Figure 8-4 is given by $v_p = c/\sin\theta$. On the other hand, the electromagnetic energy associated with the individual plane waves propagates in the direction of
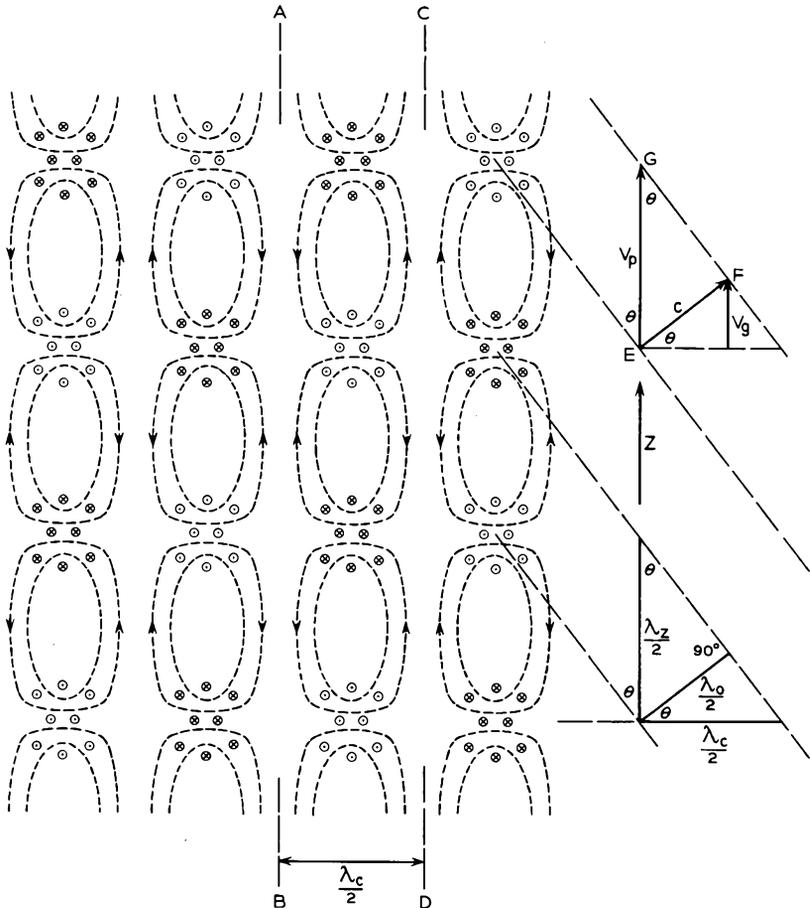


FIG. 8-4 A superposition of the two plane waves shown in Figure 8-3. With increasing time the whole pattern moves in the $z$ direction with *phase velocity* $v_p = c/\sin\theta$. The wavelength in the $z$ direction is given by $\lambda_z = \lambda_o/\sin\theta$. The energy associated with the two separate waves propagates in the direction of travel of the wave fronts of the separate waves. This direction makes an angle of $90° - \theta$ with the $z$ direction. Hence, the energy associated with the above pattern propagates in the $z$ direction with a *group velocity* given by $v_g = c \cos (90° - \theta = c \sin\theta$.

travel of the wave fronts and hence at an angle of $90 - \theta$ degrees with respect to the $z$ direction. Since one wave transports electromagnetic energy to the right and upwards, and the other wave transports electromagnetic energy to the left and upwards, the *net* transport of energy is in the $z$ direction only. The velocity with which the energy is transported in the $z$ direction is given by the *group velocity* $v_g = c \cos (90° - \theta) = c \sin\theta$.

The wavelength $\lambda_z$ of the field pattern in Figure 8-4 is easily seen to be related to the free-space wavelength $\lambda_o$ by $\lambda_z = \lambda_o/\sin\theta$.

From Figure 8-4 it is evident that the electric field intensity is zero along the planes $AB$ and $CD$ at all times, and the magnetic field lines never cross the planes. Hence, if thin conducting sheets were inserted along these planes, the field pattern would not be disturbed. In this case the two plane waves between the conducting sheets reflect from one side to another, at the same time progressing in the $z$ direction. The plane waves outside the conducting sheets are likewise reflected from the conducting sheets, and the net result is that the field pattern of Figure 8-4 is undisturbed.

Next, let us remove the field pattern for a moment and suppose we have two conducting plates of very large area and spaced by the distance from plane $AB$ to plane $CD$ in Figure 8-4. Suppose that two plane waves are launched *between* the plates with the $E$ field parallel to the plates and with the wave fronts making an angle $\theta$ with the surface of the plates. The angle between the two wave fronts is then $2\theta$, as in Figure 8-3. The two waves are reflected from plate to plate, and the resulting field pattern is identical to that shown between the planes $AB$ and $CD$ in Figure 8-4.

Finally, suppose that the two "side plates" of the previous paragraph are joined by "top" and "bottom" plates to form a rectangular waveguide, as shown in Figure 8-5. The electric field lines now terminate on surface charges on the top and bottom plates, but the shape of the field pattern is otherwise unchanged. The waves propagate along the waveguide with phase velocity $v_p = c/\sin\theta$, and the electromagnetic energy propagates with the group velocity given by $v_g = c \sin\theta$. The axial wavelength of the field pattern in the waveguide is given by $\lambda_z = \lambda_o/\sin\theta$.

What we have done here is to find a field pattern that satisfies the boundary conditions imposed by the rectangular waveguide. These boundary conditions require that the tangential component of electric field intensity at the surface of the conducting walls be zero, and the normal component of magnetic field intensity at the surface of the waveguide be zero. From Figure 8-4 it is evident that the distance between planes $AB$ and $CD$ is determined by the angle $\theta$ and the wavelength $\lambda_o$ of the plane waves. Conversely, if we have a waveguide of a given width $a$ and a given wavelength $\lambda_o$, the angle $\theta$ is determined. If we set the width $a$ of the waveguide equal to $\lambda_c/2$, it is evident from Figure 8-4 that $\cos \theta = \lambda_o/\lambda_c$, and hence
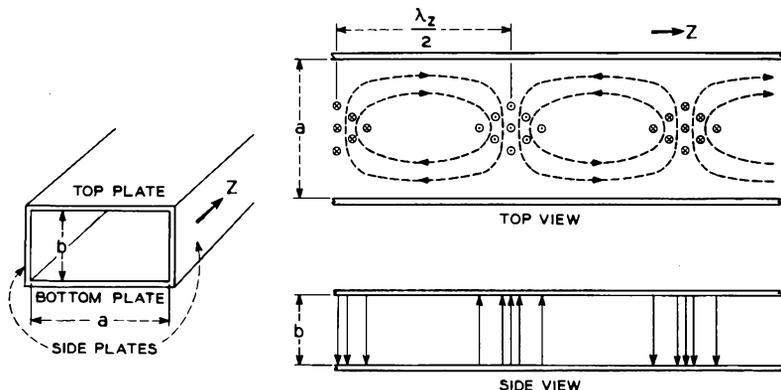
FIG. 8-5 Field patterns in a rectangular waveguide. The broken lines indicate the shapes of the magnetic field lines, and the solid lines indicate the shapes of the electric field lines.

$\sin\theta = \sqrt{1 - \cos^2\theta} = \sqrt{1 - \lambda_o^2/\lambda_c^2}$. Thus we have the relations

$$v_p = \frac{c}{\sqrt{1 - \lambda_o^2/\lambda_c^2}} \tag{8-5}$$

$$v_g = c\sqrt{1 - \lambda_o^2/\lambda_c^2} \tag{8-6}$$

and

$$\lambda_z = \frac{\lambda_o}{\sqrt{1 - \lambda_o^2/\lambda_c^2}} \tag{8-7}$$

So far we have described only the most frequently used dominant mode of propagation of waves in a rectangular waveguide. From Figure 8-4 it is evident that waves in this mode can propagate only if $\lambda_o$ is less than $\lambda_c$. We also note that Equations (8-5) and (8-6) indicate that $v_p$ and $v_g$ become imaginary quantities when $\lambda_o > \lambda_c$. The wavelength $\lambda_c$ is called the cutoff wavelength and is a characteristic of the waveguide and the mode of propagation. Signals of wavelength shorter than $\lambda_c$ can propagate, but signals of longer wavelength cannot propagate. (We assume here that the dimension $b$ of the waveguide is smaller than $a$.)

Finally, let us return once more to Figure 8-4. Suppose the plane $CD$ were translated to the right a distance $\lambda_c/2$. The field pattern between the planes would then consist of two side-by-side patterns similar to the one described above for the dominant mode of the rectangular waveguide. Clearly this field pattern also satisfies the boundary conditions of an *enlarged* waveguide, that is, one twice as wide as we have previously considered. Or, conversely, for a given waveguide width, such a field pattern

can be established in the waveguide provided the free space wavelength of the signal is sufficiently short. In a similar way, we see that an infinite number of field patterns or modes of propagation can be established in a rectangular waveguide. As the field pattern becomes more complex, the cutoff wavelength becomes shorter, and the signal frequency must be higher. Furthermore, since waves can be reflected from the top and bottom of the rectangular waveguide, there is a second infinite set of modes of propagation in which the electric field lines are parallel to the $a$ dimension of the waveguide. Also, for a given signal frequency, a mode with the **E** field parallel to the $a$ dimension can be superimposed on a mode with the **E** field parallel to the $b$ dimension. The resulting field patterns, therefore, can be very complex. Later in the chapter we shall find that there are still other modes of propagation in which the **E** field has a $z$ component and the **H** field is entirely transverse to the $z$ direction.

The final section of the chapter considers the propagation of electromagnetic waves along transmission lines which are characterized by phase velocities that are less than the velocity of light. Such transmission lines form integral parts of a number of microwave tubes, such as traveling-wave tubes, backward-wave oscillators, and magnetrons.

## 8.1 Maxwell's Equations and the Wave Equation

Some of the laws governing the behavior of static electric and magnetic fields were discussed in Chapter 1. The equations from that chapter which are pertinent to our present discussion are listed below.

Equation (1.1-4):
$$\oint_{\text{closed path}} \mathbf{E} \cdot d\mathbf{l} = 0 \qquad (8.1\text{-}1)$$

Equation (1.4-3):
$$\nabla \cdot \mathbf{D} = \rho \qquad (8.1\text{-}2)$$

Equation (1.5-2):
$$\nabla \cdot \mathbf{B} = 0 \qquad (8.1\text{-}3)$$

Equation (1.5-5):
$$\nabla \times \mathbf{H} = \mathbf{J} \qquad (8.1\text{-}4)$$

We shall consider these equations one by one to see what form they take when time-varying fields are present.

First, it may be stated that Equations (8.1-2) and (8.1-3) are true as they stand for time-varying fields and charges as well as for static fields and charges.

Let us next consider Equation (8.1-1) as it applies to a closed loop of resistance wire. The experiments of Faraday have shown that, if the loop is linked by changing magnetic fields, there will be current flow around the loop and hence a voltage drop around the loop. That is to say, the right-

hand side of Equation (8.1-1) is not equal to zero in such a time-varying field. Faraday's law may be stated mathematically as

$$\oint_{\text{closed path}} \boldsymbol{\varepsilon} \cdot d\mathbf{l} = -\frac{\partial}{\partial t} \int_{\text{surface}} \boldsymbol{\mathfrak{B}} \cdot \mathbf{n} dS \qquad (8.1\text{-}5)$$

where the surface of integration is taken as any surface bounded by the closed path. Script letters are used for the time-varying field vectors to distinguish them from the dc vectors used previously. Physically the law states that the total voltage induced in a closed loop is given by the time rate of change of magnetic flux linking the loop. We can convert this equation to a more useful form by applying it to a small loop of area $\Delta A$, the loop being small enough that $\boldsymbol{\mathfrak{B}}$ can be taken as uniform in magnitude and direction. Let the component of $\boldsymbol{\mathfrak{B}}$ normal to the plane of the loop be denoted by $\boldsymbol{\mathfrak{B}}_n$. Dividing both sides of the equation by $\Delta A$ and taking the limit as $\Delta A \rightarrow 0$, we have

$$\lim_{\Delta A \rightarrow 0} \frac{1}{\Delta A} \oint_{\text{closed path}} \boldsymbol{\varepsilon} \cdot d\mathbf{l} = -\frac{\partial \boldsymbol{\mathfrak{B}}_n}{\partial t} \qquad (8.1\text{-}6)$$

But the left-hand side is equal to the component of $\nabla \times \boldsymbol{\varepsilon}$ normal to the plane of the loop, so that

$$\nabla \times \boldsymbol{\varepsilon} = -\frac{\partial \boldsymbol{\mathfrak{B}}}{\partial t} \qquad (8.1\text{-}7)$$

Expressions for the curl in rectangular and cylindrical coordinate systems are given in Appendix XII.

Maxwell's great contribution to these fundamental laws was a recognition of the fact that ac magnetic fields are set up not only by real currents consisting of charges in motion, but also by so-called displacement currents. The displacement current density is given by the time rate of change of the electric flux density vector $(\partial/\partial t)\, \boldsymbol{\mathfrak{D}}$, so that Equation (8.1-4) becomes

$$\nabla \times \boldsymbol{\mathfrak{K}} = \boldsymbol{\mathfrak{g}} + \frac{\partial \boldsymbol{\mathfrak{D}}}{\partial t} \qquad (8.1\text{-}8)$$

If we take the divergence of both sides of Equation (8.1-8), we obtain

$$\nabla \cdot \left( \boldsymbol{\mathfrak{g}} + \frac{\partial \boldsymbol{\mathfrak{D}}}{\partial t} \right) = 0$$

since the divergence of a curl is identically zero. Using Gauss's Law, Equation (8.1-2), this may be written as

$$\nabla \cdot \boldsymbol{\mathfrak{g}} + \frac{\partial \rho}{\partial t} = 0$$

the equation of continuity, Equation (1.3-2). Thus, the introduction of the displacement current density in Equation (8.1-8) is necessary to satisfy the equation of continuity.

We may thus summarize our results for ac and dc fields in the four equations known as Maxwell's equations:

$$\nabla \times \boldsymbol{\varepsilon} = -\frac{\partial \boldsymbol{\mathfrak{B}}}{\partial t} \tag{8.1-9}$$

$$\nabla \times \boldsymbol{\mathfrak{K}} = \boldsymbol{\mathfrak{J}} + \frac{\partial \boldsymbol{\mathfrak{D}}}{\partial t} \tag{8.1-10}$$

$$\nabla \cdot \boldsymbol{\mathfrak{D}} = \rho \tag{8.1-11}$$

$$\nabla \cdot \boldsymbol{\mathfrak{B}} = 0 \tag{8.1-12}$$

Equations (8.1-9) and (8.1-10) when taken together, with conduction current density $\boldsymbol{\mathfrak{J}}$ set to zero, form a very interesting pair. Equation (8.1-9) states that a changing magnetic field gives rise to an electric field, and Equation (8.1-10) in turn states that a changing electric field gives rise to a magnetic field. Thus it is clear how wave propagation and standing-wave phenomena are obtained: each type of electromagnetic field vector acts as a source for the other. A change in one produces the other, and vice versa. Thus, energy oscillates continuously from the electric fields to the magnetic fields and back.

In all our discussions of microwave tubes we shall describe physical behavior for a simple sinusoidal variation at a fixed frequency. In all cases we shall be dealing with linear phenomena, and hence we can represent any arbitrary input or response by a superposition of sinusoidal inputs and responses. We can therefore use the *phasor* notation to describe the currents and field vectors:

$$\boldsymbol{\varepsilon} = \text{Re } \mathbf{E}\epsilon^{j\omega t}, \qquad \boldsymbol{\mathfrak{J}} = \text{Re } \mathbf{J}\epsilon^{j\omega t}, \text{ etc.}$$

Then

$$\frac{\partial \boldsymbol{\varepsilon}}{\partial t} = \text{Re } j\omega \mathbf{E}\epsilon^{j\omega t}, \text{ etc.} \tag{8.1-13}$$

Thus, if all quantities vary sinusoidally at a single frequency, we have the following form of Maxwell's Equations:

$$\nabla \times \mathbf{E} = -j\omega \mathbf{B} \tag{8.1-14}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + j\omega \mathbf{D} \tag{8.1-15}$$

$$\nabla \cdot \mathbf{D} = \rho \tag{8.1-16}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{8.1-17}$$

In rectangular coordinates Equations (8.1-14) and (8.1-15) can be written as

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -j\omega\mu\mu_o H_x$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -j\omega\mu\mu_o H_y$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu\mu_o H_z \qquad (8.1\text{-}18)$$

and

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = J_x + j\omega\varepsilon\varepsilon_o E_x$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = J_y + j\omega\varepsilon\varepsilon_o E_y$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = J_z + j\omega\varepsilon\varepsilon_o E_z \qquad (8.1\text{-}19)$$

where we have substituted $\mathbf{B} = \mu\mu_o\mathbf{H}$ in Equation (8.1-14) and $\mathbf{D} = \varepsilon\varepsilon_o\mathbf{E}$ in Equation (8.1-15).

In addition to Equations (8.1-14) through (8.1-17), it should be noted that two other equations from Chapter 1 are valid for time-varying fields as well as for dc fields. These are the equations for the force on an electron, Section 1.2, and the equation of continuity, Equation (1.3-2), which in phasor notation becomes

$$\nabla \cdot \mathbf{J} = -j\omega\rho \qquad (8.1\text{-}20)$$

Let us now use Equations (8.1-14) and (8.1-15) to derive the wave equation for an electromagnetic wave in a region in which there are no free charges and no conduction currents. In this case the equations reduce to

$$\nabla \times \mathbf{E} = -j\omega\mu\mu_o\mathbf{H} \qquad (8.1\text{-}21)$$

$$\nabla \times \mathbf{H} = j\omega\varepsilon\varepsilon_o\mathbf{E} \qquad (8.1\text{-}22)$$

Taking the curl of both sides of the first equation and combining the result with the second equation, we obtain

$$\nabla \times (\nabla \times \mathbf{E}) = -j\omega\mu\mu_o\nabla \times \mathbf{H} = \omega^2\mu\mu_o\varepsilon\varepsilon_o\mathbf{E} \qquad (8.1\text{-}23)$$

Now from Equation 13 of Appendix XII,

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} \qquad (8.1\text{-}24)$$

The first term on the right here is zero since, from Equation (8.1-16),

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon\varepsilon_o} \qquad (8.1\text{-}25)$$

and we have assumed $\rho = 0$ in the region of space under consideration, Equation (8.1-23) then can be written as

$$\nabla^2\mathbf{E} + k^2\mathbf{E} = 0 \qquad\qquad (8.1\text{-}26)$$

where $k^2 = \omega^2\mu\mu_o\varepsilon\varepsilon_o$. This equation is known as the wave equation for an electric field. Equations (8.1-21) and (8.1-22) also can be used in a similar manner to derive the wave equation for a magnetic field, namely

$$\nabla^2\mathbf{H} + k^2\mathbf{H} = 0 \qquad\qquad (8.1\text{-}27)$$

Equations (8.1-26) and (8.1-27) describe the propagation of an electromagnetic wave in a region of free space in which there are no free charges or conduction currents.

Perhaps the simplest application of Equations (8.1-26) and (8.1-27) is in the description of a plane electromagnetic wave, such as one might obtain at a very large distance from a radiating antenna. Let us assume that the electric field intensity of the wave is directed only in the $x$ direction and is given by $E_x$. For a wave propagating in the $z$ direction, Equation (8.1-26) then reduces to

$$\frac{d^2E_x}{dz^2} + k^2E_x = 0 \qquad\qquad (8.1\text{-}28)$$

This has the solution

$$E_x = E_{xo}\epsilon^{\pm jkz} \qquad\qquad (8.1\text{-}29)$$

Now $k^2 = \omega^2\mu\mu_o\varepsilon\varepsilon_o$, and in free space $\mu = \varepsilon = 1$. We shall set $k = \omega/c$ for a wave in free space, where $c = 1/\sqrt{\mu_o\varepsilon_o}$. If time dependence is included, and if we assume the propagating medium is free space, the expression for the electric field intensity becomes

$$\mathcal{E}_x = \mathrm{Re}\, E_{xo}\epsilon^{j(\omega t \pm kz)} = \mathrm{Re}\, E_{xo}\epsilon^{j\omega[t \pm (z/c)]} = E_{xo}\cos\omega\left(t \pm \frac{z}{c}\right)$$

$$(8.1\text{-}30)$$

Here the plus sign in the term $\cos\omega[t \pm (z/c)]$ applies to a wave propagating in the negative $z$ direction, and the minus sign applies to a wave propagating in the positive $z$ direction. We see that the quantity $c = 1/\sqrt{\mu_o\varepsilon_o}$ is the velocity of propagation of the plane wave, equal to $3 \times 10^8$ meters/sec.

By setting $E_y = E_z = 0$ in Equation (8.1-18), we find that $H_x = H_z = 0$, and

$$H_y = \frac{j}{\omega\mu_o}\frac{\partial E_x}{\partial z} = \frac{k}{\omega\mu_o}E_x = \sqrt{\frac{\varepsilon_o}{\mu_o}}E_x \qquad\qquad (8.1\text{-}31)$$

where we have assumed a wave propagating in the positive $z$ direction and have used a minus sign in the exponent on the right-hand side of Equation

(8.1-29). We have also assumed that $\mu = \varepsilon = 1$. Equation (8.1-31) indicates that *for a plane wave propagating in free space, the ratio of the electric field intensity to the magnetic field intensity is given by* $\sqrt{\mu_o/\varepsilon_o}$. This ratio has the dimensions of an impedance and is numerically equal to 377 ohms.

## 8.2   Energy Stored in Electric and Magnetic Fields; Power Flow in an Electromagnetic Wave

Here we shall first derive expressions for the energy stored per unit volume in electric and magnetic fields. The expressions apply to both static and time-varying fields.[3]

### (a)   Electric Fields

Consider a capacitor that is charged to a voltage of $v$ volts. If an incremental amount of charge $dq$ is added to the charge already on the capacitor, the work done in adding the incremental charge is $vdq$. This work is converted to energy stored in the electric field of the capacitor. Now from Equation (1.4-7) we have $q = Cv$, and hence $dq = Cdv$. Thus the work done in adding the charge $dq$ to the capacitor is $Cvdv$. If the capacitor is charged from zero volts to $v$ volts, the energy stored in the electric field of the capacitor is given by

$$\text{energy stored} = \int_{o}^{v} Cvdv = \tfrac{1}{2}Cv^2 \qquad (8.2\text{-}1)$$

If the capacitor is a parallel-plate device in which the plates are of area $A$ and spacing $d$, and if edge effects are neglected, $C = \varepsilon\varepsilon_o A/d$, and the energy stored per unit volume between the plates is $\tfrac{1}{2}Cv^2/Ad = \tfrac{1}{2}\varepsilon\varepsilon_o(v/d)^2$. Setting $v/d = \mathcal{E}$, where $\mathcal{E}$ is the electric field intensity between the plates, we obtain

$$\text{energy stored per unit volume} = \tfrac{1}{2}\varepsilon\varepsilon_o\mathcal{E}^2 \qquad (8.2\text{-}2)$$

We see that the expression for the energy stored per unit volume depends only on the magnitude of the electric field intensity and is independent of the geometry of the electrodes that generate the field.

### (b)   Magnetic Fields

Equation (8.1-5) indicates that the voltage induced in a loop of resistance wire by a changing magnetic field is equal to the time rate of change of the magnetic flux linking the loop. Consider a toroidal coil, such as that shown in Figure 1.5-3. If the coil has $N$ turns and all are linked by the flux $\phi$, the

---

[3]References 8a, 8c, 8d.

voltage induced in the coil by a changing $\phi$ is given by

$$v = N\frac{d\phi}{dt} \tag{8.2-3}$$

Now the inductance $L$ of the coil is equal to the number of flux linkages per ampere of current passing through the coil. Hence $N\phi = Li$, and

$$Nd\phi = Ldi \tag{8.2-4}$$

Equation (8.2-3) then can be written as

$$v = L\frac{di}{dt} \tag{8.2-5}$$

This equation states that when the current through the coil is changing, there is a voltage $v$ developed across the coil proportional to the rate of change of the current. (The resistive losses in the coil are neglected here.)

The rate at which work is being done to change the current in the coil is $vi$. Thus the work done in an increment of time $dt$ during which the current changes by $di$ is $vidt = Lidi$. This work is converted to energy stored in the magnetic field of the coil. The total energy stored in the magnetic field when the current in the coil is increased from zero to $i$ is then

$$\text{energy stored} = \int_0^i Lidi = \tfrac{1}{2}Li^2 \tag{8.2-6}$$

In the case of the toroidal coil shown in Figure 1.5-3, $L = \pi r^2 \mu \mu_o n N$, where $r$ is the radius of the individual turns of wire, $\mu$ is the relative permeability of the medium filling the coil, $n$ is the number of turns per unit length around the coil, and $N$ is the total number of turns in the coil. If $R$ is the mean radius of the toroid, $n = N/2\pi R$. The volume within which the magnetic energy is stored is approximately given by $(\pi r^2)(2\pi R)$. Hence the energy stored per unit volume within the coil is given by $\tfrac{1}{2}Li^2/(\pi r^2)(2\pi R)$, which reduces to

$$\text{energy stored per unit volume} = \tfrac{1}{2}\mu\mu_o(ni)^2 = \tfrac{1}{2}\mu\mu_o\mathcal{K}^2 \tag{8.2-7}$$

where we have substituted $\mathcal{K} = ni$ from Equation (1.5-11). Again we see that the energy stored per unit volume depends only on the magnitude of the magnetic field intensity and is independent of the field configuration.

### (c)   *Power Flow in an Electromagnetic Wave*

Here we shall examine the power flow associated with a plane wave propagating in free space. We shall assume that the wave propagation is in the $z$ direction and consists of an electric field component $\mathcal{E}_x$ and a magnetic field component $\mathcal{K}_y$. Consider a pillbox element of volume with faces of area $A$

lying parallel to the $x$-$y$ plane and having thickness $dz$ in the $z$ direction. The energy stored in this volume will vary with time as the wave propagates past the volume element. From Equations (8.2-2) and (8.2-7), the instantaneous stored energy in the volume element is given by

$$\mathcal{W} = \tfrac{1}{2}(\varepsilon_o\mathcal{E}_z{}^2 + \mu_o\mathcal{IC}_y{}^2)A\,dz \qquad (8.2\text{-}8)$$

The rate of change of energy stored in the volume element is

$$\frac{\partial\mathcal{W}}{\partial t} = A\,dz\left(\varepsilon_o\mathcal{E}_z\frac{\partial\mathcal{E}_z}{\partial t} + \mu_o\mathcal{IC}_y\frac{\partial\mathcal{IC}_y}{\partial t}\right)$$

$$= -A\,dz\left(\mathcal{E}_z\frac{\partial\mathcal{IC}_y}{\partial z} + \mathcal{IC}_y\frac{\partial\mathcal{E}_z}{\partial z}\right)$$

$$= -A\,dz\frac{\partial}{\partial z}(\mathcal{E}_z\mathcal{IC}_y) \qquad (8.2\text{-}9)$$

where we have substituted $\varepsilon_o(\partial\mathcal{E}_z/\partial t) = -\partial\mathcal{IC}_y/\partial z$ from Equation (8.1-10) and $\mu_o(\partial\mathcal{IC}_y/\partial t) = -\partial\mathcal{E}_z/\partial z$ from Equation (8.1-9).

Thus the time rate of change of the energy stored in the volume element is equal to the change in the quantity $A\mathcal{E}_z\mathcal{IC}_y$ in the distance $dz$. Since energy flows only in the $z$ direction, we see that $\mathcal{E}_z\mathcal{IC}_y$ is of the nature of a *power density*, or rate of flow of energy per unit area. It is customary to represent the power density by a vector $\mathbf{S}$, which is directed in the direction of propagation. In the present case,

$$|\mathbf{S}| = S_z = \mathcal{E}_z\mathcal{IC}_y \qquad (8.2\text{-}10)$$

More generally, whenever there is propagation of electromagnetic energy, the power density can be represented by a vector $\mathbf{S}$ such that

$$\mathbf{S} = \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{IC}} \qquad (8.2\text{-}11)$$

The vector $\mathbf{S}$ is called the *Poynting vector* after the man who discovered it. The power density is measured in watts per square meter.

Equation (8.2-10) can be written in another useful form as follows:

$$S_z = \frac{1}{2}(\mathcal{E}_z\mathcal{IC}_y + \mathcal{E}_z\mathcal{IC}_y) = \frac{1}{2}\left[\sqrt{\frac{\varepsilon_o}{\mu_o}}\mathcal{E}_z{}^2 + \sqrt{\frac{\mu_o}{\varepsilon_o}}\mathcal{IC}_y{}^2\right] = c\left[\frac{1}{2}\varepsilon_o\mathcal{E}_z{}^2 + \frac{1}{2}\mu_o\mathcal{IC}_y{}^2\right]$$

$$(8.2\text{-}12)$$

where we have substituted $\mathcal{E}_z = \sqrt{\mu_o/\varepsilon_o}\ \mathcal{IC}_y$ from Equation (8.1-31) and $c = 1/\sqrt{\mu_o\varepsilon_o}$. This states that the energy stored in the electric and magnetic fields of the plane wave propagates in the $z$ direction with velocity $c$, as we might expect.

## 8.3 Boundary Conditions

Maxwell's Equations constitute a set of differential equations which can be solved in a given region subject to imposed boundary conditions. In

many cases, the region over which a solution is sought can be divided up into several subregions, and appropriate matching of fields is made at the boundaries between these regions. Let us therefore consider the relationships between the fields adjacent to a boundary but on either side of it.

## (a)  Electric Fields

Figure 8.3-1(a) shows the boundary between two regions of different relative dielectric constants $\varepsilon_A$ and $\varepsilon_B$. A single electric field line passing through the boundary between the two regions is shown. In region $A$ the electric field has magnitude $E_A$, and in region $B$ it has magnitude $E_B$. A small rectangle is drawn about the point where the field line crosses the boundary. The rectangle is $\Delta z$ units long in the direction parallel to the boundary and $\Delta x$ units wide in the direction normal to the boundary. Half of the rectangle is in each region. We shall assume that $\mathbf{E}_A$ can be resolved into two components, one parallel to the boundary $E_{||A}$, and one normal to the boundary $E_{\perp A}$. Similarly, $\mathbf{E}_B$ can be resolved into components $E_{||B}$ and $E_{\perp B}$.
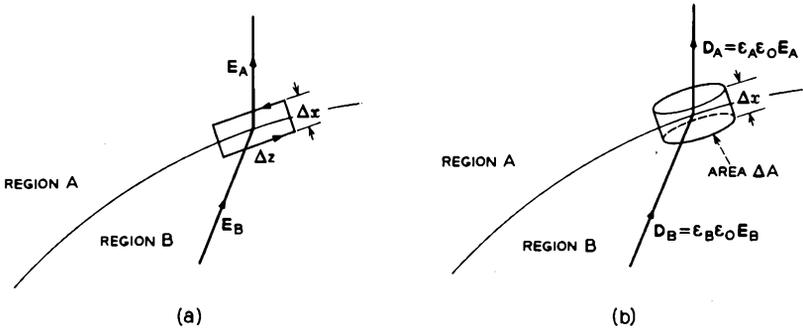


FIG. 8.3-1  Electric field vectors at a point on the boundary between two regions. Region $A$ has permittivity $\varepsilon_A \varepsilon_o$ and region $B$ has permittivity $\varepsilon_B \varepsilon_o$.

Let us evaluate Equation (8.1-5) for the region defined by the rectangle $\Delta x \Delta z$ in Figure 8.3-1(a). Substituting $\mathcal{E} = \text{Re } \mathbf{E}\, \epsilon^{j\omega t}$ and $\mathcal{B} = \text{Re } \mathbf{B}\, \epsilon^{j\omega t}$ in the equation, we obtain

$$\oint_{\text{closed loop}} \mathbf{E} \cdot d\mathbf{l} = -j\omega \int_{\text{surface}} \mathbf{B} \cdot \mathbf{n} dS \qquad (8.3\text{-}1)$$

If both $\Delta x$ and $\Delta z$ are assumed to be very small, $\mathbf{E}$ in region $A$ or region $B$ will be constant in magnitude and direction over the part of the rectangle included in the region. Let the average value of the component of $\mathbf{B}$

normal to the plane of the rectangle be $B_n$. Equation (8.3-1) can then be written

$$\oint_{\text{rectangle}} \mathbf{E} \cdot d\mathbf{l} = E_{\|B}\Delta z + E_{\perp B}\frac{\Delta x}{2} + E_{\perp A}\frac{\Delta x}{2} - E_{\|A}\Delta z - E_{\perp A}\frac{\Delta x}{2} - E_{\perp B}\frac{\Delta x}{2}$$

$$= -j\omega B_n\Delta x\Delta z \qquad (8.3\text{-}2)$$

Next let $\Delta x \to 0$ in such a way that the rectangle is still centered about the boundary. The right-hand side of Equation (8.3-2) then approaches zero, and the equation reduces to

$$E_{\|B}\Delta z - E_{\|A}\Delta z = 0 \qquad (8.3\text{-}3)$$

from which

$$E_{\|A} = E_{\|B} \qquad (8.3\text{-}4)$$

Thus the components of $\mathbf{E}$ parallel to the boundary are equal on both sides of the boundary, despite the fact that the two regions have different dielectric constants.

Next let us consider the field perpendicular to the boundary. We shall work with the $\mathbf{D}$ vector in this case and show that the normal component of $\mathbf{D}$ is continuous at the boundary. Figure 8.3-1(b) shows an electric field line which passes through the boundary between regions $A$ and $B$. In region $A$, $\mathbf{D}_A = \varepsilon_A \varepsilon_o \mathbf{E}_A$, and in region $B$, $\mathbf{D}_B = \varepsilon_B \varepsilon_o \mathbf{E}_B$. A small pillbox-shaped volume surrounds the point where the field line passes through the boundary. The pillbox has area $\Delta A$ on the faces parallel to the boundary and thickness $\Delta x$. We assume that $\mathbf{D}_A$ can be resolved into a component $D_{\perp A}$ perpendicular to the boundary and $D_{\|A}$ parallel to the boundary. Similarly, $\mathbf{D}_B$ can be resolved into $D_{\perp B}$ and $D_{\|B}$.

Equation (8.1-16) can be written in the form

$$\int_{\text{volume}} \nabla \cdot \mathbf{D}\, dv = \int_{\text{volume}} \rho\, dv \qquad (8.3\text{-}5)$$

Using Gauss's theorem (Appendix XII) this may be written as

$$\int_{\text{closed surface}} \mathbf{D} \cdot \mathbf{n}\, dS = \int_{\text{volume}} \rho\, dv \qquad (8.3\text{-}6)$$

This is the same as Equation (1.4-2). Let us now apply this equation to the pillbox-shaped volume in Figure 8.3-1(b). If we let the thickness $\Delta x$ of the box become vanishingly small, and if we assume there is no surface charge at the boundary,

$$\int_{\text{closed surface}} \mathbf{D} \cdot \mathbf{n}\, dS = D_{\perp A}\Delta A - D_{\perp B}\Delta A = \int_{\text{volume}} \rho\, dv = 0 \qquad (8.3\text{-}7)$$

Then

$$D_{\perp A} = D_{\perp B} \qquad (8.3\text{-}8)$$

Thus the perpendicular component of electric flux density is continuous at a boundary.

### (b) Magnetic Fields

Next, consider the boundary between regions of relative permeability $\mu_A$ and $\mu_B$, as shown in Figure 8.3-2(a). The figure shows a single magnetic field line which passes through the boundary. A rectangle of dimensions $\Delta x$
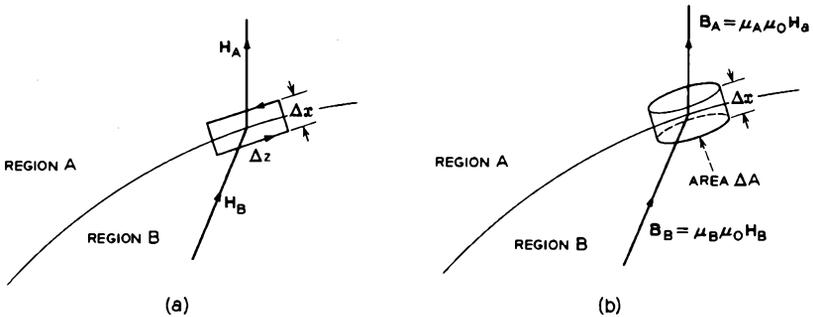


FIG. 8.3-2 Magnetic field vectors at a point on the boundary between two regions. Region $A$ has permeability $\mu_A \mu_o$ and region $B$ has permeability $\mu_B \mu_o$.

by $\Delta z$ surrounds the point where the field line crosses the boundary and is centered about the boundary so that the rectangle lies in each region. We assume that $\mathbf{H}_A$ can be resolved into components $H_{\perp A}$ and $H_{||A}$, perpendicular and parallel to the boundary. Similarly, $\mathbf{H}_B$ can be resolved into components $H_{\perp B}$ and $H_{||B}$.

Let us evaluate Equation (8.1-15) for the region defined by the rectangle $\Delta x \Delta z$ in Figure 8.3-2(a). The equation can be written in the integral form

$$\int_{\text{surface}} (\nabla \times \mathbf{H}) \cdot \mathbf{n} dS = \int_{\text{surface}} (\mathbf{J} + j\omega \mathbf{D}) \cdot \mathbf{n} dS \qquad (8.3\text{-}9)$$

Applying Stoke's theorem (see Appendix XII), we obtain

$$\oint_{\text{closed loop}} \mathbf{H} \cdot d\mathbf{l} = \int_{\text{surface}} (\mathbf{J} + j\omega \mathbf{D}) \cdot \mathbf{n} dS \qquad (8.3\text{-}10)$$

If both $\Delta x$ and $\Delta z$ are very small, $\mathbf{H}$ in region $A$ or $B$ will be constant in magnitude and direction over the part of the rectangle included in the

region, and **J** and **D** will be uniform over the area of the rectangle. Let the average values of the components of **J** and **D** normal to the plane of the rectangle be $J_n$ and $D_n$. Equation (8.3-10) can then be written as

$$\oint_{\text{closed loop}} \mathbf{H} \cdot d\mathbf{l} = H_{||B}\Delta z + H_{\perp B}\frac{\Delta x}{2} + H_{\perp A}\frac{\Delta x}{2} - H_{||A}\Delta z$$

$$-H_{\perp A}\frac{\Delta x}{2} - H_{\perp B}\frac{\Delta x}{2} = (J_n + j\omega D_n)\Delta x \Delta z \quad (8.3\text{-}11)$$

If we now let $\Delta x$ approach zero, the right-hand side of the equation approaches zero, and the equation reduces to

$$H_{||B}\Delta z - H_{||A}\Delta z = 0 \qquad (8.3\text{-}12)$$

or

$$H_{||A} = H_{||B} \qquad (8.3\text{-}13)$$

Hence the tangential component of the magnetic field intensity vector is continuous at a boundary.

Finally, let us consider the normal components of magnetic field at the boundary. We shall start with Equation (8.1-17) in the integral form:

$$\int_{\text{volume}} \nabla \cdot \mathbf{B} dv = 0 \qquad (8.3\text{-}14)$$

Using Gauss's theorem (Appendix XII) gives us

$$\int_{\text{closed surface}} \mathbf{B} \cdot \mathbf{n} dS = 0 \qquad (8.3\text{-}15)$$

If this equation is applied to the pillbox-shaped volume shown in Figure 8.3-2(b) and if we let the thickness $\Delta x$ of the volume become vanishingly small, we obtain

$$\int_{\text{closed surface}} \mathbf{B} \cdot \mathbf{n} dS = B_{\perp A}\Delta A - B_{\perp B}\Delta A = 0 \qquad (8.3\text{-}16)$$

Hence

$$B_{\perp A} = B_{\perp B} \qquad (8.3\text{-}17)$$

Thus the perpendicular component of magnetic flux density is continuous at a boundary.

In summary, at an infinitesimally thin boundary between two regions which have different permeability and permittivity, the tangential components of **E** and **H** are continuous, and the perpendicular components of **D** and **B** are continuous.

## 8.4  Ohm's Law and Skin Effect

### (a)  Ohm's Law

Ohm's Law is perhaps the first learned and most basic of the experimental laws of electricity. At low frequencies this law states that the ratio of voltage drop to current in a resistor is a constant. At microwave frequencies, the current density throughout a resistor or a conductor is usually not constant, and Ohm's Law is best stated in the form

$$\mathbf{J} = \sigma\mathbf{E} \tag{8.4-1}$$

where $\mathbf{J}$ is the current density, $\mathbf{E}$ is the electric field intensity, and $\sigma$ is the conductivity of the medium.

This equation can be related to the more familiar form of Ohm's Law in the following way. Consider a conductor of length $l$ and cross-sectional area $A$, as shown in Figure 8.4-1. A voltage $V$ is applied over the length $l$, and a current density $J$ flows parallel to the length $l$. We assume the current density is uniform over the cross section of the conductor. The electric field intensity within the conductor is of magnitude $E$. Then $J = \sigma E$, and the total current flowing in the conductor is given by
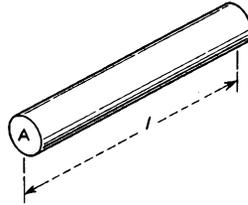


Fig. 8.4-1 A uniform cylindrical conductor of conductivity $\sigma$.

$$I = JA = \sigma EA = \sigma\frac{V}{l}A = \frac{V}{R} \tag{8.4-2}$$

or

$$V = IR \tag{8.4-3}$$

where $R = l/\sigma A$ is the resistance of the conductor over the length $l$. Equation (8.4-3) expresses the more familiar form of Ohm's Law.

### (b)  Skin Effect

Here we shall derive the distribution of current density in a semi-infinite conductor when an rf electric field is applied parallel to the surface of the conductor.[4] Figure 8.4-2 shows a portion of the conductor. We shall assume that the electric field is applied in the $z$ direction only and that it does not vary in magnitude in the $y$ and $z$ directions. Let the electric field

---

[4]Reference 8a.

just outside the conductor be $E_{zo}$.  From the discussion in Section 8.3, we
know that this field will be continuous across the boundary between the
conductor and free space, and hence that $E_{zo}$ also will be the electric field
intensity just inside the surface of the conductor.  We shall first determine
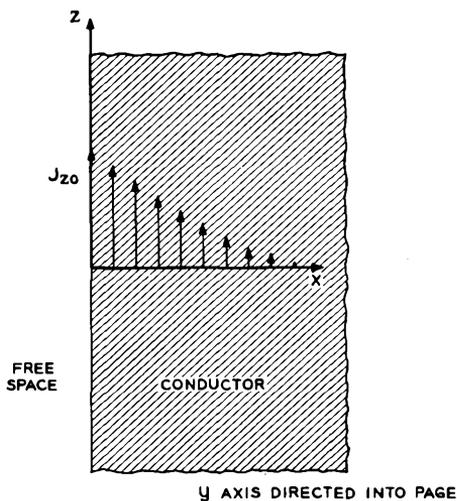


FIG. 8.4-2 Current flow near the surface of a conductor at microwave frequencies.

the variation of $E_z$ with distance $x$ into the conductor, and since $J_z = \sigma E_z$,
we shall note that $J_z$ varies in a similar manner with distance into the
conductor.

Let us use Equations (8.1-18) and (8.1-19) to derive an equation for $E_z$
within the conductor.  Since $E_x = E_y = \partial E_z/\partial y = 0$, Equations (8.1-18)
reduce to

$$H_x = H_z = 0 \qquad (8.4\text{-}4)$$

and

$$j\omega\mu\mu_o H_y = \frac{\partial E_z}{\partial x} \qquad (8.4\text{-}5)$$

We shall set $J_z = \sigma E_z$ and $J_x = J_y = 0$ in Equations (8.1-19).  Then

$$\frac{\partial H_y}{\partial z} = 0 \qquad (8.4\text{-}6)$$

and

$$\frac{\partial H_y}{\partial x} = (\sigma + j\omega\varepsilon\varepsilon_o)E_z \qquad (8.4\text{-}7)$$

Substituting for $H_y$ in Equation (8.4-7) from Equation (8.4-5), we obtain

$$\frac{\partial^2 E_z}{\partial x^2} = j\omega\mu\mu_o(\sigma + j\omega\varepsilon\varepsilon_o)E_z \qquad (8.4\text{-}8)$$

Consideration of the actual values of $\sigma$, $\omega$, and $\varepsilon\varepsilon_o$ for conductors at microwave frequencies shows that

$$\sigma \gg \omega\varepsilon\varepsilon_o \qquad (8.4\text{-}9)$$

Hence to a good approximation

$$\frac{\partial^2 E_z}{\partial x^2} = j\omega\mu\mu_o\sigma E_z \qquad (8.4\text{-}10)$$

This is the wave equation for an electric field in a conducting medium. The equation is analogous to Equation (8.1-26), which applies to electromagnetic waves in free space or in a dielectric medium. Equation (8.4-10) has the solution

$$E_z = E_{zo}\epsilon^{-(1+j)\sqrt{\omega\mu\mu_o\sigma/2}\,x} \qquad (8.4\text{-}11)$$

Finally, substituting $J_z = \sigma E_z$ and $J_{zo} = \sigma E_{zo}$, we obtain

$$J_z = J_{zo}\epsilon^{-(1+j)\sqrt{\omega\mu\mu_o\sigma/2}\,x} \qquad (8.4\text{-}12)$$

This equation shows that not only does the current density decay in magnitude away from the surface, but it also experiences a progressive phase shift. Although this relationship has been derived for a plane surface of infinite depth, it may be applied to curved surfaces of finite depth as long as the current decays in a distance small compared with the thickness and radius of curvature of the conductor.

It is convenient to write Equation (8.4-12) in the form

$$J_z = J_{zo}\epsilon^{-(1+j)x/\delta} \qquad (8.4\text{-}13)$$

where

$$\delta = \frac{1}{\sqrt{\pi f \mu\mu_o\sigma}} \qquad (8.4\text{-}14)$$

The length $\delta$ is known as the skin depth. The skin depth $\delta$ is a measure of the rate at which the current density decays into the metal. In a distance $\delta$ from the surface, the current density has dropped to $1/\epsilon$ of its value at the surface. This is a very rapid decay at microwave frequencies for most metals. Table 8.4-1 gives the number of skin depths in a 1.59 mm (1/16 inch) thick wall of several metals commonly used in microwave transmission

TABLE 8.4-1. NUMBER OF SKIN DEPTHS IN 1.59 MM THICKNESS
WALLS OF VARIOUS MATERIALS

| Metal | Number at 3 Gc | Number at 9 Gc |
|---|---|---|
| Silver............................ | 1352 | 2340 |
| Copper........................... | 1318 | 2280 |
| Gold............................. | 1106 | 1913 |
| Molybdenum..................... | 726 | 1258 |
| Nickel........................... | 583 | 1010 |
| Stainless steel (nonmagnetic)........ | 181 | 314 |

lines and electron tubes. This thickness is used for the wall of several stand-
ard size waveguides at microwave frequencies. Because the skin depth is so
small at microwave frequencies, one normally assumes negligible currents
exist on the outer surface of a waveguide or cavity. For example, at 3 Gc
a copper waveguide of 1.59-mm wall thickness has current densities on the
outer surface which are only $10^{-572}$ times the current density on the inner
surface. Thus, in effect, perfect shielding is accomplished. Since the
current density decays so rapidly with distance, the bulk of the metal in
microwave conducting structures is not used to provide a path for current
flow but rather is used for structural rigidity. It is an excellent approxima-
tion to visualize the wall currents in microwave structures as consisting
solely of surface currents.

The imaginary part of the exponent in the right-hand side of Equation
(8.4-11) gives the phase change of the electric field intensity as it propagates
into the conductor. We see that the skin depth $\delta$ corresponds to $1/2\pi$
wavelengths of the type of wave propagation that takes place in the
conductor.

One can use Equation (8.4-13) to determine the total ohmic power loss
per unit surface area of the conductor for a given tangential component of
magnetic field in free space just outside the conductor. The ohmic power
loss in an element of volume having unit length parallel to the surface, unit
width parallel to the surface, and thickness $dx$ in the direction normal to the
surface is $(1/2\sigma) |J_z|^2 dx$. The total power loss per unit area of the surface
is then

$$P_\square = \frac{1}{2\sigma} \int_0^\infty |J_z|^2 dx = \frac{\delta}{4\sigma} |J_{zo}|^2 \qquad (8.4\text{-}15)$$

where we have substituted for $J_z$ from Equation (8.4-13).

Often it is more convenient in using this equation to express $J_{zo}$ in terms
of the magnetic field in free space just outside the conductor. Within the

conductor we can rewrite Equation (8.3-10) in the form

$$\oint_{\text{closed loop}} \mathbf{H} \cdot d\mathbf{l} = \int_{\text{surface}} (\sigma + j\omega\varepsilon\varepsilon_o)\mathbf{E} \cdot \mathbf{n}dS \approx \int_{\text{surface}} \sigma\mathbf{E} \cdot \mathbf{n}dS$$

$$= \int_{\text{surface}} \mathbf{J} \cdot \mathbf{n}dS \qquad (8.4\text{-}16)$$

where we have again made the approximation that $\sigma \gg \omega\varepsilon\varepsilon_o$ within the conductor. Figure 8.4-3 shows a cross-sectional view of the conductor near the
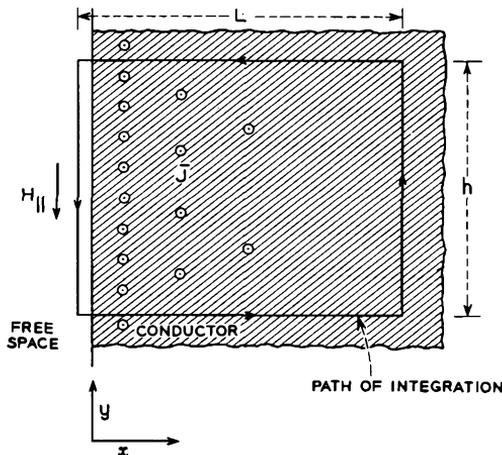


FIG. 8.4-3 Current flow in a conductor at microwave frequencies. The length $L$ is much greater than $\delta$. The component of magnetic field parallel to the path of integration is therefore zero along the right-hand side of the path of integration.

surface. The current flow $J_z$ is assumed to be normal to the page and directed out of the page. Let us evaluate the left-hand side of Equation (8.4-16) for the path of integration shown in the figure. We assume that the length $L$ is very large so that $J_z$ is essentially zero at the right-hand side of the path. For symmetry reasons there is no net contribution to the line integral for a component of $\mathbf{H}$ parallel to the top and bottom sides of the path of integration (assuming $h$ is infinitesimally short). Then

$$\oint_{\text{closed loop}} \mathbf{H} \cdot d\mathbf{l} = H_{||}h = \int_{\text{surface}} \mathbf{J} \cdot \mathbf{n}dS = J_{zo}h \int_o^\infty \epsilon^{-(1+j)x/\delta}dx$$

$$= \frac{J_{zo}h\delta}{1+j} \qquad (8.4\text{-}17)$$

where $H_{||}$ is the magnetic field parallel to the surface just outside the conductor. Then

$$H_{||} = \frac{J_{so}\delta}{1 + j} \qquad (8.4\text{-}18)$$

Substituting into Equation (8.4-15), we obtain

$$P_\square = \frac{|H_{||}|^2}{2\sigma\delta} \qquad (8.4\text{-}19)$$

This equation is most useful in the sense that ohmic losses in cavity or waveguide walls can be computed directly from the magnetic fields in the free space adjacent to the metal without resorting to calculations of the currents within the metal conductors themselves.

Equation (8.4-14) shows that $\delta$ is inversely proportional to the square root of the frequency. Hence, by Equation (8.4-19), the ohmic loss is proportional to the square root of the frequency, for a constant conductivity and for a given magnetic field at the surface.

## (c)   The Perfect Conductor

The concept of a perfect conductor is often used in the study of microwave components. In essence, this concept assumes, for purposes of solving for the fields in regions not containing metal, that the conductivity of the metal is infinite. Now, infinite conductivity implies that charges could travel instantaneously to neutralize any electric field which would tend to be set up within a conductor; thus the electric field within a perfect conductor is zero. Since the tangential component of electric field is continuous at the surface of the conductor, the tangential component of electric field outside the conductor must be zero adjacent to the surface. However, electric field lines can terminate on surface charges on the conductor, the field lines being perpendicular to the surface at the point of intersection. On the other hand, magnetic field lines cannot pass through a perfect conductor or terminate on it. Thus, there can only be a tangential component of magnetic field just outside a perfect conductor.

The errors involved in using the concept of a perfect conductor to find the external field distribution are of the same order of magnitude as the ratio of the skin depth to the other cavity or waveguide dimensions, and generally they may be considered to be negligible. Thus, for most purposes, the electromagnetic fields within a cavity or waveguide can be found under the assumption that the metal walls are perfect conductors. The fact that the conductor is imperfect affects only the power loss or attenuation, and this may be accounted for by using the concept of skin depth together with Equation (8.4-19).

## 8.5 Waveguides

In this section we shall discuss solutions of the wave equation for the case of electromagnetic wave propagation in a waveguide. Figure 8.5-1 shows an end view of a waveguide. We shall assume that the waveguide is of infinite length in the $z$ direction and its cross-sectional dimensions remain constant with $z$. The walls of the waveguide are perfect conductors, so that the solutions we obtain must satisfy the boundary conditions that the tangential component of electric field and the normal component of magnetic field be zero at the conducting surfaces. It turns out that there are an infinite number of solutions to the wave equation which satisfy these boundary conditions. These solutions are known as *modes of propagation*. This is analogous to the infinite number of possible modes of vibration for a vibrating string. Which modes are vibrated depend on how the string is plucked. In the waveguide, the manner and frequency of excitation at the input determine which modes are excited.
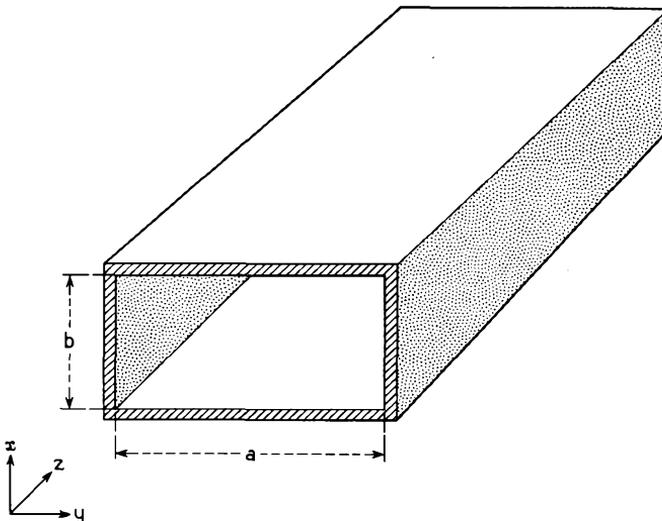


Fɪɢ. 8.5-1 Rectangular waveguide.

The infinite number of waveguide modes of propagation can be divided into two classes. Modes with $E_z$ equal to zero are known as transverse electric modes or TE modes. Modes with $H_z$ equal to zero are known as transverse magnetic modes or TM modes. As long as the waveguide is

uniformly filled with dielectric (including air or vacuum) the solutions fall into one or the other of these two classes.[5]

(a)  TE *Modes*

The solution for the TE modes can be obtained as follows. The wave equations, Equations (8.1-26) and (8.1-27), are vector equations and hence are satisfied by each component of the electric and magnetic fields. Thus the z component of magnetic field satisfies the equation

$$\nabla^2 H_z + k^2 H_z = 0 \tag{8.5-1}$$

where $k^2 = \omega^2 \mu \mu_o \varepsilon \varepsilon_o$. Expanding the Laplacian, we obtain

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + \frac{\partial^2 H_z}{\partial z^2} + k^2 H_z = 0 \tag{8.5-2}$$

A particular solution to this equation may be obtained by the method of separation of variables as

$$H_z = A \cos \frac{m\pi y}{a} \cos \frac{n\pi x}{b} \epsilon^{-j\beta z} \tag{8.5-3}$$

where $m$ and $n$ are arbitrary integers. We shall see later that this solution meets the boundary conditions imposed by the waveguide walls. Substituting Equation (8.5-3) into Equation (8.5-2), we find that the following relationship must be satisfied for Equation (8.5-3) to be a solution:

$$\beta^2 + \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 = k^2 = \omega^2 \mu \mu_o \varepsilon \varepsilon_o = \frac{\omega^2}{c^2} \tag{8.5-4}$$

The presence of the integers $m$ and $n$ in this equation indicates that there are an infinite number of solutions corresponding to an infinite number of modes of propagation.

If the variation of the magnetic field with time is included in Equation (8.5-3), the equation becomes

$$\mathcal{H}_z = \text{Re } A \cos \frac{m\pi y}{a} \cos \frac{n\pi x}{b} \epsilon^{j(\omega t - \beta z)}$$

$$= A \cos \frac{m\pi y}{a} \cos \frac{n\pi x}{b} \cos (\omega t - \beta z) \tag{8.5-5}$$

We see that the wave travels in the z direction with a phase velocity given by

$$v_p = \frac{\omega}{\beta} \tag{8.5-6}$$

[5]Reference 8.3.

where $\beta$ is determined by Equation (8.5-4) and is a characteristic of the particular mode of propagation.

The other components of the fields in the waveguide can be obtained from Equation (8.5-3) by means of Equations (8.1-18) and (8.1-19). Let us first note that (1) $E_z$ is zero, and (2) all field quantities will vary with $z$ as $\epsilon^{-j\beta z}$, so that differentiating with respect to $z$ is equivalent to multiplying by $-j\beta$. We shall assume that the waveguide is filled with air, so that to a good approximation $\mu = \epsilon = 1$. Equations (8.1-18) then give

$$j\beta E_y = -j\omega\mu_o H_x \tag{8.5-7}$$

and

$$j\beta E_x = j\omega\mu_o H_y \tag{8.5-8}$$

From Equations (8.1-19) we obtain

$$-j\beta H_x - \frac{\partial H_z}{\partial x} = j\omega\epsilon_o E_y \tag{8.5-9}$$

and

$$\frac{\partial H_z}{\partial y} + j\beta H_y = j\omega\epsilon_o E_x \tag{8.5-10}$$

Finally, combining Equations (8.5-7) and (8.5-9) as well as Equation (8.5-3) gives

$$E_y = \frac{-j\omega\mu_o}{k^2 - \beta^2}\frac{n\pi}{b}A\,\cos\frac{m\pi y}{a}\,\sin\frac{n\pi x}{b}\epsilon^{-j\beta z} \tag{8.5-11}$$

and

$$H_x = \frac{j\beta}{k^2 - \beta^2}\frac{n\pi}{b}A\,\cos\frac{m\pi y}{a}\,\sin\frac{n\pi x}{b}\epsilon^{-j\beta z} \tag{8.5-12}$$

Combining Equations (8.5-8), (8.5-10), and (8.5-3) gives

$$E_x = \frac{j\omega\mu_o}{k^2 - \beta^2}\frac{m\pi}{a}A\,\sin\frac{m\pi y}{a}\,\cos\frac{n\pi x}{b}\epsilon^{-j\beta z} \tag{8.5-13}$$

and

$$H_y = \frac{j\beta}{k^2 - \beta^2}\frac{m\pi}{a}A\,\sin\frac{m\pi y}{a}\,\cos\frac{n\pi x}{b}\epsilon^{-j\beta z} \tag{8.5-14}$$

Examination of Equations (8.5-11) through (8.5-14) shows that $H_x$ is zero when $x = 0$ and when $x = b$, and $H_y = 0$ when $y = 0$ and when $y = a$. Similarly, $E_x = 0$ when $y = 0$ and when $y = a$, and $E_y = 0$ when $x = 0$ and when $x = b$. Thus the normal component of $\mathbf{H}$ and the tangential component of $\mathbf{E}$ are zero at the inside walls of the waveguide. The field

solutions given by Equations (8.5-3) and (8.5-11) through (8.5-14) therefore satisfy the wave equation and the boundary conditions imposed by the waveguide.

From Equations (8.5-7) and (8.5-8) we note that

$$\frac{E_x}{H_y} = -\frac{E_y}{H_x} = \frac{\omega\mu_o}{\beta} \qquad (8.5\text{-}15)$$

showing that the perpendicular components of **E** and **H** are constant through the cross section of the waveguide.

In the introductory part of this chapter, we described the dominant TE mode, that is the TE mode with the lowest cutoff frequency. This is the $TE_{10}$ mode ($m = 1$, $n = 0$). The field components for this mode are

$$E_x = \frac{j\omega\mu_o aA}{\pi} \sin\frac{\pi y}{a}\epsilon^{-j\beta z} \qquad (8.5\text{-}16)$$

$$H_y = \frac{j\beta aA}{\pi} \sin\frac{\pi y}{a}\epsilon^{-j\beta z} \qquad (8.5\text{-}17)$$

$$H_z = A \cos\frac{\pi y}{a}\epsilon^{-j\beta z} \qquad (8.5\text{-}18)$$

and

$$E_y = E_z = H_x = 0 \qquad (8.5\text{-}19)$$

The field configuration for this mode is shown in Figure 8-5.

Setting $m = 1$, $n = 0$ in Equation (8.5-4) gives

$$\beta^2 + \left(\frac{\pi}{a}\right)^2 = k^2 = \frac{\omega^2}{c^2} = \left(\frac{2\pi}{\lambda_o}\right)^2 \qquad (8.5\text{-}20)$$

or

$$\beta^2 = \left(\frac{2\pi}{\lambda_o}\right)^2\left(1 - \frac{\lambda_o^2}{\lambda_c^2}\right) \qquad (8.5\text{-}21)$$

where we have set $\lambda_o = 2\pi c/\omega$, the free-space wavelength of a wave of radian frequency $\omega$, and $\lambda_c = 2a$, as in the introductory part of this chapter. The phase velocity for the $TE_{10}$ mode then becomes

$$v_p = \frac{\omega}{\beta} = \frac{\omega\lambda_o}{2\pi\sqrt{1 - \lambda_o^2/\lambda_c^2}} = \frac{c}{\sqrt{1 - \lambda_o^2/\lambda_c^2}} \qquad (8.5\text{-}22)$$

as in Equation (8-5). Similarly

$$\lambda_z = \frac{2\pi}{\beta} = \frac{\lambda_o}{\sqrt{1 - \lambda_o^2/\lambda_c^2}} \qquad (8.5\text{-}23)$$

Next let us consider the group velocity $v_g$. This is equal to the time average power flow in the waveguide divided by the energy stored per unit

length in the $z$ direction. From Equation (8.2-11), the time average power flow for the $TE_{10}$ mode is given by

$$\text{time average power flow} = \int\limits_{\substack{\text{area of} \\ \text{waveguide}}} (\text{time average of } |\mathbf{S}|)dxdy$$

$$= \frac{\beta\omega\mu_o a^3 b A^2}{4\pi^2} \qquad (8.5\text{-}24)$$

where (time average of $|\mathbf{S}|$) = (time average of $|\mathbf{\varepsilon} \times \mathfrak{K}|$) = $\frac{1}{2}|E_x H_y|$, and we have substituted for $E_x$ and $H_y$ from Equations (8.5-16) and (8.5-17) From Equations (8.2-2), (8.2-7), and Equation (6) of Appendix XIV, the average energy stored per unit length in the $z$ direction is given by

energy stored per unit length

$$= \frac{1}{\lambda_z}\int_{z=0}^{b}\int_{y=0}^{a}\int_{z=0}^{\lambda_z}\frac{1}{4}(\varepsilon_o|E_x|^2 + \mu_o|H_y|^2 + \mu_o|H_z|^2)dxdydz = \frac{\omega^2\mu_o a^3 b A^2}{4\pi^2 c^2}$$

$$(8.5\text{-}25)$$

Then

$$v_g = \frac{\text{time average power flow}}{\text{energy stored per unit length}} = \frac{\beta c^2}{\omega} = c\sqrt{1 - \lambda_o^2/\lambda_c^2} \quad (8.5\text{-}26)$$

as in Equation 8-6. Some further discussion of the group velocity is given in Appendix XIII where it is shown that

$$v_g = \frac{\partial\omega}{\partial\beta} \qquad (8.5\text{-}27)$$

From Equation (8.5-4) it is evident that $\beta\partial\beta = \omega\partial\omega/c^2$, and hence

$$v_p v_g = c^2 \qquad (8.5\text{-}28)$$

Substituting for $v_p$ from Equation (8.5-22) in this expression, we obtain $v_g = c\sqrt{1 - \lambda_o^2/\lambda_c^2}$, as in Equation (8.5-26).

The *characteristic impedance*[6] of the waveguide is defined in terms of the time average power flow and a "voltage" at the center of the waveguide given by the integral $\int_0^b E_x dx$, where $E_x$ is evaluated at $y = a/2$. From Equation (8.5-16), the mean-square value of this voltage is

$$\begin{array}{c}\text{mean-square "voltage"} \\ \text{at center of waveguide}\end{array} = \frac{\omega^2\mu_o^2 a^2 A^2 b^2}{2\pi^2} \qquad (8.5\text{-}29)$$

---

[6]This definition is not unique. Two other definitions for waveguide impedance are also used. See Reference 8b, pp. 36, 37.

The characteristic impedance for the $TE_{10}$ mode is then given by

$$Z_0 = \frac{\text{mean-square voltage}}{\text{time average power flow}} = \frac{2b}{a}\frac{\sqrt{\mu_0/\varepsilon_0}}{\sqrt{1 - \lambda_0^2/\lambda_c^2}} = \frac{754b}{a\sqrt{1 - \lambda_0^2/\lambda_c^2}}\text{ohms}$$

$$(8.5\text{-}30)$$

where we have substituted from Equation (8.5-24) for the time average power flow.

Let us plot Equation (8.5-4) as $\omega$ vs. $\beta$. We obtain the family of hyperbolas shown in Figure 8.5-2. Note that each mode has a cutoff frequency given by $\beta = 0$ in Equation (8.5-4):

$$\omega_{\text{cutoff}} = c\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

$$(8.5\text{-}31)$$

Furthermore, each curve is asymptotic to the straight line

$$\omega = \beta c$$

$$(8.5\text{-}32)$$

This straight line has a slope equal to $c$, the velocity of light. A simple geometric construction enables us to obtain the phase velocity corresponding to any frequency for any mode. Suppose we want to know the phase velocity corresponding to propagation in the $TE_{01}$ mode at a radian frequency $\omega_1$. The slope of a line drawn from the origin to a point on the $\omega$-$\beta$
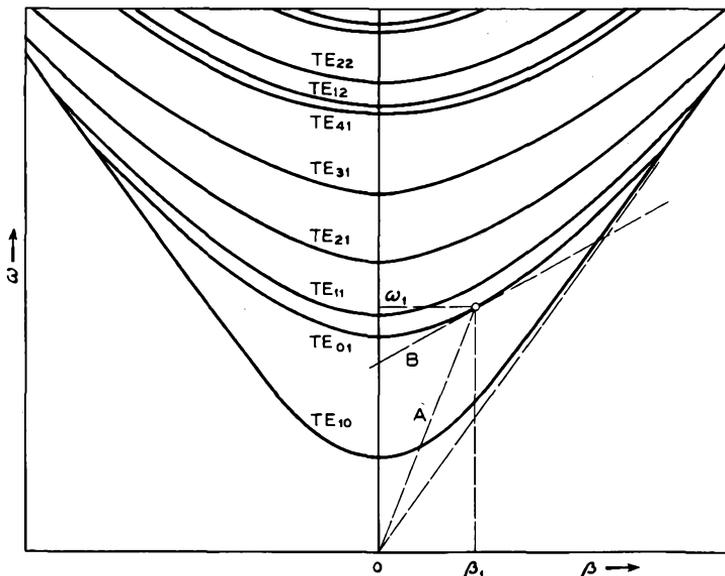


FIG. 8.5-2 $\omega$-$\beta$ diagram for the TE modes in a rectangular waveguide. All the curves are asymptotic to lines through the origin with slopes equal in magnitude to the velocity of light. $a/b = 2.3$.

curve for the correct mode and frequency (line $A$ in Figure 8.5-2) gives the phase velocity, according to Equation (8.5-22). Note that the phase velocity for all propagating frequencies and all modes is greater than the velocity of light. If the waveguide were filled with a dielectric of relative dielectric constant $\varepsilon$, the asymptote would correspond to $c = 1/\sqrt{\varepsilon\varepsilon_o\mu_o}$. The fact that the phase velocity is greater than the velocity of light is a universal property of waveguides of this type, having transverse dimensions invariant with axial position.

The group velocity $v_g = \partial\omega/\partial\beta$ is equal to the slope of a line tangent to the $\omega$-$\beta$ curve at the operating frequency (line $B$ in Figure 8.5-2). It is evident that the group velocity is always less than the velocity of light.

### (b)  TM *Modes*

So far we have considered only the transverse electric or TE modes. We shall now consider the equivalent relationships for the transverse magnetic or TM modes.

We may begin consideration of the TM modes by considering the $z$ component of Equation (8.1-26).

$$\nabla^2 E_z + k^2 E_z = 0 \qquad (8.5\text{-}33)$$

A particular solution to this equation is given by

$$E_z = A \sin \frac{m\pi y}{a} \sin \frac{n\pi x}{b}\epsilon^{-j\beta z} \qquad (8.5\text{-}34)$$

where $m$ and $n$ are integers. When this solution is substituted back into Equation (8.5-23), we find that

$$\beta^2 + \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 = k^2 = \omega^2\mu\mu_o\varepsilon\varepsilon_o \qquad (8.5\text{-}35)$$

as in the case of TE modes (Equation (8.5-4)).

The other field components may be obtained from Equation (8.5-34) by application of Equations (8.1-18) and (8.1-19), in which case we set $H_z = 0$. Thus we obtain

$$H_x = \frac{j\omega\varepsilon_o}{k^2 - \beta^2}\frac{m\pi}{a}A \cos \frac{m\pi y}{a} \sin \frac{n\pi x}{b}\epsilon^{-j\beta z} \qquad (8.5\text{-}36)$$

$$H_y = -\frac{j\omega\varepsilon_o}{k^2 - \beta^2}\frac{n\pi}{b}A \sin \frac{m\pi y}{a} \cos \frac{n\pi x}{b}\epsilon^{-j\beta z} \qquad (8.5\text{-}37)$$

$$E_x = -\frac{j\omega\beta}{k^2 - \beta^2}\frac{n\pi}{b}A \sin \frac{m\pi y}{a} \cos \frac{n\pi x}{b}\epsilon^{-j\beta z} \qquad (8.5\text{-}38)$$

$$E_y = -\frac{j\beta}{k^2 - \beta^2}\frac{m\pi}{a}A \cos \frac{m\pi y}{a} \sin \frac{n\pi x}{b}\epsilon^{-j\beta z} \qquad (8.5\text{-}39)$$

It is easily shown that these field components satisfy the boundary condi-

tions imposed by the waveguide, and hence these are the field components associated with transverse magnetic waves. Figure 8.5-3 shows the field pattern for the $TM_{11}$ mode ($m = n = 1$).



SECTION A–A

E LINES ⟶
H LINES ---⟶
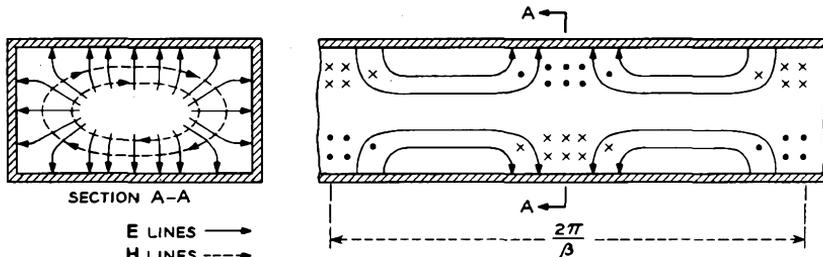
FIG. 8.5-3 $TM_{11}$ mode in a rectangular waveguide.

A complete $\omega$-$\beta$ diagram for both the TE and TM modes is shown in Figure 8.5-4. All of the TM modes are degenerate; that is, a TE mode has the same $\omega$-$\beta$ curve. The two lowest modes are TE modes and have no TM counterpart. Rectangular waveguide is normally operated in the mode of lowest cutoff frequency, the $TE_{10}$ mode.
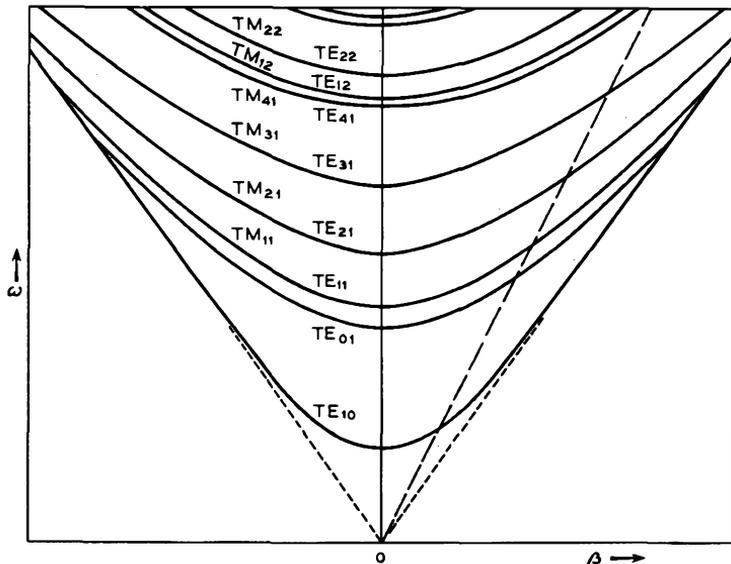


FIG. 8.5-4 Complete $\omega$-$\beta$ diagram for a rectangular waveguide. All the curves are asymptotic to the velocity of light lines. $a/b = 2.3$.

The relationships of power flow, group velocity, and phase velocity are the same for TM as for TE modes. If, for instance, we draw a straight line from the origin, as shown in Figure 8.5-4, then at the frequencies of intersection on each mode branch the phase velocities are all the same; correspondingly, by Equation (8.5-28), the group velocities are also identical at these points.

We have not considered any losses in the above discussion. Losses may be taken into account by allowing the propagation constant $\beta$ to be complex, so that the wave is attenuated in the $z$ direction. In this manner, we can allow both for resistive losses in the walls and also dielectric losses if the guide contains dielectric.

Transmission line theory may be applied directly to waveguides. For instance, a quarter wavelength away from a short circuit one sees an open circuit. Of course, in the case of a waveguide, a wavelength is no longer equal to a free space wavelength $c/f$. Rather it is a guide wavelength, given by $\lambda_z = 2\pi/\beta$, and $\beta$ is obtained from the $\omega$-$\beta$ curve; $\lambda_z$ is thus a function of frequency and of the mode of propagation.

No mention has yet been made of coupling energy in or out of a waveguide. One common method of coupling between a coaxial line and a waveguide is shown in Figure 8.5-5. The center conductor of a coaxial line is brought down through the broad wall of the waveguide. The center conductor acts like an antenna to radiate energy into the waveguide. A short is placed in the waveguide a quarter wavelength to the left of the probe causing the region to the left of the probe to look



FIG. 8.5-5 Coaxial line to waveguide transition.

like an open circuit at the probe. Hence, resultant power flow is to the right. At the receiving end of the waveguide, a similar transition may be used to couple energy back into a coaxial line.

## 8.6 Cavity Resonators

In the introductory section of this chapter, we looked at cavity resonators from the point of view of an evolution from a simple $L$-$C$ circuit. With the

discussion of waveguides behind us, we can now look at cavity resonators from a different point of view.

Let us consider the electric field solutions for the TE$_{10}$ mode in a rectangular waveguide. From Equations (8.5-16) and (8.5-19),

$$E_x = B \sin \frac{\pi y}{a} \epsilon^{-j\beta z} \qquad (8.6\text{-}1)$$

$$E_y = E_z = 0 \qquad (8.6\text{-}2)$$

Now this solution corresponds to a wave traveling in the positive $z$ direction. There is an equally valid solution corresponding to propagation in the negative $z$ direction:

$$E_x = C \sin \frac{\pi y}{a} \epsilon^{+j\beta z} \qquad (8.6\text{-}3)$$

$$E_y = E_z = 0 \qquad (8.6\text{-}4)$$

where $\beta$ is taken to be positive in both Equations (8.6-1) and (8.6-3). Physically, the wave traveling in the negative $z$ direction could be set up by an obstacle in a waveguide which reflects part of the outgoing energy back toward the source. The general solution is thus given by the superposition of the above two waves, resulting in

$$E_x = (B\epsilon^{-j\beta z} + C\epsilon^{+j\beta z}) \sin \frac{\pi y}{a} \qquad (8.6\text{-}5)$$

$$E_y = E_z = 0 \qquad (8.6\text{-}6)$$

Now, we can make a rectangular cavity resonator out of a rectangular waveguide simply by placing walls perpendicular to the $z$ axis at $z = 0$ and $z = L$. Equation (8.6-5) must then satisfy the additional boundary condition of being zero at the added walls. Setting $E_x$ to zero at $z = 0$ gives us

$$0 = B + C \qquad (8.6\text{-}7)$$

so that Equation (8.6-5) may be written

$$E_x = 2jC \sin \beta z \sin \frac{\pi y}{a} \qquad (8.6\text{-}8)$$

The additional boundary condition at $z = L$ is satisfied for

$$\beta = \frac{p\pi}{L} \qquad (8.6\text{-}9)$$

where $p$ is an integer. Since $\lambda_z = 2\pi/\beta$, this states that the cavity must be an integral number of half guide wavelengths long.

Using Equation (8.4-30) and setting $m = 1$ and $n = 0$ for the case of the

$TE_{10}$ mode, we obtain an expression for the resonant frequencies of the cavity:

$$\omega_p = c\sqrt{\left(\frac{p\pi}{L}\right)^2 + \left(\frac{\pi}{a}\right)^2}, \qquad p = 1,2,3 \ldots \qquad (8.6\text{-}10)$$

A sketch of the lowest frequency, or $p = 1$, mode of oscillation is shown in Figure 8.6-1 for the case in which $L = a$.  In a cavity at resonance, the



E LINES ———→
H LINES ---→

SECTION A–A
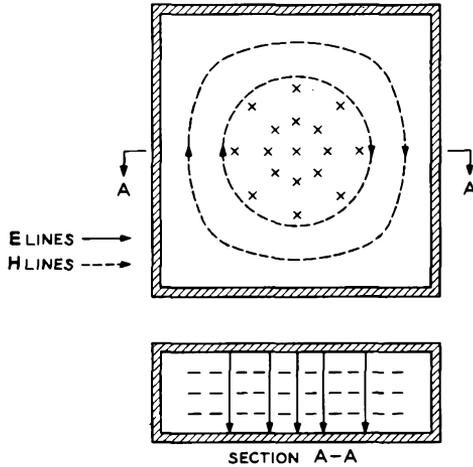
FIG. 8.6-1 Field patterns for the dominant mode of a rectangular cavity resonator.

electric and magnetic lines are 90 degrees out of time phase.  The stored energy oscillates back and forth between the two kinds of fields.  Unlike the waveguide fields, the resonator fields remain fixed in space, varying sinusoidally with time uniformly throughout the cavity.

This resonator and its field patterns may be compared with the re-entrant cavity of Figure 8-1.  One might have anticipated that the patterns of Figure 8.6-1 would occur when the heights of the posts in Figure 8-1(d) are reduced to zero.

A field analysis such as we have just carried out also enables one to obtain the resonant frequencies of all the higher-order modes.  These higher-order modes are usually of interest not because of their utility but rather because of the trouble they can cause.  For instance, in a magnetron, higher-order modes may give rise to undesirable output signals.

Resonant cavities of the type considered here are useful as microwave circuit elements. In essence, they are low-loss resonant circuits, and they may be coupled together in various ways to achieve filter-type characteristics.

As an example of a rectangular cavity resonator, let us consider a resonator of the general shape shown in Figure 8.6-1. If we assume the base to be square, Equation (8.6-10) indicates the cavity will resonate in the $TE_{101}$ mode ($m = 1$, $n = 0$, $p = 1$) at a frequency of 3000 Mc for $a = L = 7.07$ cm. The largest dimension of the re-entrant cavity of Figure 8-2 resonant at the same frequency was only 1.85 cm. Thus the effect of re-entrancy in a cavity is seen to be a decrease in overall size for the same resonant frequency. Further analysis reveals that this decrease in size is obtained at the expense of increased losses for the same stored energy in the two types of cavity.

## 8.7  Slow-Wave Structures

We have seen in Section 8.5 that wave propagation in ordinary wave-guides is characterized by a phase velocity which is greater than the velocity of light. The phase velocity is the velocity with which an observer would have to move so as to remain always in the same phase of the wave.

In the operation of traveling-wave and magnetron type devices, the electron beam must keep in step (or nearly in step) with a propagating wave Since electrons can be accelerated only to velocities which are less than the velocity of light, we must look for microwave circuits or structures capable of propagating waves with phase velocities less than the velocity of light.
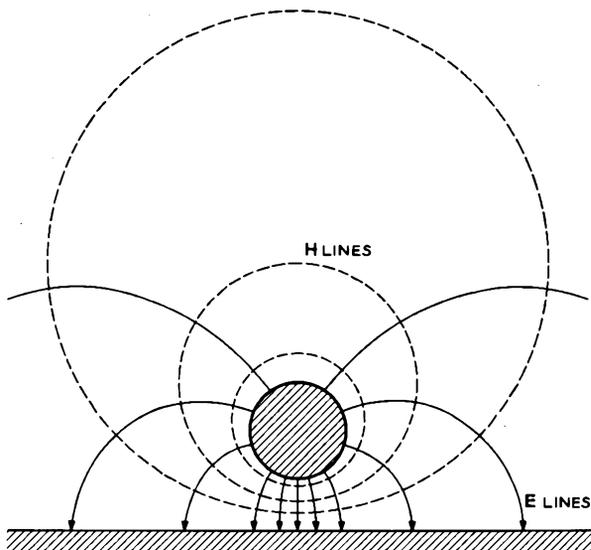


FIG. 8.7-1 Transmission line composed of a single wire above a ground plane.
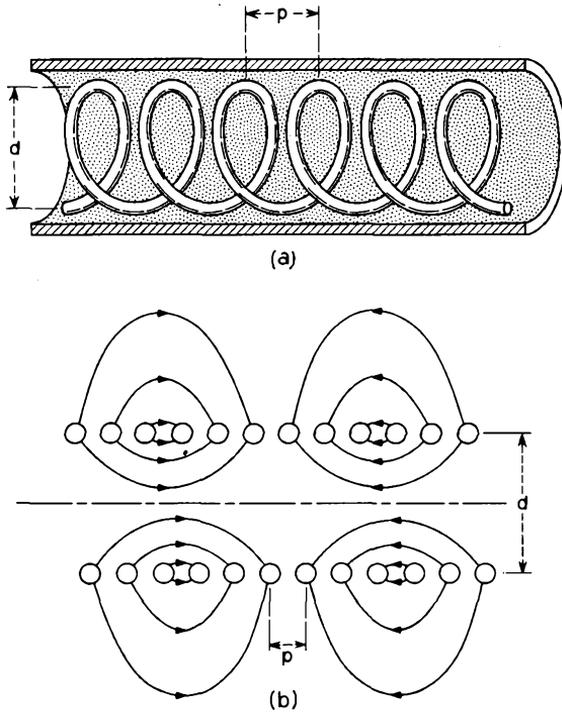
FIG. 8.7-2 The helix slow-wave structure. (a) Helical coil within a concentric conducting cylinder. This slow-wave circuit is obtained by wrapping the single-wire-above-ground into a helix, with the ground plane becoming the surrounding cylinder. (b) Electric field lines for a helix in free space.

For reasons that will become clearer in later chapters, ordinary waveguides partially or completely filled with dielectric are not satisfactory solutions to this problem. Instead, the solution will be found in a whole new class of structures appropriately called slow-wave structures or slow-wave circuits.

A simple, yet highly useful, slow-wave circuit can be demonstrated easily. Consider first a transmission line consisting of a single wire above a ground plane as shown in Figure 8.7-1. The propagation characteristics of such a line are well known.[7] An oppositely charged image of the round conductor may be constructed within the ground plane, whereby the behavior of the single-wire-above-ground line becomes identical with the common two-wire line. This line propagates a TEM mode in a direction parallel to the axis of

─────────

[7]Reference 8.2.

the wire at the velocity of light. The TEM designation means that both the electric and magnetic field lines lie entirely in the transverse plane.

Now it is intuitively obvious that gradual bends or twists of the wire above the ground plane, keeping the spacing from wire to ground constant, will have only a minor effect on propagation characteristics of the line. The field lines will faithfully follow the wire, despite such bends. Thus we can imagine the line distorted into the helical coil shown in Figure 8.7-2(a). The requirement that the spacing from wire to ground remain constant is met by having the ground plane become a cylinder enclosing the coil. If the spacing from wire to cylinder is much less than the cylinder diameter and much less than the spacing between turns, the electric field lines from each wire will terminate almost entirely on the adjacent cylinder surface, and the field pattern will be similar to that of Figure 8.7-1.

Since the wave follows the wire at very nearly the velocity of light, the resultant velocity along the axis of the cylinder must be less than the velocity of light. Consequently, an electron can be shot along the cylinder axis at a velocity which enables it to keep in step with the wave. The velocity at which the "in step" electron moves is the phase velocity of the slow-wave circuit. From geometrical considerations, this phase velocity is easily shown to be approximately given by

$$v_p = c\frac{p}{\sqrt{p^2 + (\pi d)^2}} \qquad (8.7\text{-}1)$$

where $d$ and $p$ are the helix diameter and pitch, respectively.

Helices are commonly used as slow-wave circuits in low and medium power traveling-wave tubes. However, generally they are employed without the attendant conducting cylinder surrounding the helix. This causes some quantitative changes in the physical picture presented above, but the basic nature of the slowing process is unchanged. Figure 8.7-2(b) shows the approximate shape of the electric field lines when no outer cylinder is present. For the particular case chosen in the figure, the free-space wavelength of the signal is approximately equal to the length of wire in twelve turns of the helix.

For a structure to be a slow-wave circuit, it is necessary that it possess physical periodicity in the axial direction. That is, there is a finite length, called the period, by which the infinitely long structure must be translated in the axial direction so that one obtains the same structure back again, point for point. In the case of the helical circuit of Figure 8.7-2, for instance, a translation back or forth through a distance of one pitch length results in identically the same structure again. Thus, the period of this helical slow-wave structure is the same as its pitch.

Only periodic structures can propagate slow waves when filled with air or vacuum. It can be shown that smooth, air- or vacuum-filled, nonperiodic structures such as the waveguides of Section 8.4 propagate fast waves only.

## (a)  Floquet's Theorem

Since slow-wave structures are necessarily periodic structures,[8] let us examine some general theorems concerning the solutions of Maxwell's Equations and the relations between phase velocity, group velocity, stored energy, and power flow in periodic structures.

Floquet's Theorem concerns the nature of the single-frequency solutions for the electromagnetic fields obtained from Maxwell's Equations. It may be stated as follows for a periodic structure consisting of identical cells of periodic length $L$ placed end to end.

> The steady-state solutions for the electromagnetic fields of a single propagating mode in a periodic structure have the property that fields in adjacent cells are related by a multiplicative complex constant, this constant being the same for all pairs of adjacent cells.

Mathematically the theorem may be stated as

$$\mathbf{E}(x,y,z - L) = \Gamma\mathbf{E}(x,y,z) \qquad (8.7\text{-}2)$$

where $L$ is the length of one period of the structure, and $\Gamma$ is a complex constant. The direction of propagation is along the $z$ axis, as before. The same expression can be written with $\mathbf{E}$ replaced by $\mathbf{H}$.

The proof of Floquet's Theorem may be obtained by use of the uniqueness theorem[9] of electromagnetic theory which states that the field solutions in two identical microwave structures, operating at the same frequency, can differ only by a complex multiplicative constant, corresponding physically to two different levels of excitation. An analogous situation occurs in ordinary circuit theory where two identical circuits are excited by two different sources at the same frequency. The corresponding phasor currents in the two circuits can differ only by a complex constant, equal to the ratio of the phasors representing the two sources.

Consider the infinitely long periodic structures shown schematically in Figure 8.7-3(a). Each cell is numbered for identification purposes. Assume that the solutions for the electromagnetic fields for a wave propagating to the right have been obtained. Thus, the electric field in cell $n$ may be designated

$$\mathbf{E}_{an}$$

[8]Reference 8f.
[9]Reference 8c, pp. 486–488.

where the first subscript indicates that the solution pertains to the circuit in Figure 8.7-3(a), and the second subscript identifies the cell number.

Now let us consider a second structure obtained from the first by a linear translation in the axial direction of one periodic length as shown in



(a)

(b)
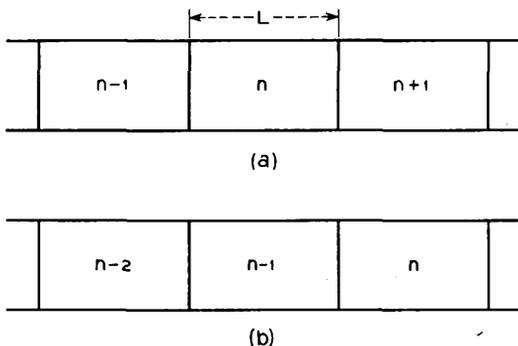
FIG. 8.7-3 Infinitely long periodic structures having identical boundary conditions. (a) The original structure. (b) Structure obtained from the original structure by a linear translation of one period in the axial direction.

Figure 8.7-3(b). Because of the translational symmetry, the new structure will appear identical to the old structure. The uniqueness theorem requires that the fields of structure $b$ be identical to those of structure $a$, except for a constant complex multiplier. That is,

$$\mathbf{E}_{b(n-1)} = \Gamma \mathbf{E}_{an}$$
$$\mathbf{E}_{bn} = \Gamma \mathbf{E}_{a(n+1)}, \text{ etc.} \qquad (8.7\text{-}3)$$

Now we identify structure $b$ by its true nature; it is, after all, merely a translated version of structure $a$ so that the field pattern in structure $b$ is the same as in structure $a$ but translated one period to the right.

$$\mathbf{E}_{b(n-1)} = \mathbf{E}_{a(n-1)}$$
$$\mathbf{E}_{bn} = \mathbf{E}_{an}, \text{ etc.} \qquad (8.7\text{-}4)$$

Combining Equations (8.7-3) and (8.7-4), we get

$$\mathbf{E}_{a(n-1)} = \Gamma \mathbf{E}_{an}$$
$$\mathbf{E}_{an} = \Gamma \mathbf{E}_{a(n+1)}, \text{ etc.} \qquad (8.7\text{-}5)$$

This proves the theorem, since $n$ is, of course, arbitrary.

This simple and highly useful theorem is analogous to theorems concerned with wave propagation in other types of periodic ensembles. For instance, the currents and voltages in an infinite chain of identical filter sections are

governed by the same basic rule; that is, the currents and voltages of one section are equal to the corresponding quantities in the preceding section multiplied by a complex constant. This analogy is often put to use when a microwave periodic structure is represented by an equivalent circuit consisting of such a chain of filter sections.[10]

Let us now rewrite the complex constant $\Gamma$ in Equation (8.7-2) using the defining relationship

$$\Gamma \equiv \epsilon^{j\beta_o L} \tag{8.7-6}$$

so that Equation (8.7-2) becomes

$$\mathbf{E}(x,y,z-L) = \epsilon^{j\beta_o L}\mathbf{E}(x,y,z) \tag{8.7-7}$$

Now, $\beta_o$ could in general be complex. If it were a pure real quantity, it is clear that Equation (8.7-7) implies only a phase shift from one cell to the next. A negative imaginary part to $\beta_o$ would imply a decay in the strength of the fields with distance along the structure, corresponding to ohmic losses. For simplicity, let us assume a lossless structure, so that $\beta_o$ is real. Our results can be generalized later by allowing $\beta_o$ to be complex, if we wish to take losses into account.

Now we shall postulate that the solution to Maxwell's Equations in a periodic structure can be written in the following form

$$\mathbf{E}(x,y,z) = \mathbf{E}_p(x,y,z)\epsilon^{-j\beta_o z} \tag{8.7-8}$$

where $\mathbf{E}_p(x,y,z)$ is a periodic function of $z$ with period $L$. A similar expression holds when $\mathbf{E}$ is replaced by $\mathbf{H}$. Equation (8.7-8) can be proven to be the solution if two conditions are fulfilled. First, it must satisfy the wave equation for the electric field, Equation (8.1-26) and the proper boundary conditions; and second, it must satisfy Floquet's Theorem, Equation (8.7-2). Let us first show that the latter condition is satisfied.

Equation (8.7-8) can be rewritten with $z$ replaced by $z - L$.

$$\mathbf{E}(x,y,z-L) = \mathbf{E}_p(x,y,z-L)\epsilon^{-j\beta_o(z-L)} \tag{8.7-9}$$

Since $\mathbf{E}_p$ is a periodic function with period $L$,

$$\mathbf{E}_p(x,y,z-L) = \mathbf{E}_p(x,y,z) \tag{8.7-10}$$

so that Equation (8.7-9) becomes

$$\mathbf{E}(x,y,z-L) = \mathbf{E}_p(x,y,z)\epsilon^{-j\beta_o z}\epsilon^{+j\beta_o L} \tag{8.7-11}$$

Equation (8.7-8) may be used in the right-hand side of this equation, obtaining

$$\mathbf{E}(x,y,z-L) = \mathbf{E}(x,y,z)\epsilon^{j\beta_o L} \tag{8.7-12}$$

---

[10]Reference 8h, Chapter 4.

But this expression is the mathematical statement of Floquet's Theorem, Equation (8.7-7). Therefore, Equation (8.7-8) does indeed satisfy Floquet's Theorem.

The requirement that the right-hand side of Equation (8.7-8) should satisfy the wave equation will be applied later after we write Equation (8.7-8) in a more convenient form. Since $E_p(x,y,z)$ is periodic in $z$ with period $L$, we can express it by means of a Fourier series:

$$E_p(x,y,z) = \sum_{-\infty}^{\infty} E_n(x,y)\epsilon^{-j(2\pi n/L)z} \tag{8.7-13}$$

This equation is a vector equation, and it is merely a shorthand way of writing three separate equations, one for each vector component. The quantities $E_n$ in the Fourier sum are the usual Fourier coefficients, except that they are functions of the transverse coordinates $x$ and $y$. This may seem strange at first to one who is more familiar with the usual Fourier series in time, where the Fourier coefficients are constants. From this more conventional point of view, Equation (8.7-13) actually represents an infinite number of Fourier series, one for each choice of $x$ and $y$.

Using Equation (8.7-13), the solution for a propagating wave in a periodic structure, Equation (8.7-8) can be written

$$E(x,y,z) = \sum_{-\infty}^{\infty} E_n(x,y)\epsilon^{-j(\beta_0+2\pi n/L)z} \tag{8.7-14}$$

Defining

$$\beta_n = \beta_o + \frac{2\pi n}{L} \tag{8.7-15}$$

we have

$$E(x,y,z) = \sum_{-\infty}^{\infty} E_n(x,y)\epsilon^{-j\beta_n z} \tag{8.7-16}$$

The quantities $E_n(x,y)\epsilon^{-j\beta_n z}$ are known as space harmonics by analogy with time-domain Fourier series. Now we can impose the necessary condition that our solution should satisfy the wave equation, Equation (8.1-26). Substituting Equation (8.7-16) into the wave equation, we obtain

$$\nabla^2\left[ \sum_{-\infty}^{\infty} E_n(x,y)\epsilon^{-j\beta_n z} \right] + k^2\left[ \sum_{-\infty}^{\infty} E_n(x,y)\epsilon^{-j\beta_n z} \right] = 0 \tag{8.7-17}$$

Since the wave equation is linear, we can interchange the order of differentiation and summation, obtaining

$$\sum_{-\infty}^{\infty} [\nabla^2 E_n(x,y)\epsilon^{-j\beta_n z} + k^2 E_n(x,y)\epsilon^{-j\beta_n z}] = 0 \tag{8.7-18}$$

From this equation we see that if each space harmonic is itself a solution of the wave equation, that is, if the bracketed term is zero for each value of $n$, the summation of space harmonics automatically satisfies the wave equation, Equation (8.7-17). Thus each space harmonic is chosen as a solution of the wave equation and consequently must also satisfy Maxwell's Equations. These statements do not imply that each space harmonic satisfies all of the boundary conditions in the structure; only the complete solution, Equation (8.7-16), satisfies this requirement. *Physically, this means that it is impossible to have wave propagation in a periodic structure consisting solely of one space harmonic; a mode of propagation must necessarily consist of an infinite number of space harmonics.*

For simplicity in this section we have considered a periodic structure of infinite length. As in the case of ordinary transmission lines, a finite length structure will have propagation properties identical to those of the infinite structure, except that forward and backward traveling waves must be superimposed to allow for mismatches at the ends of the structure. In nearly all tubes using periodic structures, the structure is matched at both ends so as to eliminate reflected waves.

## (b)   *Field Solutions in a Particular Slow-Wave Structure*

A simple example may help to clarify some of the above points. Let us investigate wave propagation in the periodic structure shown in Figure 8.7-4.[11] This structure consists of two parallel infinite conducting planes.
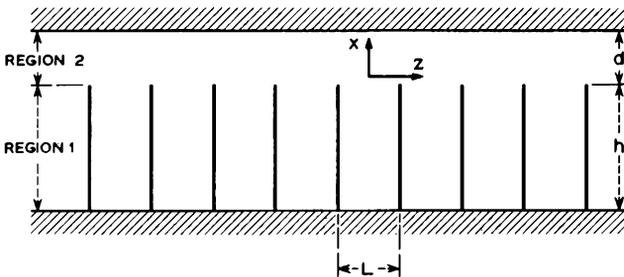


FIG. 8.7-4 A slow-wave structure consisting of thin fins mounted perpendicular to one plate of a parallel-plate line. The direction of propagation is to the right or left.

On the bottom plane are mounted infinitesimally thin conducting fins of height $h$ and infinite width (in the direction perpendicular to the page). The separation from the top of the fins to the top plane is $d$.

---

[11]Reference 8g.

A set of coordinate axes may be chosen as shown, with the origin at the center between two fins. The periodic spacing is $L$. Slow-wave propagation can exist in the $z$ direction.

As in any microwave structure there are many modes of propagation possible. The mode of lowest frequency is often the simplest to analyze, and in most structures it is the most commonly used mode. We shall thus content ourselves with studying the simplest mode of the structure shown in Figure 8.7-4. Since the structure itself is invariant in the $y$ direction, we shall assume the electromagnetic fields are also invariant in this direction. A consequence of this assumption is that spatial derivatives in the $y$ direction must be zero.

It will be convenient to divide the space between the planes into two regions. Region 1 is the space for which $-h \leq x \leq 0$, the region of the vanes. Region 2 is the gap above the vanes for which $0 \leq x \leq d$. We can then solve Maxwell's Equations separately in the two regions, and finally, we can equate the tangential components of electric and magnetic field at the boundary between the two regions, that is, at $x = 0$. Continuity of the tangential electric and magnetic field vectors is necessary, as discussed in Section 8.3. In each region we shall choose our solutions so that the boundary condition at a perfect conductor of zero tangential electric field is satisfied.

Region 1 will be considered first. Consider the unit cell bounded by the two vanes at $z = \pm L/2$. The simplest solution here is a standing-wave solution to Maxwell's Equations consisting of $E_z$ and $H_y$ components only. The desired solution for $E_z$ is

$$E_z = A \sin k(x + h) \tag{8.7-19}$$

where $A$ is an arbitrary constant and $k = \omega/c$. It may be verified that this solution satsifies the wave equation, Equation (8.1-26), and the boundary condition $E_z = 0$ at $x = -h$. The solution for the magnetic field may be obtained from Equation (8.7-19) by use of the second of Equations (8.1-18),

$$j\omega\mu_o H_y = \frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} \tag{8.7-20}$$

Using Equation (8.7-19) and the fact that $E_x$ is zero, we obtain

$$H_y = -j\sqrt{\frac{\varepsilon_o}{\mu_o}}A \cos k(x + h) \tag{8.7-21}$$

It may be verified that the other components of **H** are zero.

Floquet's Theorem, Equation (8.7-7), may be used to find the fields in region 1 in between the other pairs of vanes. If the gaps are numbered in

order with $N = 0$ corresponding to the gap centered at $z = 0$, $N = 1$ to the gap centered at $z = L$, etc., we then have from Floquet's Theorem, in all of region 1,

$$E_z = A \sin k(x + h) \, \epsilon^{-jN\beta_o L} \qquad (8.7\text{-}22)$$

$$H_y = -j\sqrt{\frac{\varepsilon_o}{\mu_o}} A \cos k(x + h) \, \epsilon^{-jN\beta_o L} \qquad (8.7\text{-}23)$$

Next, we proceed to solve Maxwell's Equations in region 2. The general solution is.given by Equation (8.7-16). Let us consider the $z$ component:

$$E_z = \sum_{-\infty}^{\infty} E_{zn}(x)\epsilon^{-j\beta_n z} \qquad (8.7\text{-}24)$$

Each space harmonic will satisfy Maxwell's Equations, or equivalently, the wave equation, Equation (8.1-26), which in our case can be written

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2\right)E_{zn}(x)\epsilon^{-j\beta_n z} = 0 \qquad (8.7\text{-}25)$$

Performing the $z$ differentiation, we obtain

$$\left(\frac{\partial^2}{\partial x^2} - \beta_n{}^2 + k^2\right)E_{zn}(x)\epsilon^{-j\beta_n z} = 0 \qquad (8.7\text{-}26)$$

or simply

$$\left(\frac{\partial^2}{\partial x^2} - \beta_n{}^2 + k^2\right)E_{zn} = 0 \qquad (8.7\text{-}27)$$

This equation has the solution

$$E_{zn} = B_n \sinh \gamma_n(x - C_n) \qquad (8.7\text{-}28)$$

where

$$\gamma_n{}^2 \equiv \beta_n{}^2 - k^2$$

and $B_n$ and $C_n$ are arbitrary constants. The hyperbolic sine solution rather than the trigonometric sine solution has been chosen so that $\beta_n{}^2 > k^2$, since we are looking for slow waves. Since the phase velocity for a space harmonic according to Equation (8.7-24) is given by

$$v_{pn} = \frac{\omega}{\beta_n} \qquad (8.7\text{-}29)$$

we see that $v_{pn} < c$ if $\beta_n > k = \omega/c$.

Equation (8.7-24) may now be written using Equation (8.7-28):

$$E_z = \sum_{-\infty}^{\infty} B_n \sinh \gamma_n(x - C_n) \, \epsilon^{-j\beta_n z} \qquad (8.7\text{-}30)$$

The boundary condition that $E_z = 0$ at $x = d$ can be satisfied if we choose

$$C_n = d \tag{8.7-31}$$

so that

$$E_z = \sum_{-\infty}^{\infty} B_n \sinh \gamma_n (x - d) \, \epsilon^{-j\beta_n z} \tag{8.7-32}$$

Next we equate the two expressions for $E_z$ given by Equations (8.7-22) and (8.7-32) at the boundary between regions 1 and 2. We can simplify this matching technique somewhat by noting that Floquet's Theorem implies that if solutions are matched at the boundary in one cell of a periodic structure, they will be matched in all cells. Let us therefore match over the range $-L/2 \leq z \leq L/2$. We obtain

$$- \sum_{-\infty}^{\infty} B_n \sinh \gamma_n d \, \epsilon^{-j\beta_n z} = A \sin kh$$

for

$$-\frac{L}{2} \leq z \leq \frac{L}{2} \tag{8.7-33}$$

The coefficients $B_n$ can be obtained by the following process. Multiply both sides of the equation by $\epsilon^{+j\beta_m z}$ and integrate over a period:

$$- \sum_{-\infty}^{\infty} B_n \sinh \gamma_n d \int_{-L/2}^{L/2} \epsilon^{j(\beta_m - \beta_n)z} dz = A \sin kh \int_{-L/2}^{L/2} \epsilon^{j\beta_m z} dz \tag{8.7-34}$$

The right-hand side is easily integrated. The left-hand side can be manipulated as follows, using Equation (8.7-15):

$$\int_{-L/2}^{L/2} \epsilon^{j(\beta_m - \beta_n)z} dz = \int_{-L/2}^{L/2} \epsilon^{j(2\pi/L)(m-n)z} dz = \begin{cases} 0 \text{ for } n \neq m \\ L \text{ for } n = m \end{cases} \tag{8.7-35}$$

Equation (8.7-34) thus becomes

$$-B_m L \sinh \gamma_m d = A \frac{\sin \dfrac{\beta_m L}{2}}{\dfrac{\beta_m}{2}} \sin kh \tag{8.7-36}$$

By substituting Equation (8.7-36) into Equation (8.7-32), we obtain for $E_z$ in region 2

$$E_z = -A \sin kh \sum_{-\infty}^{\infty} \frac{\sin \dfrac{\beta_n L}{2}}{\dfrac{\beta_n L}{2}} \frac{\sinh \gamma_n (x - d)}{\sinh \gamma_n d} \epsilon^{-j\beta_n z} \tag{8.7-37}$$

We thus have a complete description of the $z$ component of electric field in terms of an arbitrary amplitude factor $A$.

The other components of the electric and magnetic fields in region 2 may be obtained as follows. First, because $E_y = 0$ in region 1, it will also be zero in region 2, since tangential components of $\mathbf{E}$ are continuous at the boundary.

Next, the divergence equation, Equation (8.1-16), of Maxwell's Equations is written (using Equation (2) of Appendix XII) as

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_z}{\partial z} = 0 \qquad (8.7\text{-}38)$$

since there is no free charge in the region. This equation can be solved for $E_x$:

$$E_x = -\int \frac{\partial E_z}{\partial z} \partial x \qquad (8.7\text{-}39)$$

Performing the indicated operations on Equation (8.7-37), we obtain

$$E_x = -jA \sin kh \sum_{-\infty}^{\infty} \frac{\sin \dfrac{\beta_n L}{2}}{\dfrac{\gamma_n L}{2}} \frac{\cosh \gamma_n(x - d)}{\sinh \gamma_n d} \epsilon^{-j\beta_n z} \qquad (8.7\text{-}40)$$

The components of the magnetic field can be obtained by use of Equations (8.1-18). Since $E_y = 0$, and derivatives with respect to $y$ are also zero, we see that

$$H_x = H_z = 0 \qquad (8.7\text{-}41)$$

Equations (8.7-37), (8.7-40), and (8.1-18) together give, after simplification,

$$H_y = -j\omega\varepsilon_o A \sin kh \sum_{-\infty}^{\infty} \frac{\sin \dfrac{\beta_n L}{2}}{\dfrac{\gamma_n \beta_n L}{2}} \frac{\cosh \gamma_n(x - d)}{\sinh \gamma_n d} \epsilon^{-j\beta_n z} \qquad (8.7\text{-}42)$$

At this point we have a complete description of the fields in the slow-wave structure, assuming that we know what value of $\beta_o$ corresponds to a given frequency of operation. All the values of $\beta_n$ and $\gamma_n$ can be obtained from $\beta_o$ using Equation (8.7-15) and the relation defining $\gamma_n$,

$$\gamma_n{}^2 = \beta_n{}^2 - k^2 \qquad (8.7\text{-}43)$$

A sketch of the electric field lines for $\beta_o L = \pi/10$ is shown in Figure 8.7-5.

An equation determining $\beta_o$ from the frequency may be obtained by matching the tangential components of the magnetic field at the boundary between regions 1 and 2. However, at this point we must note that the
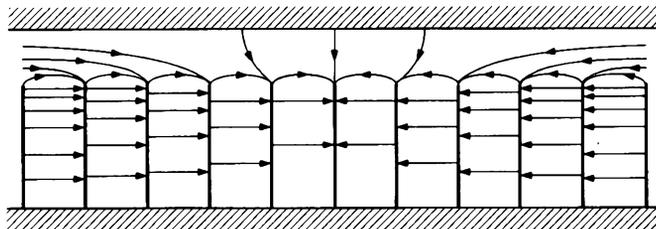
FIG. 8.7-5 Electric field lines in the slow-wave structure of Figure 8.7-4 for $\beta_o L = \pi/10$.

solutions we have obtained are only approximate, due to the neglect of fringing fields near the vane tips in region 1. Because of this approximation, the magnetic fields in the two regions will not match point for point at the boundary. Let us therefore content ourselves with matching at the mid-point of the gap, where $z = 0$.

Equating Equations (8.7-21) and (8.7-42) for $z = 0$ and $x = 0$, we obtain

$$\frac{\cot kh}{kh} = \sum_{-\infty}^{\infty} \frac{\sin \dfrac{\beta_n L}{2} \coth \gamma_n d}{\dfrac{\beta_n L}{2} \quad \gamma_n h} \tag{8.7-44}$$

where

$$k = \omega\sqrt{\mu_o \varepsilon_o}$$
$$\beta_n = \beta_o + \frac{2\pi n}{L}$$
$$\gamma_n = \sqrt{\beta_n{}^2 - k^2}$$

The solutions to this equation are obtained numerically. We shall discuss the resultant $\omega$-$\beta$ diagram in the next section.

Let us review briefly what we have accomplished in this section. We have used approximate solutions to Maxwell's Equations in a periodic structure to obtain the slow-wave propagation fields. One may wonder as to the effect of the approximations involved. It turns out that the resultant $\omega$-$\beta$ diagram is relatively insensitive to small errors in the shapes of the field solutions, so that information derived from the $\omega$-$\beta$ diagram can be taken to be quite accurate. The exact shape of the fields will be somewhat in error, but this information is usually needed only approximately.

### (c)   The Brillouin Diagram

We have seen in Sections (a) and (b) that the electric or magnetic field for a propagating mode in a slow-wave structure can be expanded as a

summation of space harmonics, as in Equation (8.7-16), which we repeat
here.

$$\mathbf{E}(x,y,z) = \sum_{-\infty}^{\infty} \mathbf{E}_n(x,y)\epsilon^{-j\beta_n z} \tag{8.7-45}$$

where

$$\beta_n = \beta_o + \frac{2\pi n}{L} \tag{8.7-46}$$

Each space harmonic propagates in the positive $z$ direction with a
different phase velocity given by

$$v_{pn} = \frac{\omega}{\beta_n} \tag{8.7-47}$$

Therefore, the mode of propagation cannot be characterized at some
frequency by a unique velocity as it was in the case of ordinary smooth
waveguides. Referring to our previous interpretation of the phase velocity,
we see that *it is no longer possible for an "observer" to move so as to be always
in the same phase of the total field. It is possible for the "observer" to move
in synchronism with only one of the space harmonics that make up the total
field.* The phases of the other space harmonics will be continually changing
as viewed by the "observer." If the "observer" takes a time average of the
total field that he sees over a sufficiently long period of time as he moves in
synchronism with one of the space harmonics, the average obtained will
be that given by the synchronous space harmonic alone, the net contri-
bution of the others being negligible in comparison. It will be useful in
later chapters to bear in mind this interpretation of the phase velocity.

Let us now plot the $\omega$-$\beta$ curve for the periodic structure of Figure 8.7-4.
We will want to make sure we include values of the propagation constant
$\beta_n$ for all of the space harmonics. This $\omega$-$\beta$ diagram is known as a Brillouin
diagram.[12] It is customary to label the abscissa as the $\beta$ axis instead of the
$\beta_n$ axis. Each branch of the Brillouin diagram is numbered according to
the space harmonic to which it refers.

The fundamental space harmonic $(n = 0)$ propagation constant is
obtained as a function of frequency from Equation (8.7-44). Since this is a
transcendental equation, there will be an infinite number of frequencies
or modes of propagation for each value of $\beta$. This infinite number of modes
should come as no surprise, since we first encountered them in the analysis
of ordinary waveguides. The $\omega$-$\beta$ curves for the fundamental space har-
monics are shown in Figure 8.7-6, including the higher-order modes. We have

[12]After L. Brillouin who studied extensively wave propagation in periodic structures.
See Reference 8f.

included only the modes propagating in the positive $z$ direction for the time being. The negative propagating modes have branches which are the mirror images of these about the $\omega$ axis, as in Figure 8.5-4. We have labeled the



lower branch with a zero, indicating that it corresponds to the fundamental $(n = 0)$ space harmonic. We shall omit labeling the higher-order modes. Physically, these higher modes correspond approximately to additional half wavelengths in region 1 of Figure 8.7-4.

Now from Equation (8.7-46), we see that the Brillouin diagram branches for the other space harmonics are obtained by taking the fundamental space harmonic branches in Figure 8.7-6 and translating them parallel to the $\beta$ axis through distances which are integral multiples of $2\pi/L$. Figure 8.7-7 shows this construction utilized to obtain the plus-one and minus-one space harmonics.

FIG. 8.7-6 The branches of the Brillouin diagram corresponding to the fundamental space harmonics of modes propagating in the positive $z$ direction.

It would be enlightening at this point to consider the group velocity in a periodic structure. It is defined as in Section 8.5 for ordinary waveguides (Equation (8.5-27)):

$$v_g = \frac{\partial \omega}{\partial \beta} \qquad (8.7\text{-}48)$$

It has the same physical significance as before; that is, it is the velocity at which energy is transported down the periodic structure. Since all of the space harmonics must be taken together to constitute a mode of propagation, we would expect all of them to have the same group velocity, corresponding to the velocity of energy transport. A glance at Figure 8.7-7 shows that this is indeed the case. All of the branches for any mode of propagation have the same slope at any given frequency, hence, the same group velocity. Group and phase velocities are measured by geometrical constructions as in Figure 8.5-2. It should be noted that no equation similar to Equation (8.5-28) exists for slow-wave structures.

Figure 8.7-7 also shows that the minus space harmonics $(n = -1, -2,$ etc.) have phase velocities that are negative, albeit the group veloc-

FIG. 8.7-7 Branches of the Brillouin diagram for the three lowest space harmonics of modes propagating in the positive $z$ direction.

ity is positive. This is an interesting property of periodic structures which has no parallel in the smooth waveguide case. This means that an electron "observer" can remain in synchronism with a wave which is actually transporting energy in the opposite direction. This remarkable property has made possible the backward-wave oscillator and M-Carcinotron which will be described in later chapters.

Because the slow-wave structure shown in Figure 8.7-4 consists of two separate conducting members, it has no lower cutoff frequency, and in fact it propagates signals with frequencies ranging down to zero frequency, as is evident from the Brillouin diagram.

The branches marked $n = -1, 0$, and $+1$ in Figure 8.7-7 correspond to the principal mode of propagation. The broken lines in the figure have slopes corresponding to phase velocities of $+c$ and $-c$, where $c$ is the velocity of light. We see that all spatial harmonics of the principal mode of propagation lie either to the right or to the left of the $v_p = \pm c$ lines rather than between these lines. This means that the phase velocities of the fundamental and higher-order space harmonics of the principal mode are of magnitude less than the velocity of light. Furthermore, the phase velocity of the $n = 1$ space harmonic of the principal mode is less than that of the fundamental, or $n = 0$ space harmonic, and the phase velocity of the $n = 2$ space harmonic is less than that of the $n = 1$ space harmonic.

So far we have considered energy propagating solely in the positive $z$ direction. The branches for the negative propagating modes are obtained by simply reflecting all the branches of Figure 8.7-7 about the $\omega$ axis, as in Figure 8.7-8, where the complete Brillouin diagram is shown. It is seen that
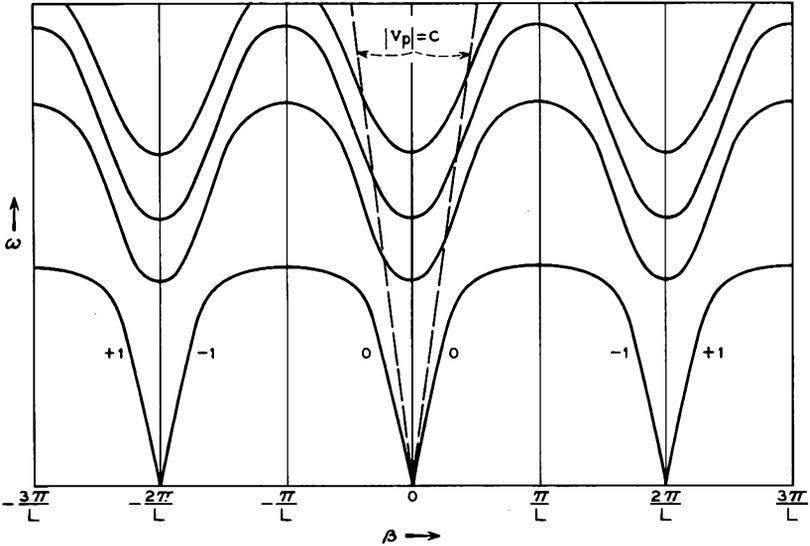


FIG. 8.7-8 Complete Brillouin diagram for the periodic structure of Figure 8.7-4.

all of these additional branches have negative group velocities, as expected. The numbers of these additional branches are chosen so as to correspond to the reflected branches.

### (d)  Power Flow

In order to complete our discussion of periodic structures we must consider a means of calculating power flow from a knowledge of the electromagnetic fields of a propagating mode.

Equation (8.5-26) states that the power flow in a smooth waveguide is given by the product of the group velocity and the energy stored per unit length.[13] Now, for a lossless periodic structure Floquet's Theorem, Equation (8.7-7), states that the fields in all cells are equal in magnitude, differing only in phase. This means that the stored energy in each unit cell

---

[13]In cases where the electromagnetic fields are known only approximately, this gives a more accurate evaluation of the power flow than does integration of the Poynting vector.

of the periodic structure is the same as that in all the others. Hence, the stored energy per unit length can be simply calculated by taking the stored energy in any one unit cell and dividing by the length of the cell. With this adaptation, Equation (8.5-26) can be used to calculate the power flow in a periodic structure. It may be written as

$$P = v_g \frac{W_L}{L} \tag{8.7-49}$$

where $W_L$ is the time average energy stored per cell and $L$ is the length of the cell.

In calculating the average stored energy per period, it is convenient to realize that the time average stored electric energy per period is equal to the time average stored magnetic energy. Thus it is necessary to calculate only the average stored energy due to either the magnetic or electric fields and multiply this by two. This relationship can be proved rigorously for a periodic structure, but the proof is rather long and complicated.[14]

As an example, we may compute the power flow per unit width in the periodic structure of Figure 8.7-4. We have already solved for the field components, finding expressions for $E_z$ and $H_y$ in region 1 and $E_x$, $E_z$, and $H_y$ in region 2, the other components being zero. In finding the stored energy it will be easier to use the magnetic field expressions, since only one component is involved.

The time average stored energy per cell (see Appendix XIV) is given by

$$W_L = \frac{\mu_o}{2} \int_{\substack{\text{unit cell}}} |\mathbf{H}|^2 dv \tag{8.7-50}$$

Since the structure is of infinite width, we shall determine only the power flow per unit width, designated

$$W_{Lw} = \frac{\mu_o}{2} \iint_{\substack{\text{unit cell}}} |H_y|^2 dz dx \tag{8.7-51}$$

where use is made of the fact that $H_x = H_z = 0$.

The contribution to this integral in region 1 is obtained by using Equation (8.7-21):

$$\begin{aligned} W_{Lw1} &= \frac{1}{2}\epsilon_o L A^2 \int_{-h}^{0} \cos^2 k(x+h) dx \\ &= \frac{1}{4}\epsilon_o L h A^2 \left[ 1 + \frac{\sin 2kh}{2kh} \right] \end{aligned} \tag{8.7-52}$$

[14]Reference 8g, pp. 10–14.

In region 2, $H_y$ is given by Equation (8.7-42), repeated below:

$$H_y = -j\omega\varepsilon_o A \sin kh \sum_{-\infty}^{\infty} \frac{\sin\frac{\beta_n L}{2}}{\frac{\gamma_n \beta_n L}{2}} \frac{\cosh \gamma_n(x - d)}{\sinh \gamma_n d}\epsilon^{-j\beta_n z} \qquad (8.7\text{-}53)$$

$|H_y|^2$ is obtained by multiplying this quantity by its conjugate,

$$H_y{}^* = +j\omega\varepsilon_o A \sin kh \sum_{-\infty}^{\infty} \frac{\sin\frac{\beta_m L}{2}}{\frac{\gamma_m \beta_m L}{2}} \frac{\cosh \gamma_m(x - d)}{\sinh \gamma_m d}\epsilon^{+j\beta_m z} \qquad (8.7\text{-}54)$$

Equation (8.7-51) becomes, for region 2,

$$W_{Lw2} = \tfrac{1}{2}\mu_o\omega^2\varepsilon_o{}^2 A^2 \sin^2 kh \sum_m \sum_n \int_0^d \int_{-L/2}^{L/2} C_m C_n \epsilon^{j(2\pi/L)(m-n)z} dz dx$$

$$(8.7\text{-}55)$$

where we have written

$$C_n = \frac{\sin\frac{\beta_n L}{2}}{\frac{\gamma_n \beta_n L}{2}} \frac{\cosh \gamma_n(x - d)}{\sinh \gamma_n d}$$
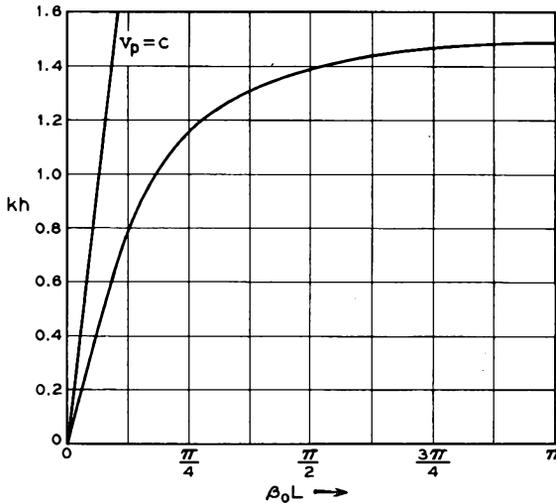


FIG. 8.7-9 Fundamental branch of the Brillouin diagram for the slow-wave structure of Figure 8.7-4, with dimensions given by $h/d = 4$ and $h/L = 5$.

Performing the $z$ integration, using Equation (8.7-35), we obtain

$$W_{Lw2} = \frac{1}{2}\mu_o\omega^2\varepsilon_o{}^2 A^2 L \sin^2 kh \sum_{-\infty}^{\infty} \frac{\sin^2\dfrac{\beta_n L}{2}}{\gamma_n{}^2\left(\dfrac{\beta_n L}{2}\right)^2} \int_0^d \frac{\cosh^2 \gamma_n(x-d)}{\sinh^2 \gamma_n d}dx$$

$$(8.7\text{-}56)$$

Integrating this expression and simplifying, we obtain

$$W_{Lw2} = \frac{1}{4}\varepsilon_o L d A^2 \sin^2 kh \sum_{-\infty}^{\infty} \frac{k^2}{\gamma_n{}^2}\left(\frac{\sin\dfrac{\beta_n L}{2}}{\dfrac{\beta_n L}{2}}\right)^2 \frac{1 + \dfrac{\sinh 2\gamma_n d}{2\gamma_n d}}{\sinh^2 \gamma_n d} \quad (8.7\text{-}57)$$

The total time average stored energy is the sum of that given by Equations (8.7-52) and (8.7-57). Although Equation (8.7-57) is complicated, it is easily evaluated, since the series converges quite rapidly.

Next, we can compute the power flow, using Equation (8.7-49) for a particular geometry of the finned structure shown in Figure 8.7-4. Consider a structure with dimensions chosen such that $h/d = 4$ and $h/L = 5$. Equa-
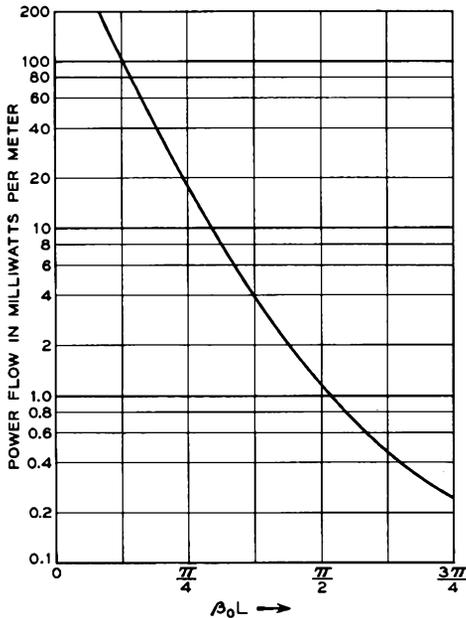


Fig. 8.7-10 Power per unit width in the structure of Figure 8.7-4 for a vane tip-to-tip voltage of one volt. $h/d = 4$ and $h/L = 5$.

tion (8.7-44) is solved numerically to obtain the fundamental space har-
monic branch of the Brillouin diagram for the lowest frequency mode.
This is shown in Figure 8.7-9. A line with slope equal to the velocity of
light is drawn in for reference. We note, as before, that the phase velocity
at every point on the curve is less than the velocity of light. The upper
cutoff frequency is approximately that for which the vanes are a quarter
wavelength long. We might have expected this, since the input impedance
to a quarter wavelength shorted line is infinite, presenting an open circuit as
far as axial current flow is concerned.

Having thus determined the relationship between $\omega$ and $\beta_o$, we can pro-
ceed to calculate the power flow. It will be more interesting here to deal
with an actual structure, designed for a specific operating frequency. Let
us choose dimensions such that the mode cuts off at 10 Gc. This occurs for

$$h = 0.706 \text{ cm}$$

$$L = 0.141 \text{ cm} = h/5$$

$$d = 0.1766 \text{ cm} = h/4$$

The power flow per unit width of the structure is presented in Figure
8.7-10 as a function of $\beta_o L$. This curve is obtained by multiplying the
group velocity by the time average stored energy per unit length, where
the former quantity is obtained by measuring slopes in Figure 8.7-9. The
power flow given is that amount required to produce a peak voltage of
one volt from one vane tip to the next. From Equation (8.7-19) we see that
this occurs for

$$E_z(0)L = AL \sin kh = 1 \text{ volt}$$

The power flow goes to zero as $\beta L$ approaches $\pi$ because the group velocity
goes to zero. On the other hand, as $\beta L$ goes to zero, we approach the dc
condition where the top plate is all at one potential and the fins are all at
the opposite potential. It becomes more and more difficult to maintain a
voltage difference of one volt from one vane to the next, and the power
required becomes infinitely large.

### PROBLEMS

8.1 The equivalent circuit for the cavity in Figure 8-2 is given by a resistance,
capacitance, and inductance in parallel. (a) Calculate the values of the capacitance
and inductance at 3000 Mc for the dimensions given in the text. (b) The magnetic
field in the inductive sections of the cavity can be written

$$H_\theta = \frac{I}{2\pi r} \frac{\cos\frac{\omega}{c}z}{\cos\frac{\omega}{c}l}$$

where $z$ is measured from the top or bottom wall, respectively, and $I$ is the current through the equivalent circuit inductance. Using Equation (8.4-19), find an expression for the ohmic power loss in the cavity at resonance for the dimensions given in the text. From this power loss find an expression for the resistance in the equivalent circuit. Neglect losses in the capacitive region of the cavity.

8.2 Scaling laws apply exactly to microwave structures whose walls are perfect conductors and approximately to others. The scaling law may be stated mathematically as follows. If $\mathbf{E}(x,y,z,\omega t)$ is a solution to the wave equation, then $\mathbf{E}(Kx,Ky,Kz, K\omega t)$ is also a solution, where $K$ is a numerical constant. Demonstrate the validity of this statement.

8.3 Suppose that the cavity of Problem 8.1 is scaled to be resonant at $K$ times 3000 Mc. (a) What are the resistance, capacitance, and inductance of the equivalent circuit for the new cavity, assuming the cavity walls are made of the same material? (b) What is the ratio of the $Q$'s of the two circuits, where $Q = R/\omega L$.

8.4 By applying Stoke's theorem to Equation (8.1-14) show that the component of magnetic field perpendicular to a perfect conductor is zero, given that the parallel component of electric field is zero.

8.5 Show that the resistive power loss in a conductor may be derived assuming a uniform current density in a wall of a thickness equal to the skin depth, where the wall current per unit width is given by

$$I_s = \int_0^\infty J_s dx$$

and $J_s$ is given by Equation (8.4-13).

8.6 Slow-wave structures may sometimes be represented by an equivalent circuit consisting of a uniform lossless transmission line periodically loaded by either a series or a shunt reactance. If $Z_o = 1/Y_o$ is the characteristic impedance and $\varphi_o$ is the phase shift per period of the unloaded line, then periodic shunt loading due to a susceptance $B$ results in the relation

$$\cos \beta_o L = \cos \varphi_o - \frac{B}{2Y_o} \sin\varphi_o$$

where $\beta_o L$ is the periodic phase shift of the periodically loaded structure. Similarly, for periodic series loading due to a reactance $X$, one obtains

$$\cos \beta_o L = \cos \varphi_o - \frac{X}{2Z_o} \sin \varphi_o$$

Prove either one of these relationships using the fact that corresponding voltages and currents in adjacent cells are related by the factor $\epsilon^{-j\beta_o L}$. Use the results of uniform transmission line theory which state that the input and output voltages and currents for a line of electrical length $\varphi_o$ are related by:

$$V_{in} = V_{out}\cos \varphi_o + jI_{out}Z_o\sin \varphi_o$$

$$I_{in} = I_{out}\cos \varphi_o + jV_{out}Y_o\sin \varphi_o$$

8.7 How is Equation (8.7-37) modified if the vanes of Figure 8.7-4 have a finite thickness $\Delta$?

8.8 From Equation (8.7-37) compute the relative magnitudes of the $n = 0$, $\pm 1$, $\pm 2$ space harmonics at the frequency for which $\beta_o L = \pi/2$. Compute them at the value of $x$ for which they are a maximum.

8.9 A slow-wave structure has an equivalent circuit consisting of a cascade of filter sections whose shunt arm is a pure capacitance $C$ and whose series arm is an open-circuited transmission line. The transmission line has a characteristic admittance of $\frac{1}{2}\omega_o C$, where $\omega_o$ is the lowest radian resonant frequency of the line; i.e., the susceptance of the series arm is given by $B_1 = \frac{1}{2}\omega_o C \tan (\pi\omega/2\omega_o)$. Sketch the Brillouin diagram over the range $0 \leq \omega \leq 4\omega_o$ and $-2\pi \leq \beta L \leq 2\pi$. Determine the cutoff frequencies accurately and then qualitatively sketch in the curves. Make use of the filter formula $\cos \beta_o L = 1 + B_2/2B_1$, where $B_2$ is the susceptance of the shunt arm.

8.10 By studying the symmetries of a slow-wave structure one may deduce certain facts about its space harmonics. In the figure is shown an interdigital line, assumed to be infinitesimally thin in the $y$ direction. This structure has a symmetry such that a translation of $L/2$ in the $z$ direction accompanied by a reflection about the $y$-$z$ plane results in the structure mapping back onto itself.
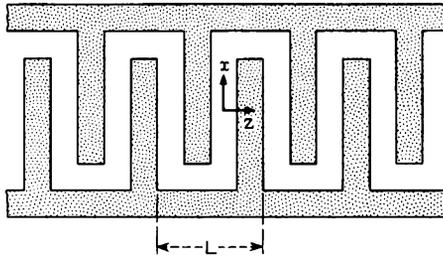


Problem 8.10

The solution for the electric field near $x = 0$ is of the form

$$E_z = \sum_{-\infty}^{\infty} (A_n \cos kx + B_n \sin kx)\epsilon^{-|\gamma_n y|}\epsilon^{-j\beta_n z}$$

Because of the symmetry described above we can replace $z$ by $z + (L/2)$ and $x$ by $-x$, and the resulting expression for $E_z$ can differ from the original only by a complex constant. Use these facts to show that either $A_n = 0$ for $n$ odd and $B_n = 0$ for $n$ even (symmetric mode), or else $A_n = 0$ for $n$ even and $B_n = 0$ for $n$ odd (antisymmetric mode).

## REFERENCES

Four general references on electromagnetic field theory and its application to waveguides and cavities are:

8a. S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio,* John Wiley and Sons, Inc., New York, 1959.

8b. C. G. Montgomery, R. H. Dicke, and E. M. Purcell, "Principles of Microwave Circuits," *MIT Radiation Laboratory Series*, Vol. 8, McGraw-Hill Book Co., Inc., New York, 1948.

8c. J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill Book Co., Inc., New York, 1941.

8d. H. H. Skilling, *Fundamentals of Electric Waves*, John Wiley and Sons, Inc., New York, 1948.

8e. G. C. Southworth, *Principles and Applications of Waveguide Transmission*, D. Van Nostrand Co., Inc., Princeton, N. J., 1950.

The following references contain material on slow-wave structures:

8f. L. Brillouin, *Wave Propagation in Periodic Structures*, Dover Publications, Inc., New York, 1953.

8g. D. A. Watkins, *Topics in Electromagnetic Theory*, Chapter 1, John Wiley and Sons, Inc., New York, 1958.

8h. J. R. Pierce, *Traveling-Wave Tubes*, D. Van Nostrand Co., Inc., Princeton, N. J. 1950.

8i. J. C. Slater, *Microwave Electronics*, Chapter 8, D. Van Nostrand Co., Inc., Princeton, N. J., 1950.

8j. A. H. W. Beck, *Space-Charge Waves and Slow Electromagnetic Waves*, Chapter 3, Pergamon Press, Inc., New York, 1958.

8k. E. Belohoubek, "Propagation Characteristic of Slow-Wave Structures Derived from Coupled Resonators," *RCA Review* **19**, 283-310, June, 1958.

Other references covering specific items discussed in the chapter are:

8.1 K. Fujisawa, "General Treatment of Klystron Resonant Cavities," *Trans. IRE*, **MTT–6**, no. 4, 344–358, October, 1958.

8.2 H. H. Skilling, *Electric Transmission Lines*, McGraw-Hill Book Co., Inc., New York, 1951.

8.3 R. B. Adler, "Properties of Guided Waves on Inhomogeneous Cylindrical Structures," *RLE Tech. Report No.* 102 (MIT), 1949.

8.4 W. W. Hansen, "A Type of Electrical Resonator," *J. Appl. Phys.* **9**, 654–663, October, 1938.

# Chapter 9

## KLYSTRONS

We have seen in Chapter 7 the deleterious effects that occur in conventional triodes and tetrodes as the signal frequency is increased. The factors degrading gain-bandwidth product and power output in these tubes may be divided into two categories:

1. *Circuit Factors*

   These include lead inductance, stray capacitance, and power losses due to radiation, dielectric loss factor, and resistance.

2. *Transit-Time Effects*

   These effects occur because of the finite time electrons take to travel between electrodes.

The losses due to circuit factors can be reduced by a judicious use of the microwave components discussed in the preceding chapter. On the other hand, one encounters certain fundamental difficulties in trying to minimize transit-time effects. In the triode and the tetrode, it is the cathode-to-grid transit time which is the real culprit degrading the high-frequency gain and efficiency. One can decrease this transit time by decreasing the cathode-to-grid spacing. This approach has been used successfully in the Western Electric 416B triode, described in Section 7.4. But as one can see from the dimensions of this tube, as given in Section 7.4, it is unlikely that the operating frequency could be extended much higher by further reduction of electrode spacings. Accordingly, one must seek other means for modulating the electron beams in tubes operating at high microwave frequencies.

In the present chapter we shall describe two microwave tubes which make use of a second type of modulation called *velocity modulation*. Velocity modulation is obtained by impressing a small ac component of velocity

on a dc electron beam. This can be done by allowing the beam to pass through two grids between which there is applied a small ac voltage. If the grids are spaced very close together, equal numbers of electrons emerge from the grids in equal intervals of time, but the velocity of the electrons has a small ac component. Such a beam is said to be velocity modulated. As the electrons travel away from the grids, the faster electrons move away from the slower electrons behind them and tend to overtake the slower electrons ahead of them. The axial density of electrons is therefore no longer uniform, and the beam current passing a point some distance from the grids has an ac component. In view of this, it is frequently said that the velocity modulation imparted to the beam when it passed through the grids gives rise to current modulation farther along the beam.
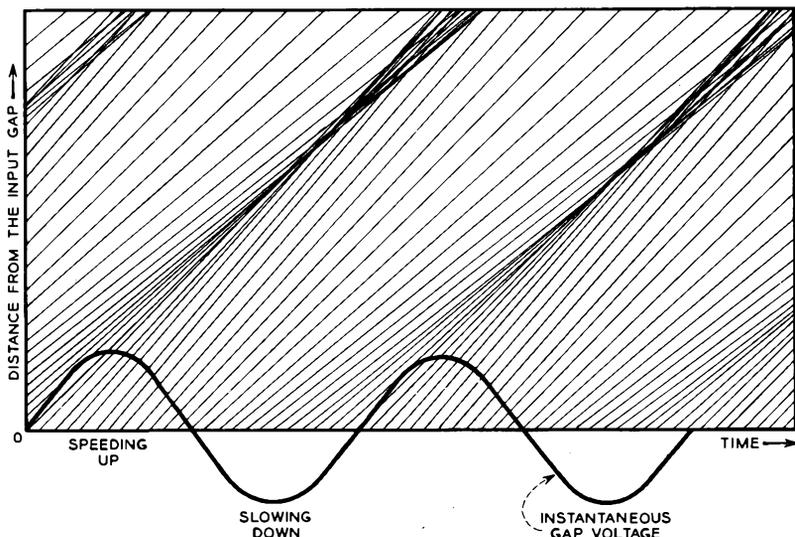


FIG. 9-1 Applegate diagram showing representative electron trajectories. The slope of each trajectory is proportional to the electron velocity. Velocity modulation is produced at the gap by the changing gap voltage. This results in density modulation beyond the gap.

The velocity modulation is illustrated in Figure 9-1, known as an Applegate diagram. In this figure, plots of distance vs. time are given for a number of representative electrons (24 per cycle). The effects of space-charge forces are neglected in drawing the figure. The electrons leave the grids spaced uniformly in time, corresponding to the lack of current modulation at this point, However, each electron has a slightly different velocity,

depending on the instantaneous rf voltage between the grids when the electron passed through the grids. The instantaneous voltage between the grids is indicated on the figure. The slope of a trajectory is proportional to the electron velocity. Because of the difference in slopes, many of the trajectories converge so as to form electron bunches at some distance from the grids. We note that the bunches tend to form about an electron which goes through the grids when the voltage is zero and increasing. Similarly, the electrons tend to move away from an electron which goes through the grids when the voltage is zero and decreasing; this electron locates what is termed the antibunch.

If the ac voltage applied between the grids is of a very high frequency, the distance along the beam between maxima and minima in velocity will, of course, be very short. This means that appreciable density variations
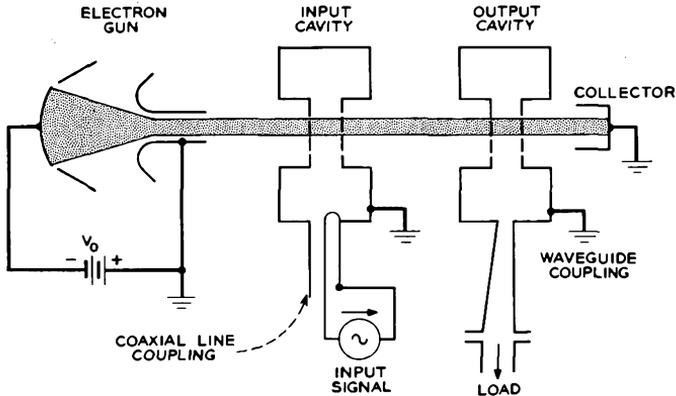


FIG. 9-2 Two-cavity klystron amplifier.

will appear after the electrons have traveled a relatively short distance from the grids. Velocity modulation, therefore, lends itself particularly to high-frequency tubes. The two velocity modulated tubes described in the present chapter, the klystron amplifier and the reflex oscillator, are generally designed for operation at frequencies above 200 Mc. Reflex klystron oscillators have been built which give useful output at frequencies greater than 100,000 Mc, or 100 Gc.

In both these tube types the beam passes through grids that are an integral part of a resonant cavity. If the cavity is excited, the voltage developed across the cavity, and hence between the grids, imparts the velocity modulation to the beam. Power is extracted from the beam in the case of the klystron amplifier by allowing the beam to pass through a second resonant cavity. The cavity is excited by the induced currents associated

with the beam just as in the case of an external resonant circuit connected between a pair of grids. In the reflex oscillator the beam is caused to return through the modulating cavity with the right phase so that it adds to the excitation energy of the cavity.

Figure 9-2 illustrates a two-cavity klystron amplifier. In this particular example the rf signal is coupled into the input cavity by means of a coaxial cable. The output cavity is coupled to the load by means of a waveguide, taking advantage of the lower attenuation inherent in waveguide.

The electron beam is produced by an electron gun of the type shown in Figure 4.5-1(a). This is a convergent Pierce gun which produces a small diameter beam from a cathode of much larger diameter. Thus, much higher beam current densities are available for a given cathode electron-emission density than in a triode or tetrode. This allows a large increase in the beam power passing through electrode gaps of a fixed area and capacitance and hence a large increase in the gain-bandwidth product which can be achieved with such a tube. The klystron is usually operated with the cathode at a negative potential and the other electrodes grounded, for reasons of convenience and safety.

Since the electrons must travel a considerable distance, the beam is prevented from spreading radially, due to the space charge repulsion, by applying an axial dc magnetic field. This field is provided by a permanent magnet or solenoid, as discussed in Section 3.4.

After passing through the output cavity, the beam strikes a collector electrode. The function of the collector electrode could be performed by replacing the second grid of the output cavity with a solid piece of metal. However, having a separate electron collector has several advantages. First, the collector can be made as large as is desired in order to collect the beam at a lower power density, thus minimizing localized heating. If the collector were part of the rf circuit, its size would be limited by the maximum gap capacitance consistent with good high-frequency performance. Second, by having a separate collector, its potential can be reduced considerably below the beam potential in the rf interaction region, thus reducing the power dissipated in the collector and increasing the overall efficiency of the device. It should be clear that the electron beam does not extract energy from any dc power supply unless the electrons are actually collected by an electrode connected to that power supply. Thus in Figure 9-2, if a separate power supply were connected between cathode and collector and if the cavity grids intercepted a negligible part of the beam, the power supply between the cathode and collector would be the only one supplying any power to the tube.

It is clear that the two-cavity klystron amplifier has considerable advantage over the conventional triode and tetrode for microwave signal ampli-

fication. Circuit losses are greatly reduced by the use of resonant cavities at the interaction gaps and by the use of microwave transmission lines for making input and output connections. Furthermore, transit-time effects, which limit the high-frequency performance of triodes and tetrodes, are largely overcome by the use of velocity modulation. In the following sections, we shall take a more quantitative look at the electron interaction process in the klystron amplifier. Later in the chapter we shall describe the reflex klystron oscillator.

### 9.1   Quantitative Theory of Klystron Interaction

The quantitative theory of klystron interaction may be conveniently divided into three parts, as follows:

1. The velocity modulation produced by a given voltage at the input cavity.
2. The current modulation at the output cavity resulting from the initial velocity modulation at the input cavity.
3. The current induced in the output cavity by the current modulation on the beam.

The first and third parts have to do with the interaction between an electron beam and the grids of a cavity. (The region between the grids of a cavity is known as the cavity gap.)

(a)   *Velocity Modulation Produced by an RF Voltage Applied to the Grids of a Cavity*
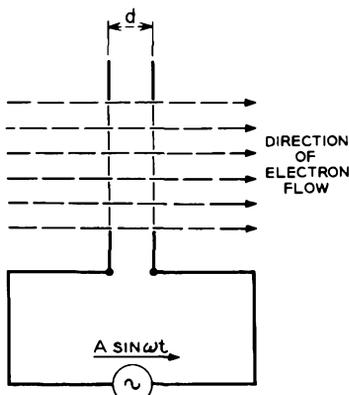


Fig. 9.1-1 Klystron buncher gap with rf voltage applied. Velocity modulation is produced on the electron beam.

The grids of the input cavity are represented in the equivalent circuit of Figure 9.1-1. An rf voltage source is shown connected to these grids. This voltage source is an equivalent source at the grids which replaces the external signal source indicated in Figure 9-2. As indicated in the introduction, this voltage will produce a velocity modulation on the beam, whose value we shall now determine.

Let the $z$ axis be taken in the direction of electron flow, with the entrance grid at the position $z = 0$. The grids are assumed to be ideal; that is, all electrons pass through

without interception, and the rf electric field lines are perpendicular to and terminate on the grids.

Although an electron is between the grids, it experiences a force due to the rf electric field. This force causes an acceleration, as in Equation (1.1-1).

$$\frac{d^2z}{dt^2} = -\frac{e}{m}E_z \qquad (9.1\text{-}1)$$

This equation holds for time-varying electric fields as well as for static fields. For the gap in Figure 9.1-1,

$$E_z = -\frac{A}{d}\sin \omega t \qquad (9.1\text{-}2)$$

where $d$ is the grid spacing, and $A \sin \omega t$ is the instantaneous gap voltage. Thus, the motion of an electron is given by the solution to the equation:

$$\frac{d^2z}{dt^2} = \frac{eA}{md}\sin \omega t \qquad (9.1\text{-}3)$$

Integrating once, we obtain

$$\frac{dz}{dt} = u_o - \frac{eA}{\omega md}(\cos \omega t - \cos \omega t_1) \qquad (9.1\text{-}4)$$

where $t_1$ is the time at which the electron passed through the first grid, and $u_o$ is the dc velocity of the electrons entering the gap. The velocity $u_o$ is given by

$$u_o = \sqrt{\frac{2e}{m}V_o} \qquad (9.1\text{-}5)$$

where $V_o$ is the dc voltage of the electron beam, as in Figure 9-2. Equation (9.1-4) gives the velocity of the electron at any instant while it is in the gap. To find the exit velocity, we must substitute the time at which the electron leaves the gap for $t$ in the above equation. Calling this time $t_2$, the exit velocity is given by

$$u(d) = u_o - \frac{eA}{\omega md}(\cos \omega t_2 - \cos \omega t_1) \qquad (9.1\text{-}6)$$

If we assume that the amplitude of the rf voltage $A$ is very small compared with the dc voltage of the beam $V_o$, the electron transit time in the gap is very nearly that given by the dc velocity alone. Thus, if $t_o$ is the instant at which the electron is at the center of the gap,

$$t_1 = t_o - \frac{d}{2u_o} \qquad (9.1\text{-}7)$$

and

$$t_2 = t_o + \frac{d}{2u_o} \qquad (9.1\text{-}8)$$

If these expressions are substituted into Equation (9.1-6) and if we simplify the resulting expression by the use of trigonometric identities, we obtain the following expression for the exit velocity:

$$u(d) = u_o + \frac{2eA}{\omega md} \sin \frac{\omega d}{2u_o} \sin \omega t_o \qquad (9.1-9)$$

If the beam-coupling coefficient $M$ is defined as in Chapter 7,

$$M = \frac{\sin \dfrac{\omega d}{2u_o}}{\dfrac{\omega d}{2u_o}} \qquad (9.1-10)$$

Equation (9.1-9) becomes

$$u(d) = u_o + \frac{eMA}{mu_o} \sin \omega t_o$$

$$= u_o \left( 1 + \frac{MA}{2V_o} \sin \omega t_o \right) \qquad (9.1-11)$$

$M$ is plotted as a function of the gap transit time in Figure 7.1-2. It is unity for zero transit time and drops off for non-zero values of transit time.

### (b)   The Bunching Process

Having discussed the process by which velocity modulation is produced on the beam at the input gap, we next consider the mechanism by which this velocity modulation causes bunching or current modulation to occur in the drift region between the two cavities.

This bunching process has already been described in connection with Figure 9-1, and we shall now seek a quantitative description of the process in order to answer important questions such as: What should the spacing be between the two cavities in order to achieve a maximum degree of bunching? What magnitude of current is induced in the output cavity?

For the moment we shall neglect the mutually repulsive forces of space charge. This approximation is reasonably valid for low-power tubes, where the electron density in the beam is relatively small. We shall further assume that all motion is in the $z$ direction. Physically, this requires either that the space-charge forces be too small to cause transverse spreading or else that the electron motion be confined by a strong dc magnetic field in the $z$ direction.

The electrons emerging from the input cavity have a velocity given by Equation (9.1-11). Since there are no accelerating fields in the drift space between the two cavities, each electron moves at a constant velocity given

by this equation for its particular value of $t_o$. This behavior has been depicted in the Applegate diagram of Figure 9-1.

Assuming a separation $l$ between the centers of the input and output cavity gaps, the time of arrival $t$ of a particular electron at the output cavity is given by the expression:

$$t - t_o = \frac{l}{u_o\left(1 + \dfrac{M}{2}\dfrac{A}{V_o}\sin \omega t_o\right)} \tag{9.1-12}$$

Let us make the simplifying assumption that the input cavity voltage amplitude is much less than the dc beam voltage. This will be true in most cases, except for some very high power tubes. The second term in the
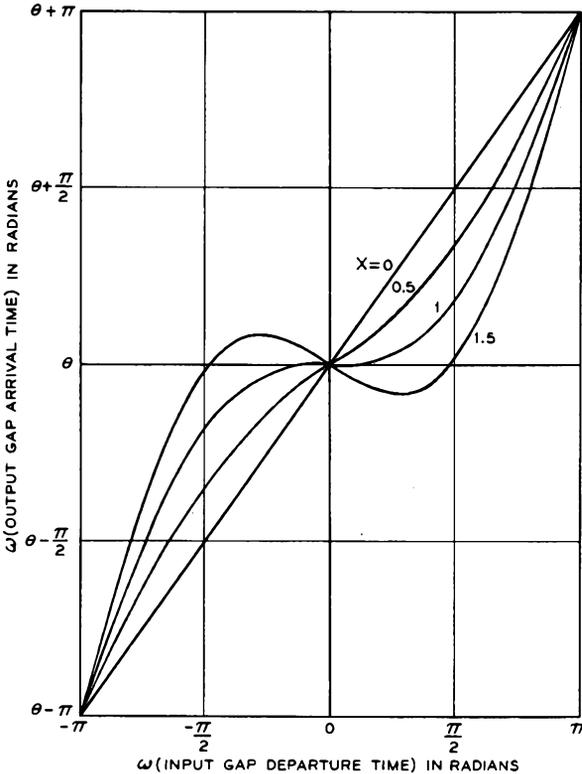


FIG. 9.1-2 Output-gap arrival time plotted vs. the time of departure from the input gap for various values of $X$, the bunching parameter, defined by Equation (9.1-14). $\theta$ is the dc transit angle. For $X$ greater than unity, some electrons leaving the input gap at three different instants arrive at the same instant at the output gap.

denominator is thus much less than unity, and Equation (9.1-12) can be approximately written as

$$t - t_o = \frac{l}{u_o}\left(1 - \frac{M}{2}\frac{A}{V_o}\sin \omega t_o\right) \qquad (9.1\text{-}13)$$

or, in terms of radians,

$$\omega t - \omega t_o = \theta\left(1 - \frac{M}{2}\frac{A}{V_o}\sin \omega t_o\right)$$

$$= \theta - X\sin \omega t_o \qquad (9.1\text{-}14)$$

where $\theta = \omega l/u_o$ is the dc transit angle between cavities, and $X = (M/2)$ $(A/V_o)\theta$ is a parameter known as the bunching parameter.

In Figure 9.1-2 are plotted curves showing output-gap arrival time as a function of input-gap departure time over one rf cycle, for various values of the bunching parameter. One notes that for values of the bunching parameter greater than unity, the departure time is a multivalued function of the arrival time for electrons near the bunch center. However, the arrival time is always a single-valued function of the departure time.

Let us first consider the situation for $X$ less than unity. The instantaneous current reaching the output cavity can be written as

$$i(t) = \frac{dq}{dt} \qquad (9.1\text{-}15)$$

where $dq$ is the amount of charge arriving at the output cavity in a time interval $dt$. In Figure 9.1-2 the ordinate and abscissa are proportional to $t$ and $t_o$, respectively. We see from this figure that the amount of charge arriving in a time $dt$ can be related to the corresponding departure time interval $dt_o$ by

$$dq = -I_o dt_o \qquad (9.1\text{-}16)$$

since electrons leave the input cavity evenly spaced at a rate given by the dc current. The minus sign is used so that $I_o$ may be a positive quantity; $dq$ is of course negative for electrons.

Substituting Equation (9.1-16) into Equation (9.1-15), we obtain

$$i(t) = -I_o\frac{dt_o}{dt} \qquad (9.1\text{-}17)$$

where the derivative is obtained simply by measuring slopes on the curve of Figure 9.1-2. Current waveforms for several values of the bunching parameter are shown in Figure 9.1-3. Infinite current peaks are obtained at the arrival times for which the curves of Figure 9.1-2 have zero slope.

For values of the bunching parameter greater than unity, the fact that the curve of Figure 9.1-2 is multivalued results in three values of slope for a
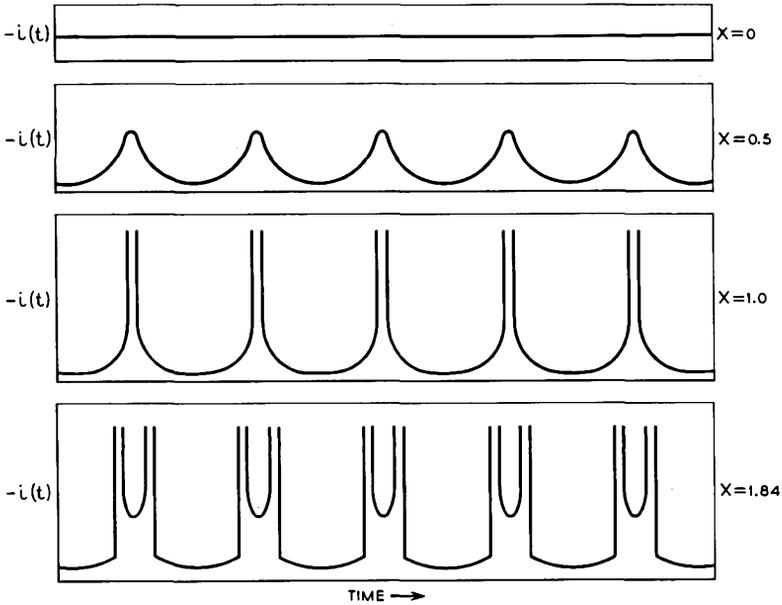
FIG. 9.1-3 Beam current waveforms at the output gap. For $X$ greater than or equal to unity, infinite peaks are obtained at the points corresponding to zero slope in Figure 9.1-2. In an actual tube, these peaks though large would remain finite because of the space-charge forces.

given arrival time near the bunch center. This situation is easily handled as follows. Since Equation (9.1-16) must include the total charge for a given arrival time, we must include a term for each of the three departure times; thus

$$dq = -I_o \left[ dt_o|_1 + dt_o|_2 + dt_o|_3 \right] \qquad (9.1\text{-}18)$$

Corresponding to Equation (9.1-17), we obtain

$$i(t) = -I_o \left[ \left| \frac{dt_o}{dt} \right|_1 + \left| \frac{dt_o}{dt} \right|_2 + \left| \frac{dt_o}{dt} \right|_3 \right] \qquad (9.1\text{-}19)$$

In each case, the absolute value of the derivative must be taken. Physically this corresponds to the fact that the charge increment $dq$ has the same sign regardless of the sequence of arrival of the electrons. A negative value of $dt_o/dt$ merely indicates that electrons which left the input cavity last arrive at the output cavity first; $dq$ always has a negative value.

The current waveforms of Figure 9.1-3 may be Fourier analyzed to determine the fundamental component and the various harmonics. This

could be done graphically. However, it is possible to solve this problem analytically. We shall proceed with such an analysis.

The current at the output gap can be written as the Fourier series:

$$i(t) = -I_o + \sum_{n=1}^{\infty} [a_n \cos n(\omega t - \theta) + b_n \sin n(\omega t - \theta)] \quad (9.1\text{-}20)$$

where

$$a_n = \frac{1}{\pi} \int_{\theta - \pi}^{\theta + \pi} i(t) \cos n(\omega t - \theta) d(\omega t)$$

and

$$b_n = \frac{1}{\pi} \int_{\theta - \pi}^{\theta + \pi} i(t) \sin n(\omega t - \theta) d(\omega t)$$

Let us consider first the situation for $X$ less than unity, so that the curves of Figure 9.1-2 are single-valued.

It will be convenient to change the variable of integration from arrival time to departure time. Equation (9.1-17) gives us

$$i(t)d(\omega t) = -I_o d(\omega t_o) \quad (9.1\text{-}21)$$

From Figure 9.1-2, we see that the limits of integration become $-\pi$ to $+\pi$. When Equations (9.1-21) and (9.1-14) are substituted into the above integrals, we obtain

$$a_n = -\frac{I_o}{\pi} \int_{-\pi}^{\pi} \cos n(\omega t_o - X \sin \omega t_o) d(\omega t_o)$$

and

$$b_n = -\frac{I_o}{\pi} \int_{-\pi}^{\pi} \sin n(\omega t_o - X \sin \omega t_o) d(\omega t_o) \quad (9.1\text{-}22)$$

$b_n$ is identically equal to zero since the integrand is an odd function of $\omega t_o$. The definite integral in the expression for $a_n$ is given by a Bessel function:[1]

$$2J_n(nX) = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos n(\omega t_o - X \sin \omega t_o) d(\omega t_o) \quad (9.1\text{-}23)$$

Equation (9.1-20) thus becomes

$$i(t) = -I_o - 2I_o \sum_{n=1}^{\infty} J_n(nX) \cos n(\omega t - \theta) \quad (9.1\text{-}24)$$

For values of $X$ greater than unity, the same expression is obtained. This is shown in Appendix XV. For small values of $X$, $J_1(X) \approx X/2$ and $J_n(nX)$

---

[1]Reference 9.1, p. 150.

is very small for $n > 1$. Equation (9.1-24) then becomes $i(t) = -I_o$ $[1 + X \cos (\omega t - \theta)]$ for small $X$ and hence for small input signals.

Equation (9.1-24) shows that the various harmonics in the bunched beam have amplitudes proportional to Bessel functions of order $n$, where $n$ is the
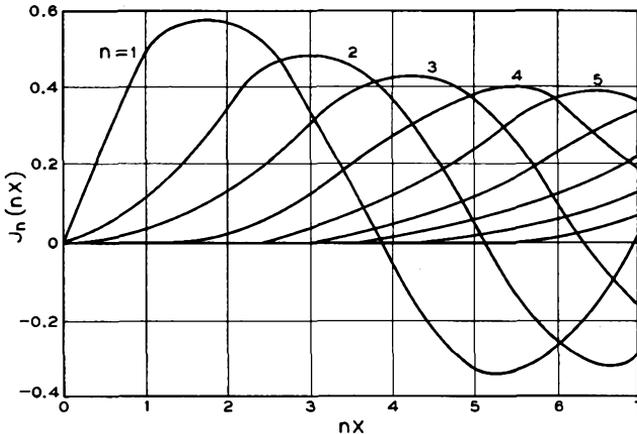


FIG. 9.1-4 Bessel functions of various orders. The maximum value of $J_1$ occurs at $X = 1.84$ and is equal to 0.582.

same as the harmonic. These Bessel functions are plotted in Figure 9.1-4. Since the abscissa in this figure is proportional to the transit time between cavities, we can adjust either the beam velocity or the distance between cavities so as to obtain a maximum amplitude for any of the harmonics. For an amplifier, we would make $X$ equal to 1.84 so as to peak the fundamental component at the output cavity. On the other hand, it is also possible to use the tube of Figure 9-2 as a harmonic generator, in which case we would choose the transit time to correspond to the peak of one of the higher-order Bessel functions of Figure 9.1-4. Since these other peaks are nearly as large as that of the fundamental, the two-cavity klystron can be a very efficient harmonic generator. Of course, the output cavity would be tuned to the harmonic frequency.

Since the electrons become bunched about an electron which passed through the input cavity when the voltage across the input cavity was changing from decelerating the electrons to accelerating them, the center of the electron bunch arrives at the output cavity delayed by the dc transit angle, but advanced by $\pi/2$. This can also be seen by comparison of the phase of the voltage applied to the input cavity with the phase of the fundamental component in Equation (9.1-24).

(c)  *Current Induced in the Output Cavity by the Bunched Beam*

To complete the description of klystron interaction, we must consider the current induced in the output cavity by the bunched beam. This problem has already been considered in Section 7.1(a), where we have considered the current induced by a modulated beam into a load connected between two grids. Although the discussion presented in Chapter 7 was applied to a tetrode, it might equally well be applied to a klystron, where we interpret the pair of grids to be the grids of a re-entrant cavity resonator.

We shall find it convenient to adopt the conventions for positive gap voltage and induced current indicated in Figure 9.1-5. The direction for positive induced current is opposite to that used in Chapters 6 and 7. In microwave tube work it is customary to assume that the ac component of beam current is positive when directed from left to right. Thus, the induced current indicated in Figure 9.1-5 is positive when the ac component of beam current is positive.

FIG. 9.1-5 Convention for positive induced current adopted in this chapter. The induced current is equal to $M$ times the beam current modulation at the center of the gap (in magnitude and phase).

From Equation (9.1-24), the dc and fundamental components of beam current at the output cavity are given by

$$i(l,t) = -I_o + i_1 \cos \omega\left(t - \frac{l}{u_o}\right) \qquad (9.1\text{-}25)$$

where

$$i_1 = -2I_o J_1(X) \qquad (9.1\text{-}26)$$

and we have substituted $\theta = \omega l/u_o$. But this is exactly the same type of wavelike behavior that was assumed for the beam current density in Section 7.1(a). From Section 7.1(a), therefore, we have the result that the induced current (with positive direction assumed as in Figure 9.1-5) is given by $M_2 i_1 \cos \omega(t - l/u_o)$, or in phasor notation we may write:

$$I = M_2 i_1 \epsilon^{-j(\omega/u_o)l} \qquad (9.1\text{-}27)$$

where $I$ and $i_1 \epsilon^{-j(\omega/u_o)l}$ are phasor quantities representing the induced current and the fundamental component of beam current at the gap center,

respectively. $M_2$ is the beam-coupling coefficient for the output cavity. Simply stated, the induced current is $M_2$ times the fundamental component of the beam current at the gap center, both in magnitude and in phase.

This is really a very important result; the induced current is independent of the loading of the output cavity. Thus, the modulated beam truly acts
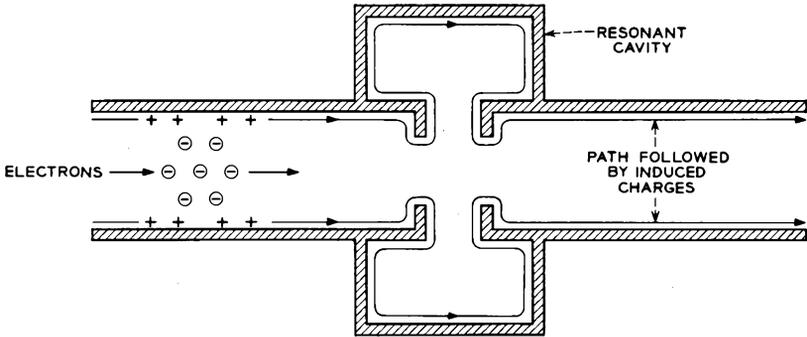
FIG. 9.1-6 Path followed by the induced charges associated with an electron bunch passing through a cavity.

as a current source. Figure 9.1-6 shows the path followed by the induced charges associated with a single bunch of electrons which passes through the output cavity.

At resonance the phase of the oscillations in the output cavity is such that maximum decelerating voltage appears across the output cavity when maximum number of electrons is crossing the gap. This follows from the fact that at resonance the cavity and load appear as a resistance connected between the grids, so that when the induced current reaches a maximum, the voltage across the cavity is also a maximum. This explains the transfer of energy from the beam to the cavity.

How much power can be delivered to the output cavity and load from the bunched beam? This question can be answered most easily from a consideration of the equivalent circuit of the output cavity, shown in Figure 9.1-7.

In this equivalent circuit, $L$ and $C$ are the inductance and capacitance of the re-entrant cavity itself, without the presence of the beam. These parameters are obtained as discussed in the introduction to Chapter 8. Similarly $G_c$ is a conductance which accounts for the resistive and dielectric losses in the cavity. These parameters determine the *cold* unloaded $Q$ of the cavity, defined by

$$Q_o = \frac{\omega_c C}{G_c} \tag{9.1-28}$$
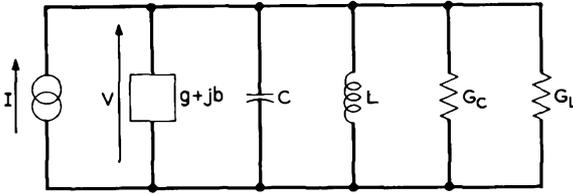
FIG. 9.1-7 Equivalent circuit for the output cavity and its load. The symbols used are defined as follows:

$$V = \text{gap voltage}$$
$$I = \text{induced current}$$
$$C = \text{gap capacitance}$$
$$L = \text{cavity inductance}$$
$$g + jb = \text{beam loading admittance}$$
$$G_c = \text{conductance representing cavity losses}$$
$$G_L = \text{load conductance (as seen from the cavity gap)}$$

where $\omega_c$ is the cavity resonant frequency in radians.

The beam-loading admittance is given by $g + jb$. The parameters $g$ and $b$ are functions of the gap transit time. For an unmodulated beam, they are given by Equations (7.1-14) and plotted in Figure 7.1-4. Physically, these parameters express the fact that for a non-zero gap transit time, some energy is exchanged between the electrons and the rf energy stored in the output cavity. It is convenient to assume the same parameters for a modulated beam. This is a small-signal approximation, useful in predicting the small-signal performance of a klystron amplifier.

In most practical tubes, the output gap transit angle is much less than $\pi$, and the susceptive component is capacitive, as shown in Figure 7.1-4. Furthermore, when the transit angle is much less than $\pi$, $b$ varies linearly with frequency, and we may write

$$b = \omega C_b \qquad\qquad (9.1\text{-}29)$$

where $C_b = g_o T_o/12$. Here $g_o$ and $T_o$ are defined as in Equation (7.1-14). The *hot* unloaded $Q$ is defined by

$$Q_u = \frac{\omega_o(C + C_b)}{G_c + g} \qquad\qquad (9.1\text{-}30)$$

where $\omega_o$ is the resonant frequency of the cavity with the beam present. $Q_u$ thus includes the intrinsic cavity parameters plus the loading due to the electron beam. It should be noted that for the small-signal approximation we are considering, $Q_u$ is independent of whether or not the beam is velocity modulated in the input cavity. Thus it could be evaluated by measuring

the characteristics of the output cavity with the beam turned on, but with no rf drive on the input cavity.[2]

In Figure 9.1-7 the conductance $G_L$ accounts for the load connected to the cavity. Since the actual load will be separated from the cavity by waveguide, transformers, and various other components, $G_L$ is an effective value as seen at the cavity gap. It is defined so as to give a correct value for the power absorbed for any given gap voltage. The external $Q$ is defined by

$$Q_e = \frac{\omega_o(C + C_b)}{G_L} \qquad (9.1\text{-}31)$$

The loaded $Q$ is the $Q$ of the whole circuit of Figure 9.1-7. It is thus given by

$$Q_l = \frac{\omega_o(C + C_b)}{G_c + g + G_L} \qquad (9.1\text{-}32)$$

From the last three equations, we note that

$$\frac{1}{Q_l} = \frac{1}{Q_u} + \frac{1}{Q_e} \qquad (9.1\text{-}33)$$

The remaining parameters in Figure 9.1-7 are the gap voltage $V$ and the induced current $I$, defined previously.

To deliver maximum power to the load, we adjust the parameters such that

$$G_L = G_c + g \qquad (9.1\text{-}34)$$

and

$$\omega^2 L(C + C_b) = 1 \qquad (9.1\text{-}35)$$

Equation (9.1-34) may also be stated as

$$Q_e = Q_u \qquad (9.1\text{-}36)$$
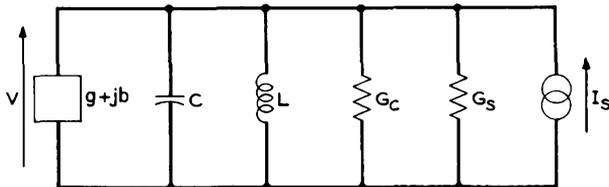


FIG. 9.1-8 Equivalent circuit for the input cavity and the signal source. The symbols used are as defined in Figure 9.1-7, with the additions:

$I_s$ = current source representing the rf input signal

$G_s$ = source conductance (as seen from the cavity gap)

---

[2]Measurements of resonant cavity characteristics are discussed in Reference 9.2.

The load conductance may be adjusted by changing the coupling of the cavity to the load. The resonance condition, Equation (9.1-35), may be obtained by varying the cavity capacitance or inductance. One common way to do this is to deform mechanically the cavity to change the gap spacing, thus changing the cavity capacitance.

The equivalent circuit for the input cavity is shown in Figure 9.1-8. Here the current source at the right represents the rf input source. The other parameters have the same meaning as in Figure 9.1-7, except that $G_L$ is replaced by $G_s$, the source conductance.

## 9.2  Reflex Klystrons

The klystron amplifier of Figure 9-2 is an extremely stable type of microwave tube, where stability refers to freedom from oscillations. In normal operation there is no feedback from the output cavity to the input cavity, except perhaps by secondary electrons produced at or beyond the output cavity which make their way back to the input cavity. In order to make this tube oscillate, it is necessary to provide an external feedback path from the output cavity to the input cavity. This can be accomplished by tapping off a portion of the output power and feeding it back into the input cavity by means of an external transmission line. The oscillation frequency and power output are then determined by the simultaneous requirement that the loop gain be unity and the loop phase shift be a multiple of $2\pi$ radians. The loop phase shift can be varied by changing either the beam voltage or the length of the feedback cable. Of course, if the frequency is varied any appreciable amount, the cavities must be retuned.

For many applications of microwave oscillators, it is necessary to change the frequency rapidly. This is most readily done if the oscillation frequency can be varied electronically. In the oscillator described above, this can be done over a limited frequency range by changing the beam voltage. However, varying the beam voltage simultaneously varies the beam power, and this results in a larger change in output power with frequency than is desirable for most applications. This drawback is eliminated in the reflex klystron described below.

A schematic drawing of the reflex klystron, together with the power supply connections, is shown in Figure 9.2-1(a). A potential profile along the electron beam is shown in Figure 9.2-1(b). The electron gun injects the electron beam through the grids of a re-entrant microwave cavity. The electrons then approach an electrode known as the repeller, which is at a lower potential than the cathode. The repelling electric field in this region causes the electrons to "turn around" and pass once again through the cavity grids, but in the opposite direction. The electrons are then collected

on the walls of the cavity or other grounded metal parts of the tube. Since this tube is a relatively low-power device, the power dissipated in the cavity by the incident electrons is not excessive. The broken lines in Figure 9.2-1
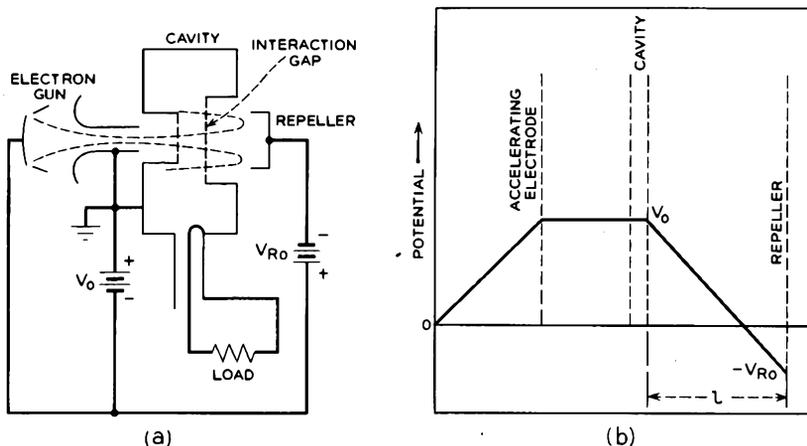


FIG. 9.2-1 Reflex klystron and its potential profile. (a) Schematic drawing of the tube showing the power supply connections. (b) Potential profile along the beam from the cathode to the repeller.

indicate the trajectories of the outermost electrons in the beam. The reflex klystron is usually operated without any magnetic field to confine the beam, and consequently the beam spreads under the influence of its own space charge. However, since the total distance traveled by the electrons is small, the total radial spreading is not excessive.

If we postulate an rf voltage on the cavity grids, the beam is velocity modulated during its first passage through the grids. While the electrons are in the repeller region, the velocity modulation is converted into current modulation, much like in the drift space between cavities of a klystron amplifier. When the current modulated beam re-enters the cavity, it induces an rf current in the walls of the cavity, as in the output cavity of the amplifier. This induced current is then the source of output power. The gap voltage is also produced by this induced current. Hence, the cavity serves a dual purpose; it is both the input cavity and the output cavity with feedback intrinsically provided.

The bunching mechanism in the repeller region can best be described by means of the Applegate diagram given in Figure 9.2-2(a). The corresponding gap voltage as a function of time is shown in Figure 9.2-2(b). Time markers are included on the abscissa in Figure 9.2-2(a), marking the
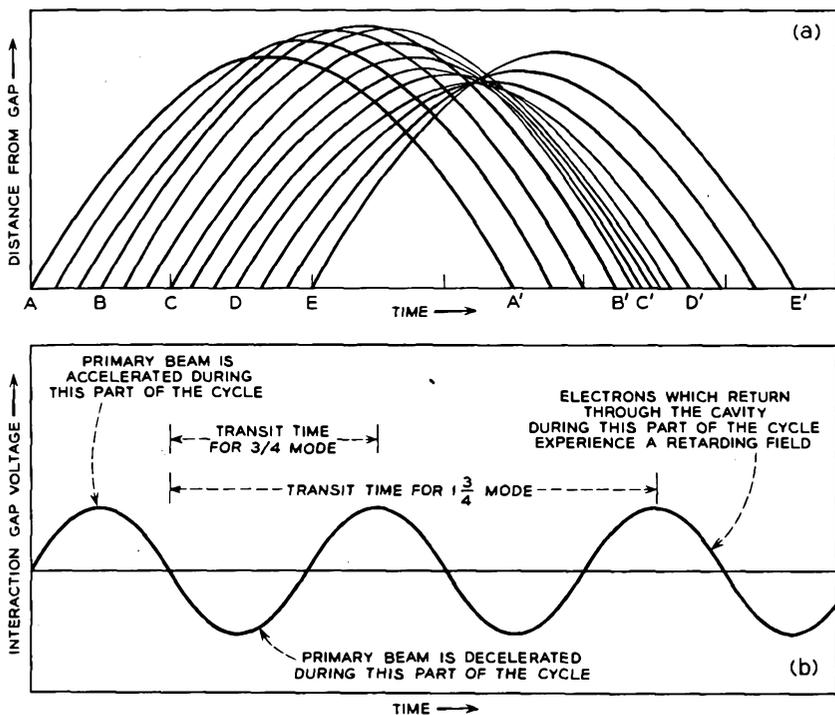
FIG. 9.2-2 Electron trajectories in a reflex klystron showing the dependence on the instantaneous gap voltage at the moment of departure. (a) Applegate diagram for the $1\frac{3}{4}$ mode of oscillation. (b) Instantaneous gap voltage.

instants for which the gap voltage is zero. Figure 9.2-2(a) shows distance-time plots for several electrons passing through the cavity at equally spaced intervals during one cycle of the cavity voltage. The electrons start out toward the repeller, are "turned around," and arrive back at the gap at a later time. Each electron leaves the gap with a different velocity, depending on the instantaneous value of the gap voltage as it passes through. Electron $B$ passes through when the voltage is a maximum, and it receives the greatest increment of velocity. It therefore penetrates farthest toward the repeller before being turned back. On the other hand, electron $D$ passes through when the gap voltage is maximally retarding, thus penetrating the least distance and arriving back in the shortest time span. Electrons $A$, $C$, and $E$ go through the gap when the voltage is zero and penetrate toward the repeller an intermediate distance.

We note that a bunch forms around electron $C$ which initially passed through the gap when the voltage was zero and decreasing. Since the bunch

arrives back at the gap at such an instant that the voltage is maximally retarding, energy is transferred from the beam to the cavity. In the diagram shown, eight of the electrons are slowed down in their second transit through the cavity, whereas only four are speeded up.

Since the dc transit time in the repeller region corresponds to $1\frac{3}{4}$ rf cycles for the condition shown in Figure 9.2-2, this mode of oscillation is known as the $1\frac{3}{4}$ mode. Alternatively, a higher magnitude of reflector voltage would cause the electrons all to return to the gap in a shorter span of time, and oscillation in the $\frac{3}{4}$ mode is possible. This occurs when the dc transit time corresponds to $\frac{3}{4}$ of an rf cycle. It is not necessary to have the transit time exactly $\frac{3}{4}$ or $1\frac{3}{4}$ cycles; oscillations are possible for slight departures from these values, but at reduced power levels and at slightly different frequencies. It is only necessary that the bunch be so phased that sufficient energy is transferred from the modulated beam to the cavity to make up for losses in the cavity and load. It is apparent that oscillations are possible for transit times corresponding to $n + \frac{3}{4}$ cycles, where $n$ is an integer.

With this much of an understanding of the qualitative behavior of the device, we proceed to take a more quantitative look.

For simplicity, the effect of space-charge forces on the electron motion will again be neglected. Because of the relatively low beam current of the reflex klystron, this is a good approximation everywhere except near the repeller where the electrons are reversing their direction of motion.

The tube has dc voltages applied as in Figure 9.2-1, with $l$ being the spacing from gap to repeller. The polarities of gap voltage and induced current are taken as in Figure 9.1-5. Assume that there exists an rf voltage across the gap given by

$$v(t) = A \sin \omega t \qquad (9.2\text{-}1)$$

The exit velocity of an electron is given by Equation (9.1-11),

$$u(d) = u_o\left(1 + \frac{M}{2}\frac{A}{V_o}\sin \omega t_o\right) \qquad (9.2\text{-}2)$$

where $t_o$ is the time at which the electron passes through the center of the gap.

We shall assume that there is a uniform electric field between the repeller and the cavity so that the electrons experience a uniform force directed toward the cavity while traveling in this region. The resulting acceleration of the electrons is given by the differential equation:

$$\frac{d^2z}{dt^2} = -\frac{e}{m}\frac{V_o + V_{Ro}}{l} \qquad (9.2\text{-}3)$$

This equation can be integrated once, obtaining

$$\frac{dz}{dt} = u(d) - \frac{e}{m} \frac{V_o + V_{Ro}}{l}(t - t_o)$$ (9.2-4)

where $t_o$ is the time at which the electron first passed the mid-point of the cavity. Strictly speaking, the exit time from the gap rather than $t_o$ should be used in Equation (9.2-4).

Let us now find an expression for the time $t$ at which the electron returns to the gap. This value is obtained by solving Equation (9.2-4), with

$$\frac{dz}{dt} = -u(d)$$ (9.2-5)

since the electron neither gains nor loses net energy in the repeller region. We obtain

$$t - t_o = \frac{2mlu_o}{e(V_o + V_{Ro})}\left[1 + \frac{M}{2} \frac{A}{V_o} \sin \omega t_o\right]$$ (9.2-6)

where use has been made of Equation (9.2-2). Multiplying through by the radian frequency, we obtain

$$\omega t - \omega t_o = \theta\left[1 + \frac{M}{2} \frac{A}{V_o} \sin \omega t_o\right]$$ (9.2-7)

where

$$\theta = \frac{2m\omega l u_o}{e(V_o + V_{Ro})}$$ (9.2-8)

is the dc transit angle in the repeller region.

To find an expression for the instantaneous current on the beam as it re-enters the cavity, we proceed in the same manner as in Section 9.1, except that there is a different relationship between the departure and arrival times of an electron, that is, Equation (9.1-14) is replaced by (9.2-7). Comparing these two equations, we note that they are identical except for the algebraic sign of the term containing the bunching parameter $X$, where

$$X = \frac{M}{2} \frac{A}{V_o}\theta$$ (9.2-9)

The result for the reflex klystron therefore is given by Equation (9.1-24), except that $X$ must be replaced by $-X$. We thus have the following result for the beam current injected into the cavity gap from the repeller region:

$$i(t) = -I_o - 2I_o \sum_{n=1}^{\infty} (-1)^n J_n(nX) \cos n(\omega t - \theta)$$ (9.2-10)

where use has been made of the identity[3]

---

[3]Reference 9.1, p. 128.

$$J_n(-x) = (-1)^n J_n(x) \qquad (9.2\text{-}11)$$

The fundamental component of beam current is

$$i_1(t) = 2I_o J_1(X) \cos (\omega t - \theta) \qquad (9.2\text{-}12)$$

In the notation of Figure 9.1-5, the electrons travel from right to left through the grids, so that the induced current phasor is given by

$$I = -2M I_o J_1(X)\epsilon^{-j\theta} \qquad (9.2\text{-}13)$$

which is $-M$ times the phasor representing the current of Equation (9.2-12).

The equivalent circuit of the cavity is given as in the klystron amplifier by Figure 9.1-7, with one change. The beam-loading admittance is denoted by $g' + jb'$ instead of $g + jb$, since in a reflex klystron the results of Section 7.1(b) are not applicable. This is true for two reasons. First, the beam traverses the cavity twice and on its second transit is highly bunched. Second, the secondary electrons due to the beam impact contribute significantly to beam loading.

The induced current $I$ acts as a current source producing the gap voltage $V$. This voltage in turn is the cause of the original velocity modulation on the electron stream.

When the tube is oscillating in the steady state, we may write

$$-\frac{I}{V} + G + jB = 0 \qquad (9.2\text{-}14)$$

where

$$G = g' + G_e + G_L$$

and

$$B = b' + \omega C - \frac{1}{\omega L}$$

The amplitude and frequency of oscillation are determined by the condition that this equation be satisfied for both the real and imaginary parts. Let us define an electronic admittance by

$$Y_e = -\frac{I}{V} \qquad (9.2\text{-}15)$$

Using Equations (9.2-1), (9.2-9) and (9.2-13), we have the following formula for the electronic admittance:

$$Y_e = \frac{2M I_o J_1(X)\epsilon^{-j\theta}}{A\,\epsilon^{-j\frac{\pi}{2}}}$$

$$= \frac{I_o M^2 \theta}{2V_o} \frac{2J_1(X)}{X} \epsilon^{j(\frac{\pi}{2} - \theta)} \qquad (9.2\text{-}16)$$

At oscillation, this admittance appears as a negative conductance shunted by some susceptance.

The electronic admittance is nonlinear since it is proportional to the factor

$$\frac{2J_1(X)}{X}$$

and $X$, the bunching parameter, is proportional to the rf gap voltage. This factor of proportionality is shown in Figure 9.2-3. As the magnitude of the
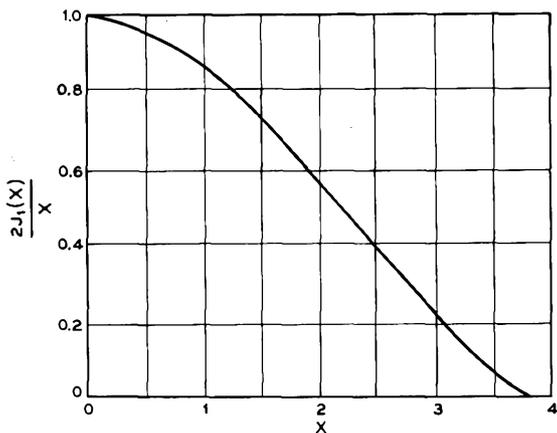


FIG. 9.2-3 Factor by which the electronic admittance varies with signal level. As the signal level goes to zero, this factor approaches unity.

rf gap voltage goes to zero, this factor approaches unity, and we obtain the small-signal value of the electronic admittance,

$$Y_{es} = \frac{I_o M^2 \theta}{2V_o} \epsilon^{j(\frac{\pi}{2}-\theta)} \qquad (9.2\text{-}17)$$

When this admittance is plotted on a rectangular admittance chart as a function of $\theta$, the dc transit angle, one obtains the admittance spiral shown in Figure 9.2-4. The admittance starts at the origin for zero transit angle and then spirals outward and clockwise as the transit angle is increased. Oscillations are possible for values of the transit angle which produce a negative conductance which is greater in magnitude than the positive conductance represented by the load and losses of the cavity. Maximum magnitudes of negative conductance are obtained for values of $\theta$ given approximately by $n + \frac{3}{4}$ cycles, where $n = 0,1,2$ etc. The electronic admittance spiral corresponding to any level of oscillation, that is, for a value of $X$
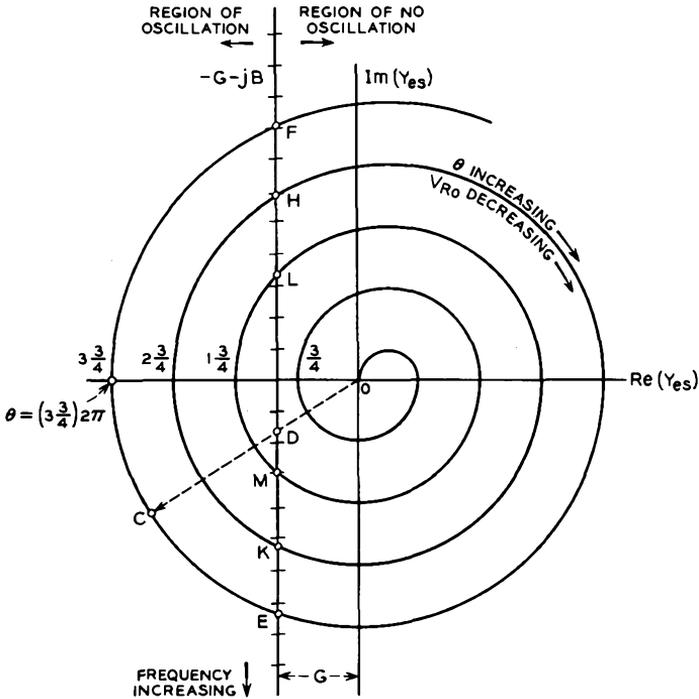
FIG. 9.2-4 Locus of small-signal electronic admittance $Y_{es}$ on a rectangular impedance plot. The locus spirals out from the origin as the dc transit angle is increased from zero. Oscillations are obtained for negative conductances greater in magnitude than the total conductance loading the cavity, that is, for points to the left of the line $-G - jB$.

greater than zero, may be obtained from the small-signal spiral by shrinking the spiral radially by the factor $2J_1(X)/X$ obtained from Figure 9.2-3.

Equations (9.2-14) and (9.2-15) may be written as

$$Y_e = -(G + jB) \qquad (9.2\text{-}18)$$

We now look for a graphical method of solving this equation so that oscillation power and frequency can be predicted as a function of the dc transit angle $\theta$. For a fixed cavity geometry, the total shunt conductance $G$ may be assumed constant over the frequency bandwidth of the cavity. The total shunt susceptance may be expanded as

$$B = \omega C_T - 1/\omega L = C_T\left(\omega - \frac{\omega_o{}^2}{\omega}\right)$$

$$= C_T\frac{(\omega + \omega_o)(\omega - \omega_o)}{\omega} \cong 2C_T\Delta\omega \cong 2GQ_l\frac{\Delta\omega}{\omega_o} \qquad (9.2\text{-}19)$$

where $C_T$ is the gap capacitance plus a capacitance to account for beam loading, and $Q_l$ is the loaded $Q$ with the beam present, defined in Equation (9.1-32). $\omega_o$ is the radian resonant frequency with the beam present. We may plot

$$-(G + jB) = -G\left[1 + j2Q_l\frac{\Delta\omega}{\omega_o}\right] \qquad (9.2\text{-}20)$$

as a function of frequency in Figure 9.2-4. The plot consists of a vertical line $G$ units to the left of the origin. Equal increments of distance along the line correspond to equal increments of cycles off resonance. As the frequency is increased, the corresponding value of cavity admittance is found at a lower point on this line. By Equation (9.2-20) the frequency range between any two values of cavity admittance is inversely proportional to $Q_l$.

By Equation (9.2-18) an operating point in Figure 9.2-4 is found as the intersection of the cavity admittance locus $EF$ and the electronic admittance spiral for the particular value of gap voltage. Thus in Figure 9.2-4 we note that small-signal oscillations ($X = 0$) are possible at the points $E$, $F$, $H$, $K$, $L$, and $M$, and the corresponding values of frequency and $\theta$ may be read off. From the values of $\theta$, values of the repeller voltage are determined from Equation (9.2-8), given the beam voltage $V_o$.

How does one determine the operating conditions for higher levels of oscillation? Assume that the tube is oscillating at point $E$ on the chart, and the repeller voltage is decreased so as to increase $\theta$ to correspond to the line $OC$. The small-signal admittance spiral has a larger magnitude of negative conductance than the positive value of cavity conductance. Hence oscillations will build up in amplitude and the admittance spiral will shrink until the conductances are equal in magnitude, that is, until point $C$ recedes to point $D$ on the shrunken spiral. An evaluation of the ratio of $OD$ to $OC$ gives the value of $2J_1(X)/X$, and hence the oscillation amplitude can be determined.

Oscillations of varying amplitude and frequency are produced continuously as one decreases the repeller voltage so as to move from point $E$ along the line $EF$ to point $F$. This whole range of operation is known as the $3\frac{3}{4}$ mode, since the center occurs at $\theta = 3\frac{3}{4}$ cycles. Similarly, the $2\frac{3}{4}$ mode of oscillation is produced over the range of transit angles needed to vary the electronic admittance from point $K$ to point $H$. The $1\frac{3}{4}$ mode exists for transit angles needed to vary the electronic admittance from point $M$ to point $L$. However, oscillations in the $\frac{3}{4}$ mode are not possible since the cavity conductance is too large; that is, the cavity is loaded too heavily to permit oscillation in this mode.

A physical description of the buildup of oscillations may also be presented with reference to Figure 9.2-4. Suppose the electron beam is suddenly

turned on with a set of electrode voltages corresponding to the line *ODC*. At this instant there exists an appreciable range of frequencies for which the net conductance is negative and also for which phase conditions are appropriate for positive feedback (note that this is *not* the frequency range corresponding to the line segment *ML*, since the latter range holds only for sinusoidal steady-state signals). Noise within this frequency range is amplified in a regenerative manner and builds up in amplitude. As the noise builds up, the negative conductance decreases in magnitude and the frequency range for regenerative amplification decreases. Finally, a stable operating· condition is obtained for which the amplification bandwidth is sufficiently narrowed so that the product of the input noise power and the amplifier gain is equal to the output signal power. This bandwidth is so narrow that it corresponds practically to a single frequency. This description of oscillator operation is useful in analyzing such quantities as signal buildup-time and oscillation line width. It applies not only to the reflex klystron but also to all other types of sinusoidal oscillators.

By means of the graphical procedure described above one can plot curves
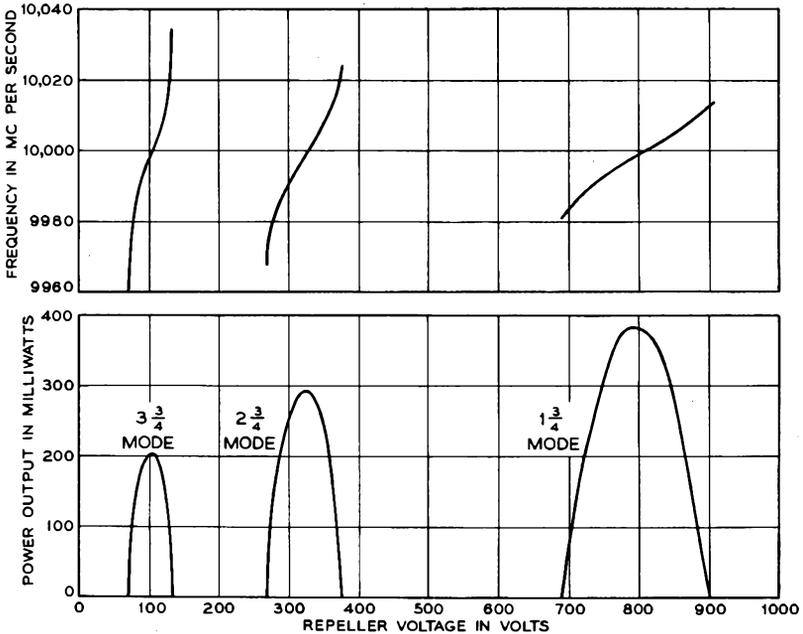


FIG. 9.2-5 Theoretical output characteristics of a reflex klystron. These curves were computed directly from Figures 9.2-3 and 9.2-4, with absolute values chosen so as to give results typical of commercially available tubes.

of power output and frequency vs. repeller voltage for any tube. Such curves are presented in Figure 9.2-5 where parameters have been chosen so as to obtain typical operating values. Three modes are obtained. For this example, modes of order higher than the $3\frac{3}{4}$ mode are not possible, since further reduction of the repeller voltage does not increase the transit time sufficiently to obtain the next higher mode. It should be noted that lower power output is obtained as the mode number is increased. From Equation (9.2-9) the gap voltage magnitude $A$ is proportional to $X/\theta$. Both $X$ and $\theta$ increase with increasing mode number, but $\theta$ increases faster, so that the ratio $X/\theta$ decreases.

The reflex klystron has been the most common microwave tube for many years. With the cavity tuned to a nominal frequency, electronic tuning over a considerable bandwidth is possible merely by varying the repeller voltage. Since the repeller draws no beam current and since the capacitance from the repeller to ground may be made very small, the output frequency can be modulated by the simplest of electronic circuitry.

The reflex klystron is frequently used as a local oscillator in a microwave receiver; the electronic tuning obtainable is ideal for automatic frequency control. It is also commonly used as a laboratory signal source and as an FM transmitter frequency–deviator tube.

The plots of power output vs. repeller voltage presented in Figure 9.2-5 are typical. As the mode number is increased, wider tuning bandwidths are obtained, but the power output is decreased. Since wide tunability is usually the most desirable feature, most reflex klystrons are designed to operate in high-order modes, typically the $3\frac{3}{4}$, $4\frac{3}{4}$, and $5\frac{3}{4}$ modes. In order to obtain large bandwidth, $Q_l$ is made quite low. However, the large value of $G$ does not permit oscillations in the lower modes.

Different center frequencies may be obtained by tuning the cavity to different resonant frequencies and adjusting the repeller voltage. Cavity tuning is often accomplished by mechanical deformation of the cavity size using a bellows type of construction. The output power can be controlled by changing the beam voltage and current. The beam current can be controlled independently of the beam voltage if a control grid is included in the electron gun. This permits amplitude modulation of the output power.

An example of a reflex klystron is the WE-449A, shown in cross section in Figure 9.2-6. This tube uses an external type of cavity in which the inductive portion is largely outside of the vacuum envelope of the tube. A ceramic window separates the internal and external portions of the cavity. This type of construction greatly simplifies the mechanical tuning adjustment since the tuning adjustment can be made outside of the vacuum. In the 449A, this adjustment is made by the use of a plunger in the external cavity. The tube envelope is constructed entirely of metal and ceramic.
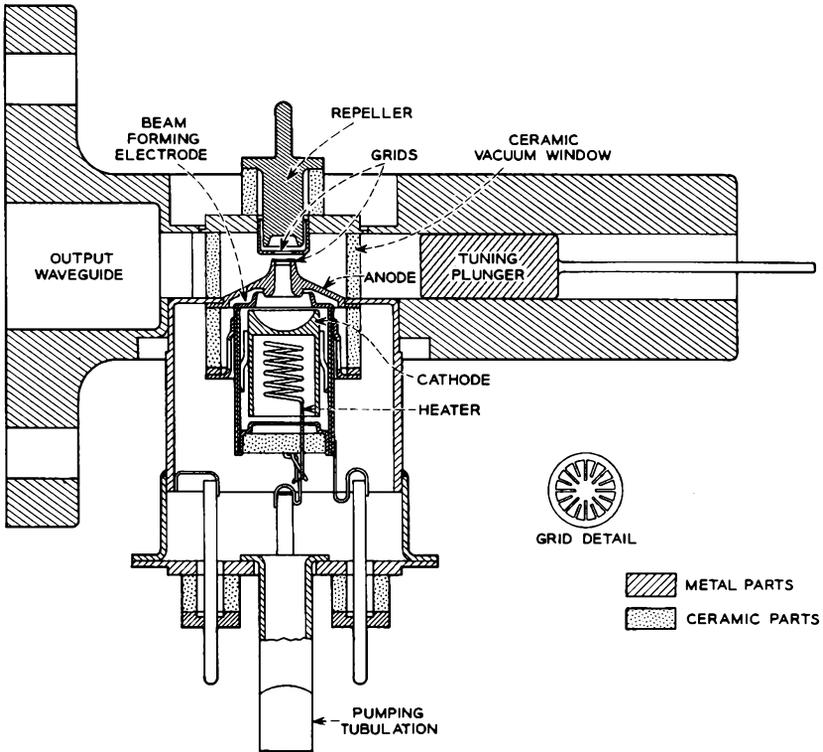
FIG. 9.2-6 WE-449A reflex klystron. This metal-ceramic tube has an overall height of approximately $6\frac{1}{2}$ cm and has the operating characteristics given in Table 9.2-1.

The electron gun is of the type shown in Figure 4.5-1(b), which generates a relatively high perveance beam of approximately 1 mm diameter. No magnetic field is used to prevent space-charge spreading of the beam, since the beam travels a relatively short distance.

The copper grids of the cavity have a geometry resembling the spokes of a

TABLE 9.2-1. WE-449A OPERATING CHARACTERISTICS

Frequency range, Mc............................5925–6425
Beam voltage, volts.............................450
Beam current, ma..............................48
Perveance, amp/volt$^{3/2}$.......................5.03 × 10$^{-6}$
Repeller voltage, volts.........................75 to 125
Mode........................................$2\frac{3}{4}$
Power output, mw..............................125
Modulation sensitivity, Mc/volt ................1.5

wheel in order to minimize electron interception. The gap spacing is 0.58 mm; the gap has a capacitance of 0.3 pf. The repeller electrode is cup-shaped in order to refocus the spreading beam so that it will pass through the grids on its return. The cavity has a hot unloaded $Q_u$ of 135 and a hot loaded $Q_l$ of 80.

The operating characteristics of the 449A are given in Table 9.2-1. The modulation sensitivity is defined as the ratio: (change in frequency of the output produced by a change in repeller voltage)/(change in repeller voltage). Greater than 40 mw of power is obtained over an electronically tunable bandwidth of 65 Mc.

### 9.3  Space-Charge Waves

Up to this point we have neglected the forces of mutual repulsion of the electrons. This has been justified insofar as we have considered low-power devices in which the density of electrons in the beam is small. However, in the next section and in further chapters, we shall consider high-power amplifier tubes in which the forces between electrons play an important role in modifying the rf performance. Therefore, we must study the forces which are produced by the bunches of electrons existing in tubes such as the klystron amplifier and the way in which these forces tend to modify the bunching process.

In Section 3.4, we have considered the forces due to dc space charge in electron beams, that is, in beams of uniform charge density with no bunching. Also discussed were the means of compensating for these forces using uniform or periodic axial magnetic fields or periodic axial electric fields. In the present section we shall consider the complementary effects of the rf bunches of electrons as a perturbation on the electron motion. We shall assume that the beam is confined to a nearly uniform diameter by one of the methods described in Section 3.4. It will be convenient to assume that the various quantities associated with an electron beam consist of a dc part plus an rf perturbation due to the electron bunches. In general, we shall assume that the rf perturbation is small compared with its dc counterpart.

### (a)  A Graphic Illustration of Space-Charge Waves

At this point it may be helpful to consider a graphic illustration of rf space-charge forces in an electron beam. Consider Figure 9.3-1. We wish to study two successive bunches of electrons as they travel down a drift tube at constant dc velocity. There is a nearly infinite magnetic field in the direction of travel so as to prevent radial excursions of the electrons. For simplicity, we shall assume that all the electrons in one transverse plane move as a unit, constituting an inflexible *disc* of charge. This is a good ap-

proximation for a thin beam. Although the electrons are moving to the right with an average velocity given by the dc voltage, we shall confine our attention to the relative motion of the discs. Thus, we shall view the motion from a frame of reference moving at the average electron velocity.

In Figure 9.3-1(a) are shown the discs comprising the two bunches at one instant of time. Discs $D$ and $H$ are at the centers of a bunch and antibunch, respectively. Thus, they have a velocity equal to the average
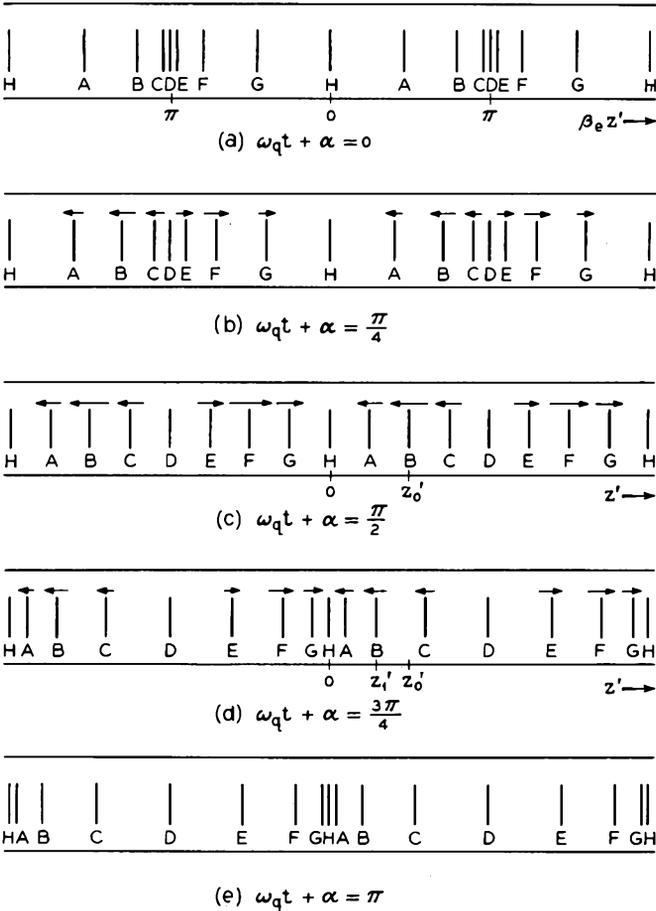


FIG. 9.3-1 Oscillations due to space-charge forces in a reference frame moving at the dc velocity of the electrons. Five successive instants of time are shown, separated by an eighth-cycle of the oscillation frequency. Instantaneous velocity vectors are shown above the charge discs.

velocity of the beam and remain stationary in our frame of reference. The center disc of the bunch exerts a repulsive force on the other discs in the bunch. Hence, discs $A$, $B$, and $C$ are pushed to the left, whereas $E$, $F$, and $G$ are pushed to the right.

Evidently the beam acts as an elastic medium, and oscillations of the discs in the axial direction will occur. A fraction of a cycle later in these oscillations the discs appear as in Figure 9.3-1(b). Discs $A$, $B$, and $C$ are moving to the left (in the reference frame) and are still being accelerated in that direction by the space-charge force of the bunch. At the same time discs $E$, $F$, and $G$ are moving and being accelerated to the right.

Still later, in Figure 9.3-1(c), the discs have attained a uniform spacing. However, they are not stationary in the reference frame. Discs $A$, $B$, and $C$ are moving to the left, and discs $E$, $F$, and $G$ to the right. Each disc sees as much space charge to the left as to the right; hence, the instantaneous acceleration is zero, and the discs have all reached their maximum velocity in the reference system. This is the instant for which the perturbation of charge density is zero; thus the rf components of velocity and charge density are 90 degrees out of phase.

An eighth of a cycle later, we have the situation depicted in Figure 9.3-1(d). A bunch is now beginning to form centered about disc $H$. The repulsive forces emanating from the center of this bunch act in such a direction as to slow down the motion to the left of discs $A$, $B$, and $C$. Similarly, the motion to the right of discs $E$, $F$, and $G$ is also slowed down.

These electrons are finally brought to rest in the reference frame at the instant depicted in Figure 9.3-1(e). At this instant the instantaneous rf velocity is zero. The repulsive bunch is pushing to the right on discs $A$, $B$, and $C$, and to the left on discs $E$, $F$, and $G$.

For the next half cycle, the discs retrace their motions in the opposite directions, appearing as in Figure 9.3-1(d), (c), (b), and (a), successively. After arriving back at the positions in Figure 9.3-1(a), the cycle repeats itself. Thus the discs oscillate back and forth in the reference frame in a simple harmonic motion about the average positions shown in Figure 9.3-1(c).

### (b) Expressions for the AC Velocity, Charge Density, and Current Density

From the preceding discussion, we can write equations describing the motion of the charge discs of Figure 9.3-1. Let us assume that the bunches are produced by an rf source of radian frequency $\omega$ at some distance to the left by some means such as a cavity gap. Since the beam velocity is $u_o$, the bunch spacing is given by

$$u_o \frac{2\pi}{\omega} = \frac{2\pi}{\beta_e} \qquad (9.3\text{-}1)$$

where $\beta_e = \omega/u_o$.

Let $z'$ be the axial coordinate in our frame of reference which moves with the dc beam velocity. We shall assume that the origin is at the location of electron disc $H$ (which we have assumed stationary in this reference frame). Let $\omega_q$ be the frequency at which the discs oscillate back and forth in the reference frame.

If we assume that the velocity and charge-density perturbations are simple sinusoidal variations in both time and position, the behavior shown in Figure 9.3-1 may be described mathematically by the two equations:

$$\rho = B \cos \beta_e z' \cos(\omega_q t + \alpha) \tag{9.3-2}$$

$$u = -C \sin \beta_e z' \sin(\omega_q t + \alpha) \tag{9.3-3}$$

where $\rho$ and $u$ are the instantaneous rf charge density and velocity perturbations. $B$ and $C$ are positive constants determined by the magnitude of the rf perturbation, and $\alpha$ is a constant which determines the phase of the oscillations. The reader may verify that these relations hold for each of the instants illustrated in Figure 9.3-1. These two perturbations are seen to be 90 degrees out of phase in both time and axial position.

Equations (9.3-2) and (9.3-3) may be written in terms of the laboratory reference frame, where $z$ is the axial position, using the relationship

$$z = z' + u_o t \tag{9.3-4}$$

obtaining

$$\rho = B \cos(\beta_e z - \omega t) \cos(\omega_q t + \alpha) \tag{9.3-5}$$

$$u = -C \sin(\beta_e z - \omega t) \sin(\omega_q t + \alpha) \tag{9.3-6}$$

The total charge density and velocity are given by

$$\rho_{tot} = -\rho_o + \rho \tag{9.3-7}$$

$$u_{tot} = u_o + u \tag{9.3-8}$$

where $\rho_o$ is the magnitude of the electron charge density. Similarly, the total current density is written as

$$J_{tot} = -J_o + J \tag{9.3-9}$$

where $J_o$ is the magnitude of the dc current density, and $J$ is the rf perturbation. We shall refer to this current density in free space as convection current density in order to distinguish it from conduction current density flowing in a conductor. The positive direction for this current density is taken in the $+z$ direction.

The instantaneous convection current density at any point is defined as the product of the instantaneous velocity and charge density at that point. Thus, we have

$$J_{tot} = \rho_{tot} u_{tot} \tag{9.3-10}$$

or

$$-J_o + J = -\rho_o u_o + u_o\rho - \rho_o u + \rho u \qquad (9.3\text{-}11)$$

We assume that the rf perturbation is small compared with the corresponding dc quantity; hence, the term $\rho u$ can be neglected in comparison with the other terms since it is the product of two small perturbations. Using the fact that $J_o = \rho_o u_o$, we obtain

$$J = u_o\rho - \rho_o u \qquad (9.3\text{-}12)$$

The equation of continuity, Equation (1.3-2), can be written for the rf perturbations as

$$\frac{\partial J}{\partial z} = -\frac{\partial \rho}{\partial t} \qquad (9.3\text{-}13)$$

since $J$ is not a function of the transverse coordinates. From Equations (9.3-5), (9.3-6), and (9.3-12), we obtain

$$\frac{\partial J}{\partial z} = -\omega B \sin(\beta_e z - \omega t) \cos(\omega_q t + \alpha) + \beta_e \rho_o C \cos(\beta_e z - \omega t) \sin(\omega_q t + \alpha)$$
$$(9.3\text{-}14)$$

and

$$-\frac{\partial \rho}{\partial t} = -\omega B \sin(\beta_e z - \omega t) \cos(\omega_q t + \alpha) + \omega_q B \cos(\beta_e z - \omega t) \sin(\omega_q t + \alpha)$$
$$(9.3\text{-}15)$$

Equating these last two equations, we obtain

$$\omega_q B = \beta_e \rho_o C \qquad (9.3\text{-}16)$$

relating the magnitudes of the velocity and charge-density rf variations. Using this relation, Equation (9.3-12) yields

$$J = u_o B \cos(\beta_e z - \omega t) \cos(\omega_q t - \alpha) + \frac{\omega_q}{\omega} u_o B \sin(\beta_e z - \omega t) \sin(\omega_q t + \alpha)$$
$$(9.3\text{-}17)$$

In practical microwave tubes $\omega_q/\omega$ is small compared with unity, as we shall see later when we evaluate $\omega_q$. Hence, the second term in this equation may be neglected in comparison with the first, and we obtain

$$J = u_o B \cos(\beta_e z - \omega t) \cos(\omega_q t + \alpha) \qquad (9.3\text{-}18)$$

(c)  *The Plasma Frequency*

Next, we shall obtain a method for determining $\omega_q$, the frequency at which the space-charge forces cause the electrons to oscillate back and forth in the reference frame. In addition, we shall verify the sinusoidal behavior

assumed for the rf variations. We shall accomplish this by examining the electron motion of Figure 9.3-1 in a more quantitative fashion.

Our results will be derived for a one-dimensional beam, that is, for a beam that is uniform in the transverse direction and infinite in diameter. The oscillation frequency for this particular case will be designated $\omega_p$. Later we shall show how the results can be altered to apply to a beam of finite cross-sectional area, with or without surrounding metal walls.

Starting with the case of a beam of infinite diameter, we note that if the discs in Figure 9.3-1 are uniformly spaced, as in Figure 9.3-1(c), there is no accumulation of charge at any position and hence no net electric field at any point. A uniform charge density $-\rho_o$ exists everywhere. Consider a typical disc, disc $B$, and let $z_o'$ be its equilibrium position shown in Figure 9.3-1(c). In Figure 9.3-1(d), disc $B$ has moved to a new position $z_1'$. This movement produces a charge excess in the region $0 \le z' \le z_1'$ over that which existed when the discs were uniformly spaced given by

$$\Delta q = -\rho_o(z_o' - z_1') \tag{9.3-19}$$

per unit area of the beam. This excess charge produces a restoring electric field at $z' = z_1'$ given by[4]

$$E_z = -\frac{\rho_o}{\varepsilon_o}(z_o' - z_1') \tag{9.3-20}$$

Thus, the acceleration of the electrons comprising a disc at any position $z'$ is given by

$$\ddot{z}' = \frac{e}{m}\frac{\rho_o}{\varepsilon_o}(z_o' - z') \tag{9.3-21}$$

using Equation (1.1-1). This equation has the solution

$$z' - z_o' = F\cos(\omega_p t + \tau) \tag{9.3-22}$$

where

$$\omega_p{}^2 = \frac{e\rho_o}{m\varepsilon_o} \tag{9.3-23}$$

The frequency corresponding to $\omega_p$ is called the plasma frequency. It is proportional to the square root of the electronic charge density. This frequency

---

[4]The rf electric field as used here and in the remainder of this section is an rf perturbation on the dc electric field. As such, it represents a departure from the equilibrium value. Thus, it need not originate on positive charge and terminate on negative charge; rather, it originates on a *deficiency* of negative charge and terminates on an *excess* of negative charge. Some authors present a clearer physical picture by assuming the electron beam to be completely neutralized by immobile positive ions, so that rf electric field lines may originate on positive charges. This approach is not used here because it implies incorrectly that space-charge neutralization by positive ions is necessary for these space-charge waves to exist.

applies only to a beam of infinite diameter. Practical beams of finite diameter are characterized by a plasma frequency which is less than $\omega_p$. This lower plasma frequency is called the reduced plasma frequency and is designated $\omega_q$. $F$ and $\tau$ are independent of time, but they may be functions of $z_o'$, the relative position in the bunch. For example, for $\beta_e z_o'$ equal to 0 or $\pi$, $F$ must be zero, since discs $H$ and $D$ are stationary in the reference frame.

The motion described by Equation (9.3-22) is simple harmonic motion of electrons about their equilibrium positions in the reference frame. Let us apply this result to the electrons leaving the input gap of a klystron amplifier. These electrons have a velocity given by Equation (9.1-11) at the exit of the gap,

$$u(d) = u_o + \frac{eMA}{mu_o} \sin \omega t_o \qquad (9.3\text{-}24)$$

where $t_o$ is the time at which the electrons pass through the gap. Since the electrons under the influence of the space-charge forces exhibit simple harmonic motion, the velocity at a later time $t$ is given by

$$u_{\text{tot}} = u_o + \frac{eMA}{mu_o} \sin \omega t_o \cos \omega_p(t - t_o) \qquad (9.3\text{-}25)$$

where $\omega_p$ corresponds to the frequency of the harmonic motion as discussed above.

The position of an electron disc is given by the product of the elapsed time and the average velocity to that instant. Since the rf perturbation of velocity is assumed small compared with the dc velocity, the average velocity is given approximately by $u_o$. Hence,

$$z = u_o(t - t_o) \qquad (9.3\text{-}26)$$

This expression may be used to eliminate $\omega t_o$ from the sine term in Equation (9.3-25). Thus, we obtain

$$u_{\text{tot}} = u_o - \frac{eMA}{mu_o} \sin(\beta_e z - \omega t) \cos(\omega_p t - \omega_p t_o) \qquad (9.3\text{-}27)$$

This result is identical to that given by Equation (9.3-6), deduced from Figure 9.3-1, if we take $\alpha = (\pi/2) - \omega_p t_o$ and $\omega_q = \omega_p$. Thus, in an infinite beam the electrons oscillate back and forth in the reference frame at the plasma frequency. In a beam of finite diameter, the restoring forces are somewhat weaker, and we shall find that the frequency of oscillation $\omega_q$ is less than $\omega_p$.

The phase angle $\alpha = (\pi/2) - \omega_p t_o$ is essentially constant for any one bunch. This can be shown as follows. Over one cycle of the modulation on the beam, and hence over one bunch, $t_o$ changes by $2\pi/\omega$, and $\omega_p t_o$ changes by $2\pi\omega_p/\omega$. Since at microwave frequencies $\omega_p/\omega$ is usually small compared

with unity, the total variation in $\omega_p t_o$ over a bunch is small. In Figure 9.3-1 we have presented the results as though $\alpha$ were constant, for simplicity.

### (d)  Beams of Finite Diameter

The results we have derived so far have been derived for an electron beam infinite in cross-sectional area. All practical beams are, of course, of finite diameter; hence, we should inquire how the foregoing results are modified for a finite beam.

Let us return to a consideration of the forces between discs of charge. It will be easier for us to visualize the rf electric field lines if we can have them begin on positive charges and end on negative charges.[5] Since these field lines are due to excess charge (from the equilibrium value) in any region, we can picture a deficiency of negative charge in any region as a positive charge. The electric field lines due to the equilibrium charge density $-\rho_o$ are not considered at all, since the beam is assumed to be confined by one of the focusing schemes of Section 3.4. In other words, the electric field we shall depict is the difference between the total instantaneous electric field and the dc electric field at any point.

The situation for an infinite beam is shown in Figure 9.3-2(a). The negative discs of charge (shown shaded) are clustered around the center of the electron bunch. The positive discs (unshaded) are clustered around the antibunch. The electric field lines in this case are given by straight horizontal lines from the positive to the negative discs.

In Figure 9.3-2(b) the analogous situation is shown for a beam of finite diameter. The electric field lines are no longer straight horizontal lines, but instead they bow out. Thus, the axial component of electric field at any disc is reduced in magnitude. This occurs for two reasons. First, the bowing out of the field lines increases the area through which the total electric flux passes. Second, the tilting of the field lines means that only a portion of the total field exists as an axial component.

As a consequence of the weakening of the axial electric field, the total restoring force on a disc is reduced. Hence, Equation (9.3-20) may be written as

$$E_z = -R^2 \frac{\rho_o}{\varepsilon_o}(z_o' - z_1')  \qquad (9.3\text{-}28)$$

where $R^2$ is a number less than unity and is a function of the beam diameter. The differential equation of motion (9.3-21) is altered by the same factor, and one thus obtains the solution:

$$z' - z_o' = F\cos(R\omega_p t + \theta)  \qquad (9.3\text{-}29)$$

---

[5]See footnote 4.
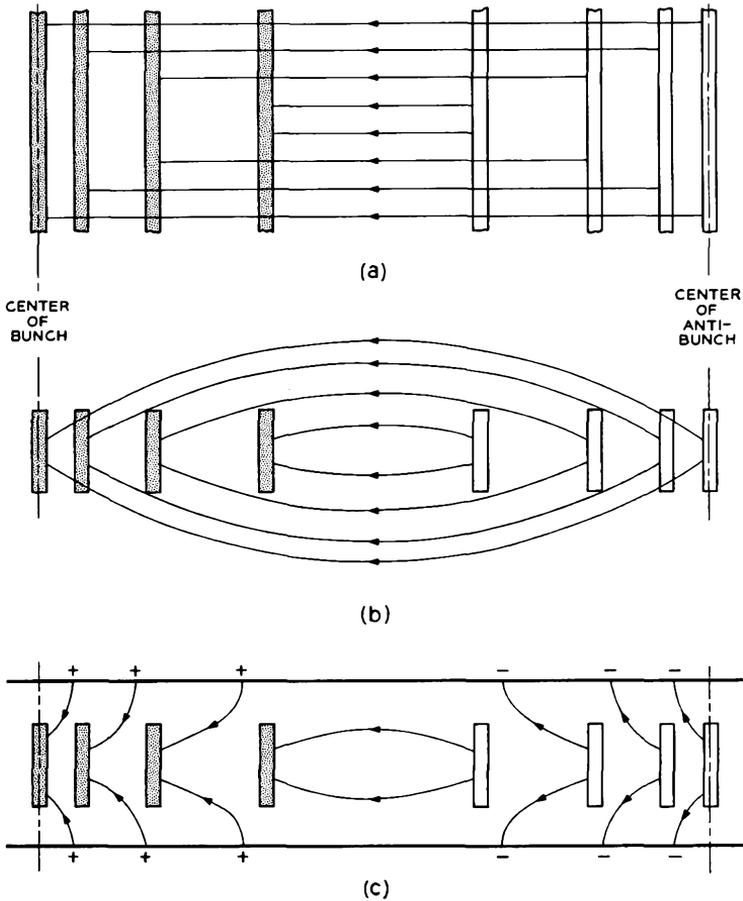
FIG. 9.3-2 RF electric field lines in various beams. Positively charged discs (unshaded discs) are used in regions where the instantaneous electron density is below the dc value. The shaded discs are negatively charged and represent electron densities above the dc value. (a) Portion of an infinite beam in which all quantities are uniform in the transverse directions. The rf electric field lines are straight and parallel to the direction of electric flow. (b) Beam of finite diameter. (c) Finite diameter beam in a drift tube.

for the finite beam, instead of Equation (9.3-22). $R$ is known as the space-charge reduction factor, and we may define the reduced plasma frequency as

$$\omega_q = R\omega_p \qquad\qquad (9.3\text{-}30)$$

All of the results derived for the infinite beam are directly applicable to the finite beam. It is necessary only to replace $\omega_p$ by $\omega_q$. Implicit in this development is the requirement of a strong axial magnetic field so that radial excursions of the electrons due to transverse rf electric fields are prevented.

In Figure 9.3-2(c) is shown the field pattern obtained when the beam is surrounded by a metal cylinder. Many of the field lines terminate on charges on the metal wall. This further reduces the restoring force which tends to bring the discs back to their equilibrium positions. Thus, the reduction factor $R$ is seen to be a function of both the beam diameter and the proximity of a metal wall. As the wall comes closer to the beam, the space-charge forces become smaller, and the frequency of oscillation depicted in Figure 9.3-1 becomes smaller.
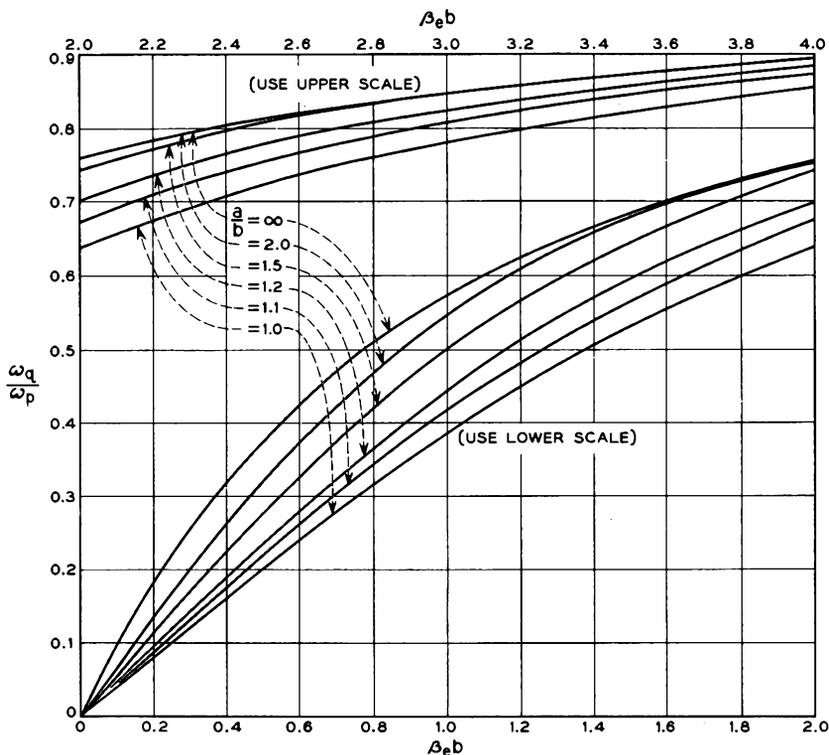


FIG. 9.3-3 Plasma frequency reduction factor vs. beam diameter for a solid, cylindrical beam of radius $b$ in a concentric, perfectly conducting cylinder of radius $a$ (Reference 9.3). (Courtesy of *Transactions IRE*)

Values of the space-charge reduction factor have been calculated for various kinds of beams with surrounding metal walls.[6] These results have been obtained by solving Maxwell's Equations inside and outside of an idealized electron beam, such as we have considered, and matching boundary conditions at the surface of the beam. Figure 9.3-3 shows the result for a solid cylindrical beam centered in a perfectly conducting metal cylinder. The reduction factor is plotted as a function of $\beta_e b = \omega b/u_o$, where $b$ is the beam radius. $a$ is the radius of the conducting cylinder. Several curves are given, each for a different ratio of cylinder diameter to beam diameter. Most microwave tubes have beams with $\beta_e b$ in the range 0.5 to 1.0, so that the reduction factor is typically $\frac{1}{3}$ to $\frac{1}{2}$.

Let us determine the ratio $\omega_q/\omega$ for a microwave tube with the following properties:

$$
\begin{array}{ll}
\text{Beam voltage, volts} \dots\dots\dots\dots\dots\dots & 2400 \\
\text{Beam current, ma} \dots\dots\dots\dots\dots\dots & 40 \\
\text{Beam diameter, mm} \dots\dots\dots\dots\dots & 1.3 \\
\text{Wall diameter, mm} \dots\dots\dots\dots\dots & 2.3 \\
\text{Frequency, Mc} \dots\dots\dots\dots\dots\dots & 6000 \\
\end{array}
$$

We shall describe a traveling-wave tube with these parameters in Chapter 10. From these data we obtain:

$$\rho_o = 1.04 \times 10^{-3} \text{ coulomb/meter}^3$$

$$\frac{\omega_p}{\omega} = 0.12$$

$$\beta_e b = 0.84$$

From Figure 9.3-3 we find that

$$R = 0.46$$

so that

$$\frac{\omega_q}{\omega} = 0.055$$

This confirms the previous assumptions that $\omega_q/\omega$ is small compared with unity in a typical case.

The rf part of Equation (9.3-27) for the instantaneous velocity can be written in a more convenient form as

$$u = -\frac{1}{2} u_o M \frac{A}{V_o} \cos \beta_q z \sin(\beta_e z - \omega t) \qquad (9.3\text{-}31)$$

---

[6]Reference 9.3.

where we have substituted for $t - t_o$ from Equation (9.3-26), and we have set

$$\beta_q = \frac{\omega_q}{u_o} \tag{9.3-32}$$

We use $\omega_q$ here instead of $\omega_p$ so that the results will be applicable to beams of finite diameter. Since $\beta_q$ is much less than $\beta_e$, Equation (9.3-31) represents a wave propagating with a phase velocity equal to the dc beam velocity and whose amplitude is slowly changing with distance. A plot of this equation vs. distance at the instant $\omega t = \pi/2$ is shown in the upper half of Figure 9.3-4.

The equation for the current density may be obtained from Equation (9.3-18) in the same manner. Thus

$$J = -\frac{1}{2}J_o\frac{\omega}{\omega_q}M\frac{A}{V_o}\sin\beta_q z \cos(\beta_e z - \omega t) \tag{9.3-33}$$
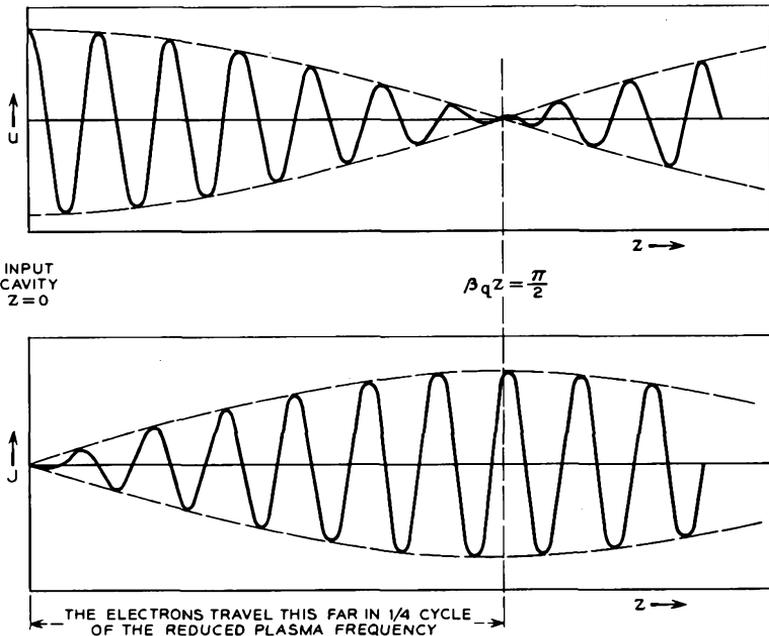


Fig. 9.3-4 Variation with distance of the rf velocity and convection current density for an electron beam which receives velocity modulation from an input cavity placed at $z = 0$. The output cavity is placed at $\beta_q z = \pi/2$ to obtain maximum induced current. The waveforms are shown at a particular instant of time. The envelopes remain stationary, whereas the sinusoidal waves within move to the right at the dc velocity.

where use has been made of Equations (9.3-6), (9.3-16), (9.3-18), (9.3-31) and the fact that $\alpha = \pi/2 - \omega_q t_o$, and $J_o = u_o \rho_o$. $J$ is plotted in the lower half of Figure 9.3-4 for the instant $\omega t = \pi/2$.

From Figure 9.3-4 we note that the velocity modulation magnitude is a maximum at the position of the input cavity gap. Beyond the gap, the velocity modulation magnitude varies as $|\cos \beta_q z|$, whereas the current modulation magnitude varies as $|\sin \beta_q z|$. At a position given by

$$\beta_q z_{\text{opt}} = \frac{\pi}{2} \tag{9.3-34}$$

the current modulation has become a maximum and the velocity modulation is zero. This distance may be termed a quarter-plasma wavelength. If one were to construct a two-cavity klystron, the output cavity would be placed at this distance from the input cavity in order to obtain maximum induced current in the output cavity.

At this point it will be well to note that the foregoing solutions are small-signal solutions, albeit space-charge forces are included. By comparison, the solutions of Section 9.1(b) are applicable under larger signal conditions — limited chiefly by the approximation in going from Equation (9.1-12) to (9.1-13) — but neglecting space-charge forces. The more general problem of large signal interaction with space charge considered requires elaborate computer calculations for solution.[7]

Thus we have from Equation (9.3-33) the result that with space charge at small signal levels the maximum current modulation occurs at the position given by Equation (9.3-34) with a magnitude

$$\frac{i_1}{I_o} = \frac{1}{2} \frac{\omega M A}{\omega_q V_o} \tag{9.3-35}$$

This may be compared with the results of Section 9.1(b) for a finite signal level, but neglecting space charge. From Equation (9.1-24) we have the result that the maximum current modulation is given by

$$\frac{i_1}{I_o} = 1.16 \tag{9.3-36}$$

at a position corresponding to

$$\beta_q z_{\text{opt}} = 3.68 \frac{\omega_c V_o}{\omega M A} \tag{9.3-37}$$

These limiting results may be compared with the numerical results obtained by Webber which include both space charge and finite signal level.[8] Figure

[7]Reference 9.4.
[8]Reference 9.4.
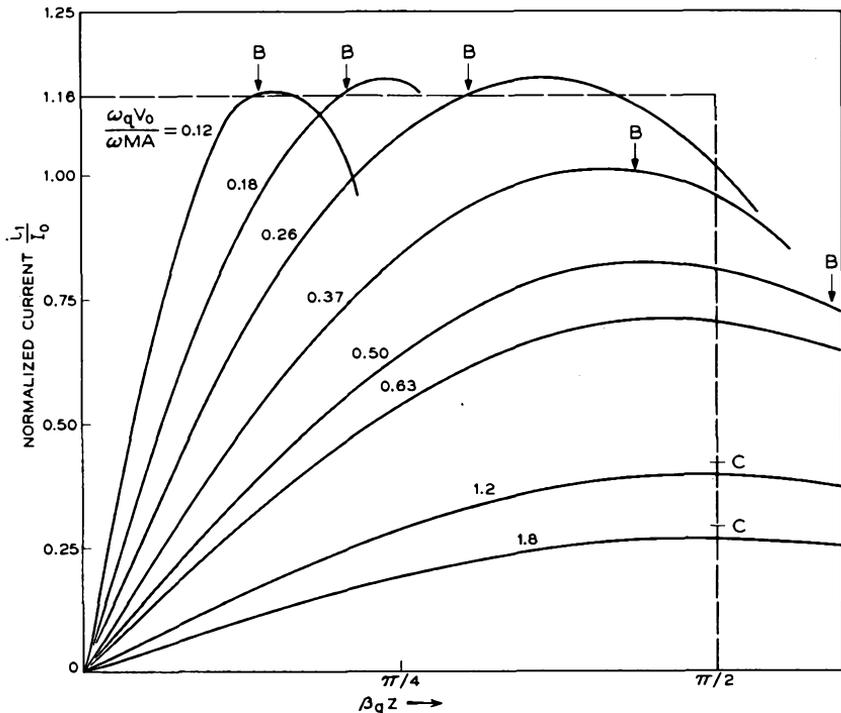
FIG. 9.3-5  Variation of the magnitude of the fundamental component of convection current with distance from the input gap, as given by computer calculations. For a beam of fixed voltage and current, the different curves represent operation at different input signal voltages, $A$. The nonlinear theory, neglecting space charge, developed in Section 9.1 predicts a nonvarying maximum value of 1.16 occurring at values of the abscissa marked $B$ for five of the curves. The small-signal, spacecharge theory developed in this section predicts maxima as given by levels marked $C$ for two of the curves, occurring at a fixed distance given by one quarter of a reduced plasma wavelength. The limits of applicability of the two theories are readily apparent (Reference 9.4). (Courtesy of *Transactions IRE*)

9.3-5 shows Webber's calculated plots of the magnitude of the fundamental component of current modulation vs. distance from the input cavity for various values of the parameter $\omega_q V_o / \omega M A$, which we have seen appears in Equations (9.3-35) and (9.3-37). This parameter may be considered to show the variation with input-cavity-signal level $A$ for a given electron beam, wherein $\omega_q$, $\omega$, $V_o$, and $M$ are constant. At small signal levels, Equations (9.3-34) and (9.3-35) are seen to hold, whereas, as the signal level is increased, the values given by Equations (9.3-36) and (9.3-37) are obtained.

(e)  *Fast and Slow Space-Charge Waves*

The variation of the magnitude of the velocity and current density with distance as shown in Figure 9.3-4 has the appearance of a standing wave. Alternatively the behavior may be written as the sum of two traveling-wave disturbances. Thus, Equations (9.3-31) and (9.3-33) may be written, respectively, as

$$u = \frac{1}{4}u_o M \frac{A}{V_o}[\sin(\omega t - \beta_f z) + \sin(\omega t - \beta_s z)] \qquad (9.3\text{-}38)$$

and

$$J = -\frac{1}{4}J_o\frac{\omega}{\omega_q}M\frac{A}{V_o}[\sin(\omega t - \beta_f z) - \sin(\omega t - \beta_s z)] \qquad (9.3\text{-}39)$$

where

$$\beta_f = \beta_e - \beta_q \qquad (9.3\text{-}40)$$

and

$$\beta_s = \beta_e + \beta_q \qquad (9.3\text{-}41)$$

These two traveling waves are known as the fast and slow space-charge waves, since their phase velocities are respectively faster and slower than the dc beam velocity.[9]

The fast and slow space-charge waves are two normal modes of excitation possible on a constant-dc-velocity electron beam in a drift tube. This means that each wave may exist by itself, or any combination of the two may exist. The input gap in a klystron excites the two waves such that the two velocity components are equal in magnitude and phase at the gap, whereas the current-density components are equal in magnitude but opposite in phase. The two different phase constants $\beta_f$ and $\beta_s$ acting over a distance given by Equation (9.3-34) bring the current density components into phase and the velocity components 180 degrees out of phase. Thus, when space-charge effects are taken into account, the klystron interaction principle may be thought of as an interference effect between the fast and slow space-charge waves.

As a final point, let us consider how the two-cavity klystron can be used to excite an electron beam with only the slow space-charge wave in the region beyond the output cavity. Let the input cavity produce the velocity and current-density variations given by Equations (9.3-31) and (9.3-33).

---

[9]Strictly speaking, the abscissa in Figure 9.3–3 should be $\beta_f b$ or $\beta_s b$ instead of $\beta_e b$ and the space-charge reduction factor $R$ should be computed separately for each space-charge wave. However, since $\beta_q \ll \beta_e$, little error is introduced by using $\beta_e b$ for this computation.

The output cavity is placed a quarter-plasma-wavelength away, and the rf convection current induces current in the cavity. Let us adjust the loading of the output cavity so that this induced current produces an output gap voltage given by

$$-A \cos\left(\omega t - \frac{\pi}{2}\frac{\omega}{\omega_q}\right) = A \sin\left(\omega t - \frac{\pi}{2}\frac{\omega}{\omega_q} - \frac{\pi}{2}\right) \qquad (9.3\text{-}42)$$

Note that this requires simply a conductance of proper value for the equivalent circuit of the output cavity and load. The velocity and current-density variations produced by this second gap voltage alone may be obtained from Equations (9.3-31) and (9.3-33) by replacing $\omega t$ by

$$\omega t - \frac{\pi}{2}\frac{\omega}{\omega_q} - \frac{\pi}{2}$$

and $\beta_q z$ by

$$\beta_q z - \frac{\pi}{2}$$

One obtains after simplification:

$$u_2 = -\frac{1}{2}u_o M \frac{A}{V_o} \sin \beta_q z \cos(\beta_e z - \omega t) \qquad (9.3\text{-}43)$$

and

$$J_2 = -\frac{1}{2}J_o \frac{\omega}{\omega_q} M \frac{A}{V_o} \cos \beta_q z \sin(\beta_e z - \omega t) \qquad (9.3\text{-}44)$$

In the region beyond the output cavity the modulation produced is the superposition of the variations produced by the two gap voltages. When Equations (9.3-31) and (9.3-43) are added and simplified, one obtains

$$u_s = -\frac{1}{2}u_o M \frac{A}{V_o} \sin(\beta_s z - \omega t) \qquad (9.3\text{-}45)$$

Similarly, Equations (9.3-33) and (9.3-44) are added to give

$$J_s = -\frac{1}{2}J_o \frac{\omega}{\omega_q} M \frac{A}{V_o} \sin(\beta_s z - \omega t) \qquad (9.3\text{-}46)$$

Comparing these results with Equations (9.3-38) and (9.3-39), we see that we have obtained purely a slow space-charge wave, whose amplitude is twice as large as that of either space-charge wave in the drift region between cavities.

## 9.4  Multicavity Klystron Amplifiers

Most klystron amplifiers are high-power tubes wherein the rf space-charge forces are quite important. Therefore, it is necessary to use the

space-charge wave theory developed in the previous section to evaluate properly the performance characteristics of these tubes. Let us first consider these relations as they apply to the two-cavity klystron amplifier of Figure 9-2.

The rf velocity and current-density variations between the input and output cavities are as shown in Figure 9.3-4. The output cavity is placed at the point where the rf current modulation is a maximum, that is, at $\beta_q z = \pi/2$. For simplicity, the two cavities are assumed to be identical. From Equation (9.3-33) the magnitude of the rf convection current at the output cavity is given by

$$|i_2| = \frac{1}{2} I_o \frac{\omega}{\omega_q} M \frac{A}{V_o} \tag{9.4-1}$$

From the discussion given in Section 9.1(c), the magnitude of the current induced into the output cavity is equal to[10]

$$|I_2| = M|i_2| = \frac{1}{2} I_o \frac{\omega}{\omega_q} M^2 \frac{A}{V_o} \tag{9.4-2}$$

The equivalent circuit of the output cavity is given in Figure 9.1-7, The magnitude of the impedance of a parallel circuit of this type can be written as[11]

$$|Z_2| = \frac{1}{G_T \sqrt{1 + 4Q_l^2 \left(\dfrac{\Delta f}{f_o}\right)^2}} \tag{9.4-3}$$

where

$G_T$ is the total shunt conductance
$Q_l$ is the hot loaded $Q$ of the cavity
$f_o$ is the resonant frequency
$\Delta f$ is the departure of the operating frequency from $f_o$

For the output circuit,

$$G_T = G_c + g + G_L \tag{9.4-4}$$

where the symbols are defined in Figure 9.1-7. The output cavity is assumed matched to the load for maximum power transfer, that is,

$$. \quad G_L = G_c + g \tag{9.4-5}$$

[10]The results of section 9.1(c) are applicable under the assumption that $\omega_q \ll \omega$ and the output gap is short. Under these conditions, the convection current may be written as in Equation (9.1-25).
[11]Reference 9.5.

The power delivered to the load is therefore given by

$$P_{\text{out}} = \tfrac{1}{2} |I_2|^2 |Z_2|^2 G_L$$

$$= \frac{1}{32}\left(\frac{I_o}{V_o}\right)^2\left(\frac{\omega}{\omega_q}\right)^2\frac{M^4 A^2}{G_{ch}}\frac{1}{1 + 4Q_l^2\left(\dfrac{\Delta f}{f_o}\right)^2} \qquad (9.4\text{-}6)$$

where $G_{ch}$ is defined in terms of the hot unloaded $Q$ of the cavities;

$$G_{ch} \equiv G_c + g \qquad (9.4\text{-}7)$$

The subscript $ch$ is used here to designate "(unloaded) cavity hot." The input power is the power delivered to the input cavity by the source. From Figure 9.1-8, we obtain

$$P_{\text{in}} = \tfrac{1}{2}A^2 G_{ch} \qquad (9.4\text{-}8)$$

The power gain is given by the ratio of $P_{\text{out}}$ to $P_{\text{in}}$:

$$\begin{matrix}\text{power gain} \\ \text{(2-cav. klystron)}\end{matrix} = \frac{M^4}{16}\left(\frac{\omega}{\omega_q}\right)^2\left(\frac{I_o}{V_o}\right)^2\frac{1}{G_{ch}^2}\frac{1}{1 + 4Q_l^2\left(\dfrac{\Delta f}{f_o}\right)^2} \qquad (9.4\text{-}9)$$

Let us examine the gain expression in detail. The last factor determines approximately the frequency dependence of the gain. The power is reduced to one half when $\Delta f/f_o = 1/2Q_l$, so that the half-power bandwidth is given by

$$\text{half-power bandwidth} = f_o/Q_l \qquad (9.4\text{-}10)$$

Thus the bandwidth varies inversely with the loaded $Q$ of the output cavity. However, by virtue of the term $1/G_{ch}^2$, the power gain is proportional to the square of the cavity $Q$'s. Therefore it is possible to trade gain for bandwidth by adjusting the cavity $Q$'s. Normally, two-cavity klystrons are designed for maximum cavity $Q$'s, obtaining maximum gain and efficiency and accepting whatever bandwidth is produced. Thus, the klystron amplifier is inherently a narrow-band device.

From Equations (9.3-23) and (9.3-30) we have the following proportionality:

$$\omega_q^2 \propto \rho_o \propto \frac{I_o}{\sqrt{V_o}} \qquad (9.4\text{-}11)$$

Consequently, the term $(\omega/\omega_q)^2 (I_o/V_o)^2$ in Equation (9.4-7) is proportional to $I_o/V_o^{3/2}$. This means that the power gain given by Equation (9.4-9) is proportional to the beam perveance, as defined in Section 4.5. Thus high perveance electron guns are needed in order to achieve high gain. Gains of 10 db or so are often obtained with two-cavity tubes.

One way of achieving higher overall gain is to connect several two-cavity amplifier tubes, in series, each tube providing approximately 10 db of gain. The power gain of two identical tubes in series is given by squaring the right-hand side of Equation (9.4-9),

$$\text{power gain} \atop (2 \text{ klystrons}) = \frac{M^8}{256}\left(\frac{\omega}{\omega_q}\right)^4\left(\frac{I_o}{V_o}\right)^4\frac{1}{G_{ch}^4}\frac{1}{\left[1 + 4Q_i^2\left(\frac{\Delta f}{f_o}\right)^2\right]^2} \qquad (9.4\text{-}12)$$

where it is assumed that input and output ports are all matched to their transmission lines. A power gain of 40 db might be achieved by connecting four 10-db tubes in series.

High gain is achieved in a much simpler fashion using the multicavity klystron amplifier. A three-cavity klystron amplifier is illustrated in Figure 9.4-1. The construction of this tube is similar to the two-cavity amplifier, differing solely in the number of cavities. The tube functions in the following manner. The input signal impresses velocity modulation on the beam at the input cavity gap. The second cavity is placed a quarter of a plasma wavelength away at the position of maximum rf convection current modulation. The induced current in this cavity produces a voltage across its gap.
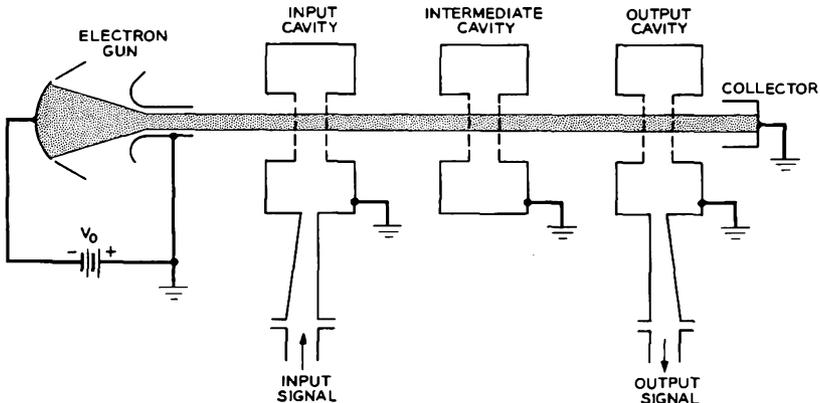


Fig. 9.4-1 Three-cavity klystron amplifier.

This second cavity voltage, which is considerably larger than the first cavity voltage, impresses velocity modulation on the beam at this point. This velocity modulation produces current modulation at the output cavity, a quarter-plasma-wavelength away. The rf convection current passing through the output cavity produces an induced current in the output cavity which causes rf power to be delivered to the load.

Thus we see that the three-cavity klystron amplifier is very much like two two-cavity tubes placed end-to-end. The basic difference is that the output cavity of the first tube and the input cavity of the second tube are combined into one intermediate cavity in the three-cavity tube.

Let us carry out a simplified analysis of the three-cavity amplifier, Assume that all three cavities are identical; that is, all have the same unloaded $Q$ and beam-coupling coefficient. The intermediate cavity is not externally loaded, but the input and output cavities are matched to their transmission lines.

If $A$ is the magnitude of the input-cavity-gap voltage, the magnitude of the rf convection current injected into the intermediate cavity gap is given by Equation (9.4-1), as for the output cavity in the two-cavity tube. The induced current in the intermediate cavity is given by Equation (9.4-2). The magnitude of the gap voltage produced by this current is given by

$$| V_2 | = \frac{1}{2} \frac{\omega}{\omega_q} \frac{I_o}{V_o} \frac{M^2 A}{G_{ch}} \frac{1}{\sqrt{1 + 4Q_u^2 \left(\frac{\Delta f}{f_o}\right)^2}} \qquad (9.4\text{-}13)$$

where $Q_u$ is the hot unloaded $Q$ of the cavity. This voltage produces a velocity modulation on the beam, which is converted into an rf convection current at the output cavity of a magnitude given by Equation (9.4-1), with $A$ replaced by $| V_2 |$, that is

$$| i_3 | = \frac{1}{4}\left(\frac{\omega}{\omega_q}\right)^2 \left(\frac{I_o}{V_o}\right)^2 \frac{M^3 A}{G_{ch}} \frac{1}{\sqrt{1 + 4Q_u^2 \left(\frac{\Delta f}{f_o}\right)^2}} \qquad (9.4\text{-}14)$$

The current induced in the output cavity is $M$ times this value. The power delivered to a matched load is given by

$$P_{\text{out}} = \frac{M^2 | i_3 |^2}{8 G_{ch}} \frac{1}{1 + 4Q_l^2 \left(\frac{\Delta f}{f_o}\right)^2} \qquad (9.4\text{-}15)$$

which, together with the expression for $|i_3|$, becomes

$$P_{\text{out}} = \frac{M^8}{128}\left(\frac{\omega}{\omega_q}\right)^4 \left(\frac{I_o}{V_o}\right)^4 \frac{A^2}{G_{ch}^3} \frac{1}{1 + 4Q_u^2 \left(\frac{\Delta f}{f_o}\right)^2} \frac{1}{1 + 4Q_l^2 \left(\frac{\Delta f}{f_o}\right)^2} \qquad (9.4\text{-}16)$$

The input power is given by Equation (9.4-8), so that the gain is

$$\begin{array}{c} \text{power gain} \\ \text{(3-cav. klystron)} \end{array} = \frac{M^8}{64}\left(\frac{\omega}{\omega_q}\right)^4 \left(\frac{I_o}{V_o}\right)^4 \frac{1}{G_{ch}^4} \frac{1}{1 + 4Q_u^2 \left(\frac{\Delta f}{f_o}\right)^2} \frac{1}{1 + 4Q_l^2 \left(\frac{\Delta f}{f_o}\right)^2}$$

$$(9.4\text{-}17)$$

Comparing this expression with Equation (9.4-12), which gives the gain of two two-cavity tubes in cascade, we note that the three-cavity tube has four times as much power gain at midband. However, we note that it also has less bandwidth by virtue of the fact that $Q_u = 2Q_l$. If we were to load the intermediate cavity by an external load so as to obtain $Q_l$ instead of $Q_u$ for this cavity, the two expressions for gain would be identical. For applications where bandwidth is not important, the higher gain made possible by using an unloaded intermediate cavity is a distinct advantage.

Still higher gain may be obtained by adding additional intermediate cavities. Multicavity klystrons with as many as seven cavities are commercially available, although the most frequently used number of cavities is four. Each of the intermediate cavities functions in the same manner as in the three-cavity amplifier. Gains of greater than 60 db are obtained when the cavities are synchronously tuned, that is, all tuned to the same frequency. However, often multicavity klystrons are operated with their cavities stagger-tuned so as to obtain greater bandwidth at some reduction in gain. This is analogous to the well-known design of wide-band IF amplifiers, wherein each stage is tuned to a slightly different frequency so as to improve the overall gain-bandwidth product. High-power klystron amplifiers with 40 to 50 db of gain and bandwidths equal to several per cent of the midband frequency are commercially available.

In high-power klystrons the cavity grids are omitted, since they would burn up due to beam interception. The beam-coupling coefficient in this case is given by a more complicated expression than Equation (9.1-10), but otherwise the interaction is unchanged.

Figure 9.4-2 shows two photographs of the Varian Associates VA-849,[12] which produces an output power of 24 kw CW (continuous wave). A cross-sectional drawing of the tube is shown in Figure 9.4-3. The tube can be purchased with cavities tuned to operate at any center frequency in the range 7125 to 8500 Mc. It is about 45 cm long and weighs 14 pounds. Figure 9.4-2(b) shows the tube in place in its electromagnet. The magnet weighs 200 pounds and dissipates a power of 1520 watts. It provides the axial dc magnetic field for focusing the beam.

The operating characteristics of the VA-849 are given in Table 9.4-1. The tube has four cavities, each with a cold unloaded $Q_o$ of about 5000. Data are presented in the table for the synchronously tuned situation and also for the case when the third cavity has its resonant frequency tuned higher. The cavity gaps are equally spaced by a distance corresponding to approximately one ninth of a plasma wavelength. This spacing, rather

---

[12]Reference 9.6.
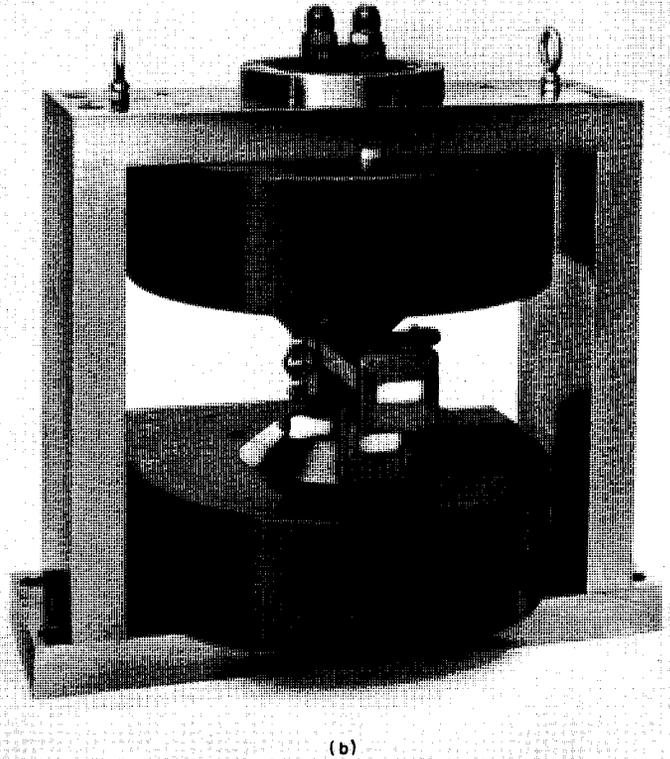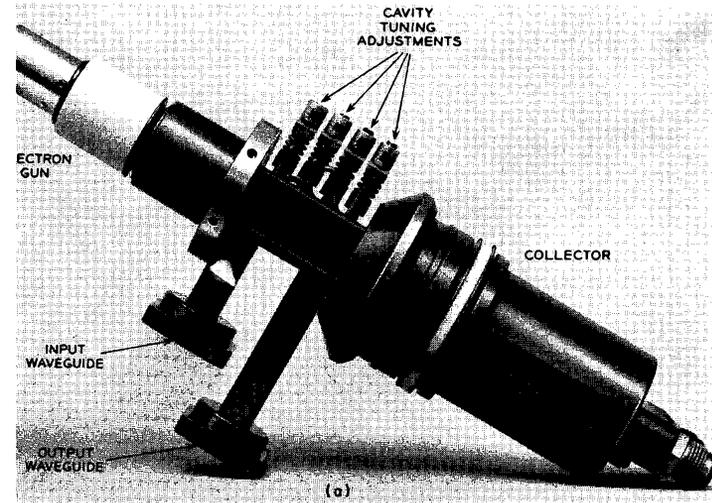
FIG. 9.4-2 VA-849 four-cavity klystron amplifier. (a) The tube without its electromagnet. The overall length of the tube is approximately 45 cm. (b) The VA-849 in place in its electromagnet. (Courtesy of Varian Associates)

CAVITY TUNERS

ANODE

CATHODE STRUCTURE

COLLECTOR

RF OUTPUT WINDOW

RF INPUT WINDOW

METAL PARTS
CERAMIC PARTS
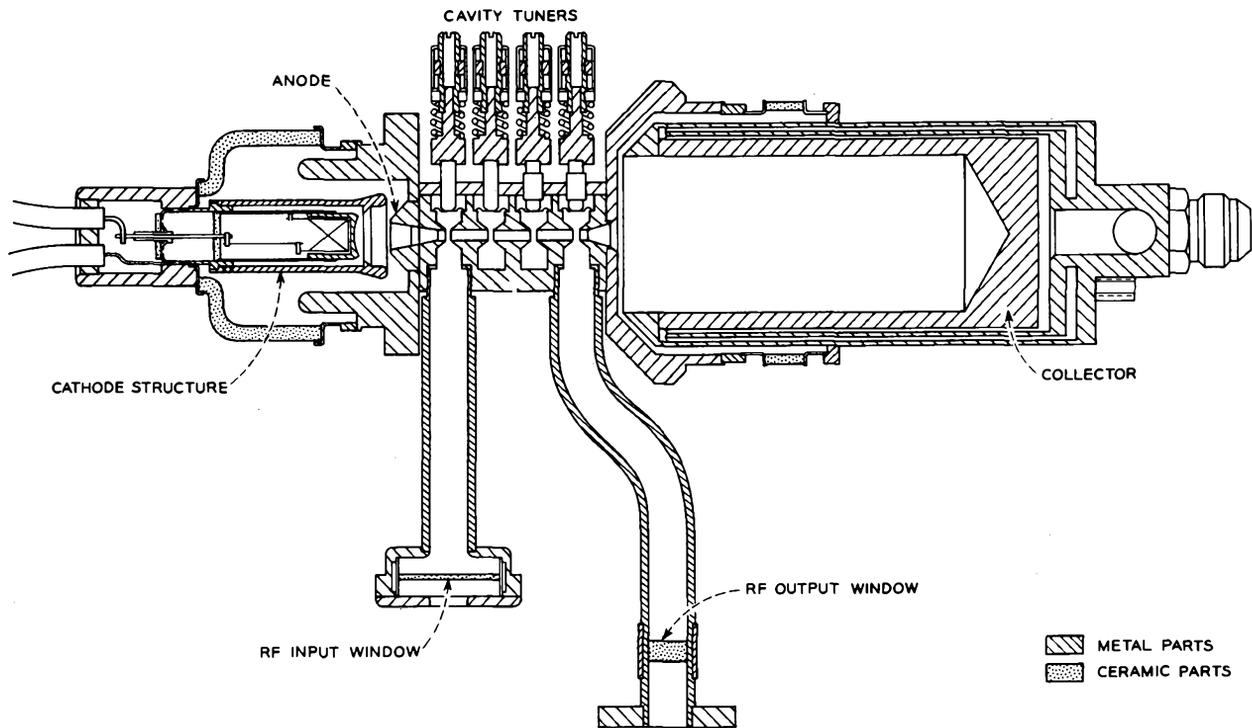
FIG. 9.4-3 Cross-sectional drawing of the VA-849. Each cavity is tuned by deforming a sidewall of the cavity. (Courtesy of Varian Associates)

than a quarter-plasma wavelength, is used in order to increase the bandwidth and efficiency of the tube. The data given in Table 9.4-1 demonstrate

TABLE 9.4-1. VA-849 OPERATING CHARACTERISTICS

Frequency range, Mc...................7125–8500
Tuning range, Mc......................60
Beam voltage, kv......................23
Beam current, amps....................2.7
Perveance, amp/volt$^{3/2}$.................0.77 × 10$^{-6}$
Beam diameter, mm....................2.8

|  | Synchronously Tuned | Third Cavity Detuned |
|---|---|---|
| Saturated power output, kw............. | 18 | 24 |
| Drive power, watt..................... | .08 | 1 |
| Gain, db............................. | 54 | 43 |
| Bandwidth (3 db), Mc................. | 18 | 30 |
| Electronic efficiency, %.............. | 29 | 38 |

that judicious staggering of the cavity resonant frequencies not only increases the bandwidth but also increases the electronic efficiency, where electronic efficiency is defined as the ratio of rf output power to the dc beam power. The reason for this increase is beyond the scope of this text, but it is an extremely important attribute of the multicavity klystron amplifier. The collector is operated at the same voltage as the cavities and drift regions.

In summary, klystron amplifiers are characterized by high gain, very good efficiency, and freedom from oscillations. On the other hand, their bandwidths are relatively small. In addition, the phase shift through the tube is directly related to the beam velocity; thus, high regulation and low ripple are required in the beam voltage power supply to avoid undesirable phase-shift variations.

## PROBLEMS

9.1 It is proposed to construct a two-cavity klystron amplifier which will amplify a 1-volt, 1000 cps signal. The input signal is applied directly to the grids of the input cavity to modulate a 100 electron volt electron beam. Explain why such a tube would be impracticable.

9.2 A two-cavity klystron is to be designed to operate as a harmonic generator. The distance between cavities may be varied so as to optimize the induced current for the desired harmonic. Calculate the magnitudes of the harmonic currents induced in the output cavity for second, third, and fourth harmonic operation. Compare these values with the magnitude of the fundamental induced current used for

two-cavity amplifier interaction. The cavity gap dimensions remain fixed; the gap transit angle is $\pi/2$ radians for the fundamental frequency.

$$Ans.: I_2/I_1 = 0.581, \quad I_3/I_1 = 0.248, \quad I_4/I_1 = 0.$$

9.3  The primary beam current passing through the cavity of a reflex klystron is $I_o$ amps. If the tube is operating in the $1\frac{3}{4}$ mode and generating power at the hot resonant frequency $f_o$ of the cavity (i.e., the resonant frequency when beam loading is included), show that the average amount of charge present in the region between the cavity gap and the repeller is given by $-7I_o/4f_o$.

9.4  The repeller electrode of a particular reflex klystron is planar and parallel to the grids of the resonant cavity, so that the electrons between the gap and the repeller experience a uniform retarding field. Voltages are applied as in Figure 9.2-1. The voltage $V_{Ro}$ is such that the electron transit angle $\theta_o$ is equal to $(n + \frac{3}{4})2\pi$.

(a) Show that if an incremental voltage $\Delta V_{Ro}$ is added to the voltage $V_{Ro}$, the incremental change in transit angle is given by

$$\Delta\theta = \theta_o\left(\frac{\Delta\omega}{\omega_o} - \frac{\Delta V_{Ro}}{V_o + V_{Ro}}\right)$$

(b) Using Figure 9.2-4, show that $\Delta\theta$ is also given by

$$\Delta\theta = -2Q_l\frac{\Delta\omega}{\omega_o}$$

(c) Show that if $Q_l$ is large compared with $\theta_o$,

$$\Delta\omega = +\frac{\omega_o}{Q_l}(n + \tfrac{3}{4})\pi\frac{\Delta V_{Ro}}{V_o + V_{Ro}}.$$

9.5  Does the plasma transit angle $\beta_q L$ between cavities vary with frequency? Here $L$ is taken to be the distance between cavities. Explain.

9.6  An electron beam, confined to flow in the $z$ direction by an infinite magnetic field, completely fills a perfectly conducting metal cylinder. $\beta_e b = \beta_e a = 1$. Space-charge waves exist on this beam as given by Equations (9.3-31) and (9.3-33). At a point where the ac component of convection current is zero (i.e., at $\beta_q z = n\pi$), the diameter of the metal cylinder is abruptly doubled. Assume that the dc beam velocity remains unchanged through this discontinuity. At the discontinuity, conservation of kinetic energy dictates continuity of the ac velocity, and the ac convection current is always continuous at a discontinuity where there is no beam interception.

Find the ratios of the maximum values of ac velocity and convection current of the second drift region to the corresponding quantities of the first. Use Figure 9.3-3 to determine the space-charge reduction factors. $Ans.:$ 1.0, 0.7.

9.7  A pure slow space-charge wave results in the beam beyond the output cavity when the output gap voltage is given by Equation (9.3-42).

(a) What is the value of output circuit conductance which results in this voltage?

(b) Show that a negative conductance of the same value results in a pure fast space-charge wave beyond the output cavity.

These results demonstrate that a slow space-charge wave is excited by extracting rf energy from an electron beam, whereas a fast space-charge wave is excited by adding rf energy to an electron beam.

9.8 A three-cavity klystron amplifier is designed to operate at a midband frequency of 9 Gc. Assume gridded gaps of 1.27-mm separation and that cavity losses are negligible. The beam voltage is 5 kv, and the beam current is 1 amp. The beam diameter is 5.08 mm, and the space-charge reduction factor $R$ is equal to 0.6.

(a) Show that the midband gain is equal to 69.3 db.
(b) What is the distance between cavity gaps?

9.9 The gain-bandwidth product of an amplifier is defined as the product of the midband voltage gain and the half-power bandwidth. In the case of a klystron amplifier, the voltage gain may be taken as the square root of the power gain.

(a) Obtain an expression for the gain-bandwidth product of two two-cavity klystron amplifiers in series from Equation (9.4-12).
(b) Obtain an expression for the gain-bandwidth product of the three-cavity klystron amplifier from Equation (9.4-17), and compare with the result for part (a).

9.10 Derive an expression for the power gain of a four-cavity klystron amplifier. All cavities are spaced a quarter-plasma wavelength apart. The two intermediate cavities have no external loading, but the input and output cavities are matched to the external circuits for maximum power transfer.

9.11 A two-cavity klystron amplifier is made into an oscillator by feeding back some of the power from the output cavity into the input cavity. The feedback factor (the ratio of input-cavity-gap voltage to output-cavity-gap voltage) is $D\epsilon^{-i\varphi}$, where $D$ and $\varphi$ are real numbers varied to change the frequency of oscillation. Neglect any cavity loading caused by the feedback path. The load is matched to the output cavity for maximum power transfer. The spacing between cavities is assumed to be a quarter-plasma wavelength for all frequencies of interest.

(a) Sketch equivalent circuits for both cavities, labeling voltages, currents and impedances in the conventional manner.
(b) Using the space-charge wave equations, find the ratio of the phasors representing the input-gap voltage and the current induced in the output cavity.
(c) From the equivalent circuit and the results of part (b), find the ratio of the phasors representing the two cavity-gap voltages and equate this to the ratio given by the feedback factor.
(d) Taking the real and imaginary parts of the result of part (c), obtain two equations relating $\omega$, $\varphi$, and $D$.
(e) Eliminate $D$ between the latter two equations and obtain $\varphi$ as a function of $\omega$, $\omega_q$, and the cavity parameters.

## REFERENCES

Three references on klystron principles and reflex klystrons are:

9a. K. R. Spangenberg, *Vacuum Tubes*, Chapter 17, McGraw-Hill Book Co., Inc., New York, 1948.
9b. J. C. Slater, *Microwave Electronics*, Chapter 10, D. Van Nostrand Co., Inc., Princeton, N. J., 1950.

9c. W. W. Harman, *Electronic Motion*, Chapter 7, McGraw-Hill Book Co., Inc., New York, 1953.

9d. A. H. W. Beck, "Velocity-Modulated Thermionic Tubes," Cambridge University Press, Cambridge, England, 1948.

Three references on space-charge waves are:

9e. A. H. W. Beck, *Space-Charge Waves and Slow Electromagnetic Waves*, Chapters 4 and 6, Pergamon Press, Inc., New York, 1958.

9f. R. G. E. Hutter, *Beam and Wave Electronics in Microwave Tubes*, Chapter 9, D. Van Nostrand Co., Inc., Princeton N. J., 1960.

9g. W. J. Kleen, "Electronics of Microwave Tubes," Chapter 11, Academic Press, Inc., New York, 1958.

Other references covering specific items discussed in the chapter are:

9.1 E. Jahnke and F. Emde, *Tables of Functions with Formulae and Curves*, Dover Publications, Inc., New York, 1945.

9.2 E. L. Ginzton, *Microwave Measurements*, Chapter 9, McGraw-Hill Book Co., Inc., New York, 1957.

9.3 G. M. Branch and T. G. Mihran, "Plasma Frequency Reduction Factors in Electron Beams," *Trans. IRE* **ED-2,** 3–11, April, 1955.

9.4 S. E. Webber, "Ballistic Analysis of a Two-Cavity Finite Beam Klystron," *Trans. IRE* **ED-5,** 98–108, April, 1958.

9.5 *Reference Data for Radio Engineers*, 4th Ed., International Telephone and Telegraph Corp., p. 242, 1956.

9.6 E. McCune, I. Maltzer, and L. T. Zitelli, "A 20 Kw CW X-Band Klystron Amplifier," *Microwave Journal* **IV,** 74–78, August, 1961.

# Chapter 10

## TRAVELING-WAVE AMPLIFIERS

The multicavity klystron amplifier described in the previous chapter was shown to have high gain and efficiency at microwave frequencies. However, its fractional bandwidth was found to be at most a few per cent. The fractional bandwidth is essentially determined by the cavity $Q$'s. Lowering these $Q$'s results in greater bandwidth, but the overall gain is reduced, as shown by Equation (9.4-17). Let us look for modifications of the multi-cavity-klystron-amplifier structure which will increase the bandwidth without greatly reducing the gain.

The three-cavity klystron amplifier is shown in Figure 10-1(a). The $Q$'s could be reduced by increasing the resistive loading of each cavity, either by increasing the cavity losses or by coupling each cavity to an external dissipative load. However, the power dissipated in this extra loading would be wasted. Instead, let us couple each cavity to a common transmission line, adjusting either the line length between cavities or the beam voltage so that the transmission line current arrives at the second and third cavities in phase with the current induced in these cavities by the electron beam. A suitable arrangement is shown in Figure 10-1(b). The transmission line loading has the effect of lowering the cavity $Q$'s without a corresponding power loss. The microwave power is fed forward along the transmission line, increasing the voltage in each cavity and finally appearing at the output cavity.

We can go one step further, introducing additional cavities between the three we already have, as in Figure 10-1(c). Successive cavity gaps are now closer than a quarter-plasma-wave length, so that the beam becomes only partially bunched between successive cavities. Nevertheless, the additional cavities result in a higher gain per unit length of the tube, as we shall see later, although the gain per cavity is less. As in Figure 10-1(b), the operating conditions are such that the transmission line current arrives at each
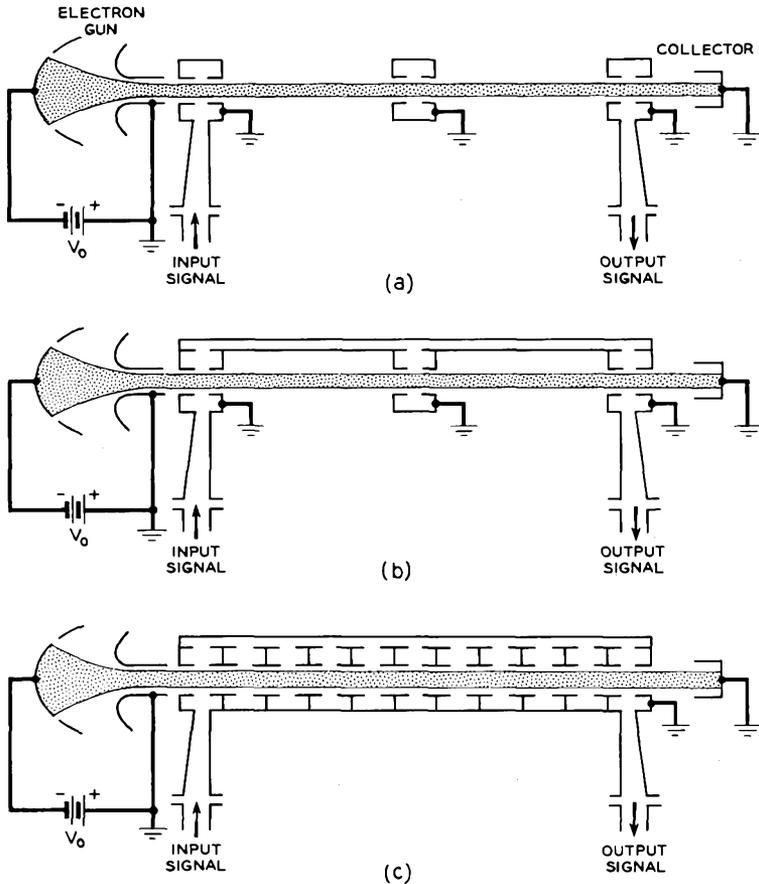
FIG. 10-1 Evolution of a traveling-wave amplifier from a multicavity klystron. (a) Multicavity klystron. (b) Multicavity klystron whose cavities are coupled by a transmission line. (c) Traveling-wave amplifier.

cavity in phase with the induced current due to the beam. The tube of Figure 10-1(c) is a traveling-wave amplifier.

One may ask: "What is the effect of the transmission line energy that travels to the left from each cavity?" In a properly designed traveling-wave amplifier, the backward-traveling energy contributions from successive cavities are of such a phase that they cancel, and thus there is no appreciable net amount of energy traveling in the backward direction.[1]

---

[1] In the backward-wave oscillator (Chapter 11), these backward-traveling contributions do add in phase, producing positive feedback and hence oscillations.

The assemblage of coupled cavities constitutes a slow-wave structure. As we have seen in Chapter 8, the slow-wave structure can be characterized by a Brillouin diagram, as in Figure 8.7-8. The phase shift per period of the fundamental wave on the circuit is given by

$$\theta_1 = \beta_o L \tag{10-1}$$

where $\beta_o$ is the phase shift per unit length, and $L$ is the period.

The electrons take a time $T = L/u_o$ in traveling from one cavity to another. If we wish a given electron to see *the same phase* of the rf signal as it passes through successive cavities, the phase change in a given cavity in time $T$ must be $\theta_1 + 2n\pi$. Hence

$$\omega = \frac{\theta_1 + 2n\pi}{T} = \frac{(\beta_o L + 2n\pi)u_o}{L} \tag{10-2}$$

Setting $\beta_e = \omega/u_o$, we obtain

$$\beta_e = \beta_o + \frac{2n\pi}{L} = \beta_n \tag{10-3}$$

where $\beta_n$ is the phase shift per unit length of the $n^{th}$ space harmonic Equation (10-3) may also be stated as

$$u_o = v_{pn} \tag{10-4}$$

where $v_{pn} = \omega/\beta_n$ is the space-harmonic phase velocity. Thus, traveling-wave amplification is obtained when the dc beam velocity is approximately equal to the phase velocity of the fundamental or any of the space harmonics with positive group velocity. Large fractional bandwidths are possible, since slow-wave structures may be designed with Equation (10-4) holding over a considerable frequency range.

Because of the large number of periods in practical slow-wave structures and the continuous nature of some (the helix, for example), it is convenient to treat the amplification process as the continuous interaction of velocity and convection-current waves on an electron beam with an electromagnetic wave propagating along the slow-wave circuit. This circuit wave is the space harmonic in synchronism with the beam, synchronism being defined by Equation (10-4). The nonsynchronous space harmonics have no net interaction with the electrons, and so they may be neglected. Historically, this was the approach first used in the discovery and analysis of the traveling-wave amplifier.

The interaction between an electron stream and a traveling-electromagnetic wave is illustrated in Figure 10-2, where disc electrons are used, as in Chapter 9. We view the interaction from a reference frame traveling in synchronism with the electromagnetic wave, for which the axial component of electric field is shown. In practice, the dc velocity of the electrons
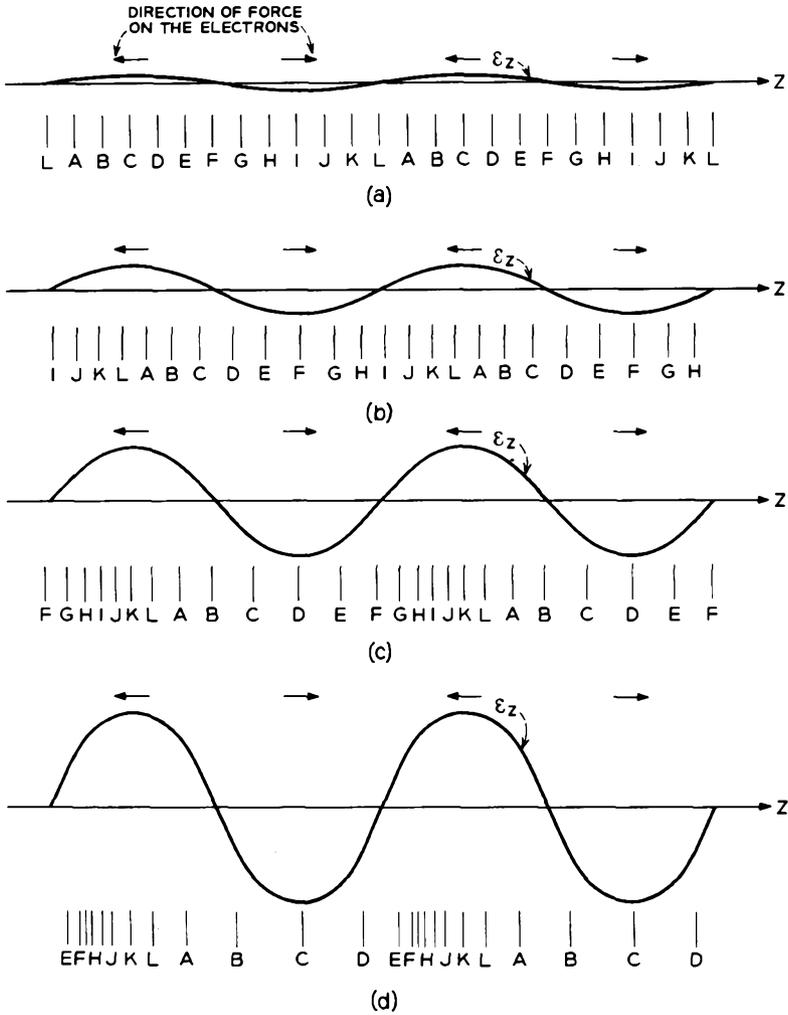
FIG. 10-2 Electron discs interacting with a traveling wave. The arrows indicate the direction of the force on the electrons due to the wave. Figure (a) corresponds to the input end of the slow-wave structure. The situations presented in Figures (b), (c), and (d) are found at positions successively farther down the tube.

is adjusted to be *slightly* greater than that of the electromagnetic wave; consequently the electrons drift to the right in the reference frame.

Figure 10-2(a) shows the conditions at the input to the slow-wave structure. The discs are uniformly spaced. Axial electric field exists due to volt-

age on the slow-wave structure from the input waveguide. This electric field exerts a force to the left on discs $A$, $B$, $C$, $D$, and $E$ and to the right on discs $G$, $H$, $I$, $J$, and $K$. These forces cause the electrons to begin to form a bunch centered about disc $L$.

Figure 10-2(b) shows the conditions farther down the tube. Since the dc velocity of the electrons is slightly greater than the wave velocity, disc $L$ has drifted into a retarding electric field, and the bunch has drifted with it. Since more electrons are in a retarding field than in an accelerating field, a net transfer of energy occurs from the beam to the electromagnetic field, thus increasing the axial electric field slightly.

In Figure 10-2(c), we observe what happens still farther down the tube. The bunch continues to become more compact and it drifts farther to the right with respect to the wave. This brings more electrons into a retarding field, and the electromagnetic wave grows in amplitude.

Figure 10-2(d) shows the conditions near the end of the tube. The bunch is even tighter, and most of the electrons are in a retarding field. An appreciable fraction of the kinetic energy of the electrons has been converted into energy stored in the electromagnetic field, and the amplitude of the field has greatly increased. The amplified wave then propagates out through the output waveguide to the load.

We note that the velocity modulation produced by the circuit electric field and the subsequent conversion to density modulation is used to transer energy from the beam to the circuit, just as in a klystron. However, unlike a klystron, velocity modulation, conversion to current modulation, and the inducing of currents into the circuit occur simultaneously and continuously along the whole length of the slow-wave structure.

In the following sections we shall consider the quantitative aspects of this interaction process.

## 10.1 Theory of the Traveling-Wave Amplifier

We shall consider the interaction process in the traveling-wave amplifier in two parts.[2] First, we shall describe the motion of the electrons which results from the rf electric fields due to both the rf circuit voltage and the rf space charge.[3] We have already considered a similar physical problem in Section 9.3, except in that case the space charge was the only source of electric field forces. The second part of this development concerns the manner in which the circuit voltage and current build up as current is induced into the circuit by the rf current in the electron beam. In the

---

[2] Reference 10a, Chapters 1 and 8. Reference 10b, Chapter 7.

[3] The forces due to the rf magnetic field are negligible in comparison with those due to the rf electric field.

development that follows, we shall derive two important equations, one giving the ac current induced on the beam by the rf field and one giving the rf field resulting from the modulated beam. These two equations can then be solved simultaneously to determine the self-consistent relations for the circuit and beam quantities.

We consider only the space harmonic of the circuit field in synchronism with the beam, as discussed in the introduction, since the other space harmonics travel at different phase velocities and have no cumulative interaction with the electrons.

As in the previous chapter, we consider only the rf forces, assuming that the dc forces are either zero or balanced by one of the focusing schemes described in Section 3.4. For simplicity, we assume rf motion of the electrons is possible only in the axial direction, as is the case for a strong axial magnetic field.

Script letters will be used in this section to denote time-varying quantities to distinguish them from phasor quantities.

The rf beam quantities are assumed small compared with their dc counterparts, as in Section 9.3. Under this small-signal assumption, nonlinear terms in the equations of motion can be neglected, and we obtain only linear differential equations. We shall therefore obtain wavelike solutions for the various quantities of the form

$$\mathcal{U} = \mathrm{Re}[u\epsilon^{j\omega t - \Gamma z}] \qquad (10.1\text{-}1)$$

where $u$ is a phasor quantity having no time or $z$ dependence. We thus introduce a generalized phasor notation. In this notation, the phasor quantity is multiplied by

$$\epsilon^{j\omega t - \Gamma z}$$

before taking the real part. Using this notation, we have

$$\frac{\partial \mathcal{U}}{\partial t} = \mathrm{Re}[j\omega u\epsilon^{j\omega t - \Gamma z}] \qquad (10.1\text{-}2)$$

and

$$\frac{\partial \mathcal{U}}{\partial z} = \mathrm{Re}[-\Gamma u\epsilon^{j\omega t - \Gamma z}] \qquad (10.1\text{-}3)$$

so that partial differentiations with respect to time and $z$ correspond respectively to multiplication of the phasor quantity by $j\omega$ and $-\Gamma$. Since all the rf quantities, both circuit and beam, have variations of the form given by Equation (10.1-1), we may write the various physical equations in the phasor notation, omitting the exponential factor.

From the description of traveling-wave interaction given in the intro-

duction, we would expect $\Gamma$ to have an imaginary part approximately equal to $j\beta_e = j\omega/u_o$.

## (a) The Electronic Equation

The total instantaneous beam velocity, convection current, and charge density are written in the mixed phasor notation as

$$u_{\text{tot}} = u_o + u \qquad (10.1\text{-}4)$$

$$i_{\text{tot}} = -I_o + i \qquad (10.1\text{-}5)$$

$$\rho_{\text{tot}} = -\rho_o + \rho \qquad (10.1\text{-}6)$$

where $u$, $i$, and $\rho$ are phasors representing the rf components in the notation of Equation (10.1-1).[4]

We shall assume for simplicity that the various quantities are uniform across any transverse plane of the beam. Equation (9.3-12) for the convection current density can be written as

$$i = (u_o\rho - \rho_o u)S \qquad (10.1\text{-}7)$$

where $S$ is the cross-sectional area of the beam. The equation of continuity (1.3-2) is written for the phasor quantities as

$$-\Gamma i = -j\omega\rho S \qquad (10.1\text{-}8)$$

Combining these two equations to eliminate $\rho$, we obtain

$$u = -\frac{j\omega - u_o\Gamma}{j\omega\rho_o S}i \qquad (10.1\text{-}9)$$

relating the rf velocity and convection current.

The acceleration of an electron is given by Equation (1.1-1) as

$$\frac{d\mathfrak{u}}{dt} = -\frac{e}{m}\mathcal{E}_{zT} \qquad (10.1\text{-}10)$$

where $\mathcal{E}_{zT}$ is the total instantaneous rf electric field as seen by the electron. This field is the sum of two contributions, space charge and the synchronous space harmonic due to the circuit voltage. Since the velocity of an electron is a function of both position and time, the total derivative in Equation (10.1-10) must be written as

$$\frac{d\mathfrak{u}}{dt} = \frac{\partial\mathfrak{u}}{\partial t} + \frac{\partial\mathfrak{u}}{\partial z}\frac{\partial z}{\partial t} \qquad (10.1\text{-}11)$$

---

[4]The use of the symbol $i$ as a phasor quantity representing the beam convection current has become standard in microwave tube work. Unfortunately, this symbol is also used for instantaneous circuit current. The latter usage will be avoided in this and succeeding chapters on microwave tubes to avoid confusion.

For small-signal levels $\partial z/\partial t$ is approximately given by the dc beam velocity. Thus

$$\frac{d\mathfrak{u}}{dt} = \frac{\partial \mathfrak{u}}{\partial t} + u_o \frac{\partial \mathfrak{u}}{\partial z} \qquad (10.1\text{-}12)$$

Equation (10.1-10) then becomes in the phasor notation:

$$(j\omega - u_o\Gamma)u = -\frac{e}{m}E_{zT}$$

$$= -\frac{e}{m}(E_{zn} + E_{zsc}) \qquad (10.1\text{-}13)$$

where $E_{zn}$ is the axial component of electric field contributed by the synchronous space harmonic, and $E_{zsc}$ is the field due to space charge.

In order to evaluate the field due to space charge, we merely need to solve Poisson's Equation, Equation (1.4-9), where $\rho$ is given by the rf space charge. This will be done using a one-dimensional beam of infinite cross section, with the effect of a finite beam diameter accounted for as in Section 9.3. Poisson's Equation for the one-dimensional beam is given in the phasor notation by

$$-\Gamma E_{zsc} = \frac{\rho}{\epsilon_o} \qquad (10.1\text{-}14)$$

Using Equation (10.1-8), we obtain a simple relationship between $E_{zsc}$ and the convection current:

$$E_{zsc} = j\frac{i}{\omega\epsilon_o S} \qquad (10.1\text{-}15)$$

For a finite beam, it was shown in Section 9.3 that the force due to space charge is reduced by the space-charge reduction factor $R^2$. Thus we have for a finite beam

$$E_{zsc} = jR^2\frac{i}{\omega\epsilon_o S} \qquad (10.1\text{-}16)$$

This may be written in terms of the reduced plasma frequency as

$$E_{zsc} = j\frac{m\omega_q^2}{e\rho_o S\omega}i \qquad (10.1\text{-}17)$$

where use has been made of Equations (9.3-23) and (9.3-30).

Finally, Equations (10.1-9), (10.1-13), and (10.1-17) can be combined to eliminate $E_{zsc}$ and $u$. Thus we obtain

$$i = \frac{j\beta_e I_o E_{zn}}{2V_o\left[(\Gamma - j\beta_e)^2 + \dfrac{\omega_q^2}{u_o^2}\right]} \qquad (10.1\text{-}18)$$

This equation is called the *electronic equation*, since it determines the ac beam convection current $i$ resulting from a given circuit voltage, as char-

acterized by the space harmonic $E_{zn}$. The time and $z$ dependence of $i$ and $E_{zn}$ are as in Equation (10.1-1).

### (b)  The Circuit Equation

The convection current in the beam causes current to be induced in the circuit. This induced current adds to current already present in the circuit, causing the circuit power to increase with distance. Shortly, we shall derive an equation expressing this relationship.

It will be convenient to have a parameter that relates the space-harmonic amplitude of the electric field to the total power flow on the slow-wave circuit. Let us define the beam interaction impedance (or beam-coupling impedance) for the $n^{th}$ space harmonic as

$$K_n = \frac{\int |E_{zn}|^2 dS}{2\beta_n^2 PS} \tag{10.1-19}$$

where $|E_{zn}|$ is the amplitude of the axial electric field of the $n^{th}$ space harmonic, $\beta_n$ is the phase constant of the $n^{th}$ space harmonic, $P$ is the average circuit power flow, $S$ is the cross-sectional area of the beam, and the area integral is taken over the cross section of the beam. For brevity, we shall refer to this quantity as the impedance. We shall find later that high impedance leads to high gain per unit length of the slow-wave circuit. In order to obtain high impedance, it is important that the electric field be concentrated in the vicinity of the electron beam. This precludes the use of dielectric loaded structures, where the electric field stored in the dielectric does not contribute to the interaction with the electron beam.

As an example, we can calculate the impedance for the fundamental space harmonic of the slow-wave structure described in connection with Figure 8.7-10. The impedance is a function of the point of operation; we shall calculate it for $\beta L = \pi/4$. Assume a very thin, cylindrical beam which just grazes the top of the fingers of the structure, shown in Figure 8.7-4. Assuming that the structure is 1 cm wide, Figure 8.7-10 gives

$$P = 0.17 \text{ mw}$$

for 1 volt from one vane tip to the next, or

$$AL \sin kh = 1 \text{ volt}$$

The magnitude of the fundamental space harmonic at the tips of the vanes is obtained from Equation (8.7-37) as

$$|E_{zo}| = A \sin kh \frac{\sin \frac{\pi}{8}}{\frac{\pi}{8}} = \frac{1}{L} \frac{\sin \frac{\pi}{8}}{\frac{\pi}{8}} \text{volt}$$

$$= 0.974 \frac{1}{L} \text{ volt}$$

Equation (10.1-19) for a thin beam becomes

$$K_o = \frac{|E_{zo}|^2}{2\beta_o{}^2 P} = \frac{(0.974)^2}{2\left(\dfrac{\pi}{4}\right)^2 (0.00017)}$$

$$= 4520 \text{ ohms}$$

For a thick beam the impedance would be lower, since $E_{zo}$ is strongest at the vane tips.

From Equation (10.1-19), we can define the *instantaneous power flow* along the slow-wave structure as

$$p = \frac{\mathcal{E}_{zn}{}^2}{\beta_n{}^2 K_n} \qquad (10.1\text{-}20)$$

where $\mathcal{E}_{zn}$ is the instantaneous value of the space harmonic averaged over the beam cross section as in Equation (10.1-19). Differentiating this expression, we obtain

$$dp = \frac{2\mathcal{E}_{zn} d\mathcal{E}_{zn}}{\beta_n{}^2 K_n} \qquad (10.1\text{-}21)$$

This equation expresses the change in instantaneous power flow due to an increment in the instantaneous value of the space harmonic.[5]
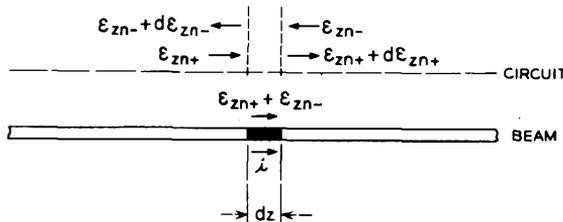


FIG. 10.1-1 Interaction of a short segment of convection current with the rf electric fields due to a slow-wave structure. The arrows above the circuit indicate directions of propagation.

Next consider Figure 10.1-1. The beam is divided up into short segments of length $dz$. Consider a typical segment in which the convection current has the instantaneous value $i$. This current segment induces currents in the circuit. These circuit currents give rise to electromagnetic waves which propagate in both directions from the point on the circuit adjacent to the

---

[5]For those troubled by the definition of Equation (10.1-20), an alternative derivation of Equation (10.1-21) consists of computing the instantaneous change in electric stored energy and applying Equation (8.5-26).

current segment, Let $\mathcal{E}_{zn-}$ and $\mathcal{E}_{zn+}$ be the instantaneous values of the $n^{th}$ order space harmonics moving to the left and to the right, respectively, at this point on the circuit. The induced currents due to this segment of convection current causes $\mathcal{E}_{zn-}$ to change by $d\mathcal{E}_{zn-}$ and $\mathcal{E}_{zn+}$ to change by $d\mathcal{E}_{zn+}$. Because of symmetry, we can say that

$$d\mathcal{E}_{zn-} = d\mathcal{E}_{zn+} \qquad (10.1\text{-}22)$$

in other words, the incremental changes in amplitude of the waves traveling in opposite directions from the point of induction are equal.

The change in instantaneous power flow for each direction is given by Equation (10.1-21). Therefore, the total instantaneous transfer of power to the circuit is given by:

$$\begin{aligned} dp &= dp_- + dp_+ \\ &= \frac{2}{\beta_n{}^2 K_n}[\mathcal{E}_{zn-}d\mathcal{E}_{zn-} + \mathcal{E}_{zn+}d\mathcal{E}_{zn+}] \\ &= \frac{2}{\beta_n{}^2 K_n}(\mathcal{E}_{zn-} + \mathcal{E}_{zn+})d\mathcal{E}_{zn-} \\ &= \frac{2}{\beta_n{}^2 K_n}(\mathcal{E}_{zn-} + \mathcal{E}_{zn+})d\mathcal{E}_{zn+} \qquad (10.1\text{-}23) \end{aligned}$$

where use has been made of Equation (10.1-22).

The incremental instantaneous power flow into the circuit from the length $dz$ of beam is given by[6]

$$dp = -i\mathcal{E}_z dz \qquad (10.1\text{-}24)$$

where $i$ is the instantaneous value of the convection current, and $\mathcal{E}_z$ is the total electric field resulting from the wave on the slow-wave circuit. As discussed previously, all space harmonics are neglected, insofar as they interact with the beam, except the one in synchronism with the beam. Thus, in our case, Equation (10.1-24) becomes

$$dp = -i(\mathcal{E}_{zn-} + \mathcal{E}_{zn+})dz \qquad (10.1\text{-}25)$$

for the convection current segment of Figure 10.1-1. Equations (10.1-23) and (10.1-25) can be combined to yield

$$d\mathcal{E}_{zn-} = d\mathcal{E}_{zn+} = -\tfrac{1}{2}\beta_n{}^2 K_n i\,dz \qquad (10.1\text{-}26)$$

This may be written in the phasor notation as

$$dE_{zn-} = dE_{zn+} = -\tfrac{1}{2}\beta_n{}^2 K_n i\,dz \qquad (10.1\text{-}27)$$

Thus, we have an expression for the incremental waves propagating away from a point on the circuit in terms of the convection current which induces them.

---

[6]This expression may be derived directly from Maxwell's Equations. See Reference 10.1.

Let $\Gamma_o = \alpha + j\beta_n$ be the complex propagation constant for the $n^{th}$ space harmonic.

Let us write an expression for the total space-harmonic field at an arbitrary point $(z = a)$. This is the sum of three contributions, as follows:

A.  The power coming from the input waveguide causes a field $E_{zn}(0)$ at the beginning of the circuit $(z = 0)$. Thus we have at $z = a$,

$$E_{znA}(a) = E_{zn}(0)\epsilon^{-\Gamma_o a} \qquad (10.1\text{-}28)$$

B.  The superposition of the incremental waves $dE_{zn+}$ arriving at $z = a$ from the left is given by

$$E_{znB}(a) = \int_0^a \epsilon^{-\Gamma_o(a-z)} dE_{zn+} \qquad (10.1\text{-}29)$$

C.  The superposition of the incremental waves $dE_{zn-}$ arriving from the right is given by

$$E_{znC}(a) = \int_a^l \epsilon^{-\Gamma_o(z-a)} dE_{zn-} \qquad (10.1\text{-}30)$$

where $l$ is the total circuit length. The total field at $z = a$ is given by the sum of these three contributions

$$E_{zn}(a) = E_{zn}(0)\epsilon^{-\Gamma_o a} - \frac{1}{2}\beta_n^2 K_n \int_0^a i\epsilon^{-\Gamma_o(a-z)} dz - \frac{1}{2}\beta_n^2 K_n \int_a^l i\epsilon^{-\Gamma_o(z-a)} dz \qquad (10.1\text{-}31)$$

where use has been made of Equation (10.1-27). We can replace the variable of integration in the definite integrals by $\tau$ and we replace $a$ by $z$, an arbitrary point, obtaining

$$E_{zn}(z) = E_{zn}(0)\epsilon^{-\Gamma_o z} - \frac{1}{2}\beta_n^2 K_n \int_0^z i(\tau)\epsilon^{-\Gamma_o(z-\tau)} d\tau - \frac{1}{2}\beta_n^2 K_n \int_z^l i(\tau)\epsilon^{\Gamma_o(z-\tau)} d\tau \qquad (10.1\text{-}32)$$

Thus we have an integral equation relating the field at any point to the convection current on the beam.

Equation (10.1-32) is readily converted to a differential equation by differentiation with respect to $z$. In differentiating we must be careful to handle the integrals correctly, since the limits are functions of $z$[7]. Two suc-

---

[7]The following formula must be used: If

$$I(z) = \int_{\alpha(z)}^{\beta(z)} F(\tau, z) d\tau$$

then

$$\frac{dI}{dx} = \int_\alpha^\beta \frac{\partial F}{\partial z} d\tau + F(\beta, z)\frac{d\beta}{dz} - F(\alpha, z)\frac{d\alpha}{dz}$$

See Reference 10.2, p. 353, or any other book on advanced calculus.

cessive differentiations of Equation (10.1-32) yield

$$\Gamma^2 E_{zn} = \Gamma_o^2 E_{zn}(0)\epsilon^{-\Gamma_o z} - \frac{1}{2}\Gamma_o^2\beta_n^2 K_n \int_0^z i\epsilon^{-\Gamma_o(z-\tau)}d\tau$$

$$- \frac{1}{2}\Gamma_o^2\beta_n^2 K_n \int_z^l i\epsilon^{\Gamma_o(z-\tau)}d\tau + \Gamma_o\beta_n^2 K_n i \qquad (10.1\text{-}33)$$

where the left-hand side results from two successive applications of Equation (10.1-3). This result may be combined with Equation (10.1-32) yielding

$$E_{zn} = \frac{\Gamma_o\beta_n^2 K_n i}{\Gamma^2 - \Gamma_o^2} \qquad (10.1\text{-}34)$$

This equation is called the *circuit equation*, since it determines how the circuit field is affected by convection current in the beam.

### (c) Solutions for Cumulative Interaction

The *electronic equation* and the *circuit equation* have been derived on the assumption that the various fields and beam quantities have a z dependence of the form

$$\epsilon^{-\Gamma z}$$

So far we have not said anything about values for $\Gamma$. It turns out that only certain values of $\Gamma$ are allowable when we require that the *circuit* and *electronic equations* be simultaneously satisfied.

Upon solving for the ratio $E_{zn}/i$ in both Equations (10.1-18) and (10.1-34) and equating the results, we obtain

$$(\Gamma^2 - \Gamma_o^2)\left[(\Gamma - j\beta_e)^2 + \frac{\omega_q^2}{u_o^2}\right] = \frac{j\beta_e\beta_n^2\Gamma_o K_n I_o}{2V_o} \qquad (10.1\text{-}35)$$

the solution of which determines the allowed values of $\Gamma$.

We can put this equation in a neater form by defining certain parameters.[8] First, we define the small-signal gain parameter $C$ by the equation

$$C^3 \equiv \frac{K_n I_o}{4V_o} \qquad (10.1\text{-}36)$$

$C$ is a small dimensionless parameter with values usually in the range 0.01 to 0.1.

Next, we define the space-charge parameter $QC$ by[9]

$$QC \equiv \frac{\omega_q^2}{4C^2\omega^2} \qquad (10.1\text{-}37)$$

---

[8]Reference 10a, Chapter 7.

[9]For large space charge, $QC$ is computed from the more exact expression $\sqrt{4QC^3} = \dfrac{\omega_q/\omega}{1 + (\omega_q/\omega)}$; see Reference 10.4.

Since $\beta_e$ and $\beta_n$ are approximately equal for synchronism, and since $\Gamma$ must have a similar imaginary part, we define the dimensionless parameters $b$, $d$, and $\delta$ by the equations

$$\Gamma_o \equiv j\beta_e(1 + Cb - jCd) \qquad (10.1\text{-}38)$$

and

$$\Gamma \equiv j\beta_e(1 + jC\delta) \qquad (10.1\text{-}39)$$

Since $\Gamma_o = \alpha + j\beta_n$, we see that $b$ and $d$ are real numbers given by

$$b = \frac{\beta_n - \beta_e}{\beta_e C} = \frac{u_o - v_{pn}}{v_{pn}C} \qquad (10.1\text{-}40)$$

and

$$d = \frac{\alpha}{\beta_e C} \qquad (10.1\text{-}41)$$

$b$ is a measure of the amount of synchronism between the electrons and the space harmonic. $d$ is directly proportional to the circuit attenuation. $\delta$ is a complex number, values of which are determined by solving Equation (10.1-35).

Equations (10.1-36) to (10.1-39) are substituted into (10.1-35). In so doing, advantage is taken of the fact that $C$ is small; that is, terms of order $C$ are neglected in comparison with terms of order unity. After simplification, we obtain

$$\delta^2 = \frac{1}{(-b + jd + j\delta)} - 4QC \qquad (10.1\text{-}42)$$

where we have used the approximation

$$\frac{\Gamma_o \beta_n^2}{j\beta_e^3} \cong 1 \qquad (10.1\text{-}43)$$

Equation (10.1-42) is a cubic equation in the unknown $\delta$. The solutions, when introduced into Equation (10.1-39), give us three allowable values for $\Gamma$, the complex propagation constant for the circuit-beam coupled system.[10] From Equation (10.1-42), we note that the solutions are functions of $b$ (the degree of synchronism), $d$ (the circuit attenuation), and $QC$ (the space charge).

From Equation (10.1-39), we note that the difference between $\Gamma$ and $j\beta_e$ is directly proportional to $C$, which from Equation (10.1-36) is a function of

---

[10]Since Equation (10.1-35) is a quartic equation, we see that one root has been lost due to the approximations we have made. This root corresponds to a backward propagating wave, which is nonsynchronous with the beam and which is not excited for proper termination of the circuit.

the amount of coupling between the beam and the circuit. The stronger the coupling (the larger $K_n$ is), the greater is this difference, as we should expect for two coupled systems.

Let us examine the nature of the solutions for the simple case:

$$b = d = QC = 0$$

Equation (10.1-42) reduces to

$$\delta^3 = -j \qquad (10.1\text{-}44)$$

with solutions

$$\delta_1 = \frac{\sqrt{3}}{2} - j\frac{1}{2}$$

$$\delta_2 = -\frac{\sqrt{3}}{2} - j\frac{1}{2}$$

$$\delta_3 = j \qquad (10.1\text{-}45)$$

The values of $\Gamma$ are then given by

$$\Gamma_1 = -\frac{\sqrt{3}}{2}\beta_e C + j\beta_e\left(1 + \frac{C}{2}\right)$$

$$\Gamma_2 = \frac{\sqrt{3}}{2}\beta_e C + j\beta_e\left(1 + \frac{C}{2}\right)$$

$$\Gamma_3 = j\beta_e(1 - C) \qquad (10.1\text{-}46)$$

Since the various fields and beam quantities propagate as

$$\epsilon^{j\omega t - \Gamma z}$$

we see that $\Gamma_1$ corresponds to a traveling wave whose amplitude grows exponentially with distance, $\Gamma_2$ to one whose amplitude decays exponentially with distance, and $\Gamma_3$ to a wave of constant amplitude. The wave corresponding to $\Gamma_1$ is termed the *growing wave*, and it is this wave which is responsible for the gain in the traveling-wave tube. We shall find that, regardless of the values of $d$ and $QC$, we shall always obtain one growing wave, one decaying wave, and one wave of nearly constant amplitude, as long as $b$ corresponds to operation sufficiently close to synchronism.[11]

We can examine the condition of synchronism for this example. It was pointed out in connection with Figure 10-2 that the electrons must travel slightly faster than the wave so that the bunch drifts into a retarding field. It is clear that the growing wave will predominate as the electrons move down the tube. Thus, the circuit field depicted in Figure 10-2 near the end

---

[11]$d$, of course, must not be so large as to neutralize the gain of the growing wave.

of the tube has $\Gamma_1$ for its propagation constant, corresponding to a phase velocity given by

$$v_{p1} = \frac{\omega}{\beta_e(1 + \frac{1}{2}C)} = \frac{u_o}{1 + \frac{1}{2}C} \qquad (10.1\text{-}47)$$

which is indeed slightly slower than the dc beam velocity. Thus, although the cold (no electrons) phase velocity $v_{pn}$ is exactly equal to $u_o$ from Equation (10.1-40), the hot phase velocity of the growing wave $v_{p1}$ is slower than $u_o$. This must be true regardless of the values of $b$, $d$, and $QC$ if we are to obtain amplification.

As a second example, we consider the condition of large space charge. Let the circuit attenuation be zero ($d = 0$), for simplicity. Under the condition that $QC$ is large, we may neglect the first term on the right-hand side of Equation (10.1-42) in comparison with the second term, obtaining

$$\delta^2 \cong -4QC \qquad (10.1\text{-}48)$$

with the solutions

$$\delta_1, \delta_3 \cong \pm j2\sqrt{QC} \qquad (10.1\text{-}49)$$

From Equation (10.1-39), we obtain

$$\Gamma_1, \Gamma_3 \cong j\beta_e(1 \pm 2C\sqrt{QC})$$
$$\cong j\beta_e\left(1 \pm \frac{\omega_q}{\omega}\right) \qquad (10.1\text{-}50)$$

for the growing wave and constant-amplitude wave, respectively. Referring back to Equations (9.3-40) and (9.3-41), we see that these propagation constants correspond to the slow and fast space-charge waves, respectively. The approximate form of Equation (10.1-48) does not allow computation of the real parts associated with these waves. Furthermore, the decaying wave propagation constant $\Gamma_2$ is not evaluated at all. Numerical calculations can be used to derive this information directly from Equation (10.1-42).[12] The result of such a calculation is given in Figure 10.1-2, for $QC = 1$. The real and imaginary parts of $\delta$ are defined by

$$\delta = x + jy \qquad (10.1\text{-}51)$$

We note from this figure that a growing wave (positive values of $x$) is obtained for $b$ approximately between 0.9 and 2.8, with a maximum value of $x_1$ equal to 0.5. The value of $b$ corresponding to this maximum value of $x_1$ determines the electron velocity for maximum gain. A useful approximate formula for the maximum value of $x_1$ as a function of $QC$ is[13]

$$x_{1\,\text{max}} \cong \frac{1}{2(QC)^{1/4}} \qquad (10.1\text{-}52)$$

---

[12] Reference 10a, Chapter 8.
[13] Reference 10.3.
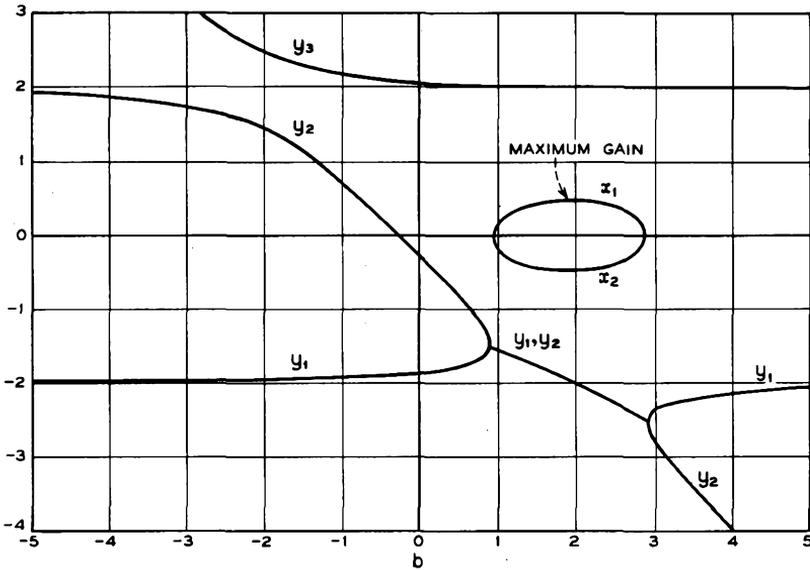
Fig. 10.1-2 Real and imaginary parts of the three solutions for $\delta$ for a traveling-wave amplifier with zero loss and $QC = 1$. (From John R. Pierce, *Traveling Wave Tubes*, D. Van Nostrand Co., Inc., Princeton, N. J., 1950)

Having determined the three allowed values of the propagation constant $\Gamma$, which correspond to waves propagating in the forward direction, one can write an expression for the rf components of the total electric field, the velocity, and the convection current as

$$E_{zT} = E_{zT1}\epsilon^{-\Gamma_1 z} + E_{zT2}\epsilon^{-\Gamma_2 z} + E_{zT3}\epsilon^{-\Gamma_3 z}$$

$$u = u_1\epsilon^{-\Gamma_1 z} + u_2\epsilon^{-\Gamma_2 z} + u_3\epsilon^{-\Gamma_3 z}$$

$$i = i_1\epsilon^{-\Gamma_1 z} + i_2\epsilon^{-\Gamma_2 z} + i_3\epsilon^{-\Gamma_3 z} \qquad (10.1\text{-}53)$$

The values of the component waves can be found from the initial conditions at $z = 0$, the input end of the slow-wave circuit. Since the beam is unmodulated at this point, we have

$$u(0) = 0$$

$$i(0) = 0 \qquad (10.1\text{-}54)$$

For the same reason, the rf space-charge field is zero and the total electric field is given by the circuit voltage alone,

$$E_{zT}(0) = E_{zn}(0) \qquad (10.1\text{-}55)$$

The latter quantity may be determined in magnitude from the rf input power using Equation (10.1-19), which can be written as

$$| E_{zn} |^2 = 2\beta_n^2 K_n P \tag{10.1-56}$$

where $| E_{zn} |$ is the rms average over the beam cross section. We assume all the quantities $u$, $i$, etc., are similar rms averages.

Before applying the above boundary conditions, we derive certain other relationships between $E_{zT}$, $u$, and $i$. First, we introduce Equation (10.1-39) into (10.1-13), obtaining

$$u_\nu = -\frac{e}{m\omega C \delta_\nu} E_{zT\nu} \tag{10.1-57}$$

for $\nu = 1,2,3$, that is, for each of the three waves. Next, we introduce Equation (10.1-39) into (10.1-9), obtaining

$$i_\nu = -\frac{j\rho_o S}{C \delta_\nu} u_\nu \tag{10.1-58}$$

which with Equation (10.1-57) becomes

$$i_\nu = \frac{j e \rho_o S}{m\omega C^2 \delta_\nu^2} E_{zT\nu} \tag{10.1-59}$$

Equations (10.1-57) and (10.1-59) may be substituted into Equations (10.1-53) so that the total electric field is the only variable. When this is done, and the boundary conditions (10.1-54) and (10.1-55) are applied, Equations (10.1-53) become

$$E_{zn}(0) = E_{zT1} + E_{zT2} + E_{zT3}$$

$$0 = \frac{1}{\delta_1} E_{zT1} + \frac{1}{\delta_2} E_{zT2} + \frac{1}{\delta_3} E_{zT3}$$

$$0 = \frac{1}{\delta_1^2} E_{zT1} + \frac{1}{\delta_2^2} E_{zT2} + \frac{1}{\delta_3^2} E_{zT3} \tag{10.1-60}$$

Since the $\delta$'s are known, these three simultaneous equations may be solved to obtain solutions for $E_{zT1}$, $E_{zT2}$, and $E_{zT3}$ in terms of $E_{zn}(0)$. The solution for the growing wave is given by

$$\frac{E_{zT1}}{E_{zn}(0)} = \frac{\delta_1^2}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)} \tag{10.1-61}$$

We define the initial loss factor $A_1$ in decibels by the expression

$$A_1 = 20 \log \left| \frac{E_{zT1}}{E_{zn}(0)} \right| \text{ db} \tag{10.1-62}$$

which can be evaluated from Equation (10.1-61).

The total power gain of the traveling-wave amplifier may be written as

$$\text{gain} = 20 \log \left| \frac{E_{zn}(l)}{E_{zn}(0)} \right| \text{db} \qquad (10.1\text{-}63)$$

where $l$ denotes the output end of the slow-wave structure. For a practical tube where the gain is 20 db or more, the fields at the output end of the structure are given very accurately by the growing wave alone, so we may write

$$\text{gain} = 20 \log \left| \frac{E_{zn1}\epsilon^{-\Gamma_1 l}}{E_{zn}(0)} \right| \text{db} \qquad (10.1\text{-}64)$$

This can be written as

$$\text{gain} = 20 \log \left| \frac{E_{zT1}}{E_{zn}(0)} \frac{E_{zn1}}{E_{zT1}} \epsilon^{-\Gamma_1 l} \right| \text{db} \qquad (10.1\text{-}65)$$

Since

$$\left| \epsilon^{-\Gamma_1 l} \right| = \epsilon^{\beta_e C x_1 l} \qquad (10.1\text{-}66)$$

Equation (10.1-65) becomes

$$\text{gain}_{db} = A_1 + 20\beta_e C x_1 l \log \epsilon + A_2 \qquad (10.1\text{-}67)$$

where

$$A_2 = 20 \log \left| \frac{E_{zn1}}{E_{zT1}} \right| \text{db} \qquad (10.1\text{-}68)$$

The quantity $A_2$ is termed the space-charge loss factor, since it is a measure of the ratio of the circuit field $E_{zn}$ to the total field, at the end of the circuit where the growing wave predominates. We further define

$$B = (40\pi \log \epsilon)x_1 = 54.6x_1 \qquad (10.1\text{-}69)$$

and the number of electronic wavelengths on the circuit $N$ by

$$N = \frac{\beta_e l}{2\pi} \qquad (10.1\text{-}70)$$

so that Equation (10.1-67) for the gain of a traveling-wave amplifier becomes

$$\text{gain}_{db} = A_1 + A_2 + BCN \qquad (10.1\text{-}71)$$

Finally, we must evaluate $A_2$, the space-charge loss factor. Equation (10.1-59) may be written for the growing wave as

$$i_1 = \frac{je\rho_o S}{m\omega C^2 \delta_1{}^2} E_{zT1} \qquad (10.1\text{-}72)$$

This equation is combined with Equation (10.1-17) written for the growing wave to eliminate $i_1$. After simplifying and introducing the definition of $QC$, Equation (10.1-37), one obtains

$$\frac{E_{zn1}}{E_{zT1}} = \frac{\delta_1{}^2 + 4QC}{\delta_1{}^2} \qquad (10.1\text{-}73)$$

recalling that $E_{zT1} = E_{zn1} + E_{zsc1}$. From Equation (10.1-68), $A_2$ is thus given by

$$A_2 = 20 \log \left| \frac{\delta_1{}^2 + 4QC}{\delta_1{}^2} \right| \text{ db} \qquad (10.1\text{-}74)$$

In conclusion, we write an expression for the gain when $b = d = QC = 0$. From Equations (10.1-45), we calculate

$$A_1 = -9.54 \text{ db}$$
$$A_2 = 0 \text{ db}$$
$$B = 47.3$$

so that Equation (10.1-71) becomes

$$\text{gain}_{db} = -9.54 + 47.3CN \qquad (10.1\text{-}75)$$

### (d) Summary of Gain Calculation

Let us review the steps in calculating the gain of a traveling-wave amplifier. Knowing the properties of the beam and the circuit, we calculate $C$, $N$, $b$, $d$, and $QC$ as follows:

$$C = \left[ \frac{K_n I_o}{4V_o} \right]^{1/3} \qquad (10.1\text{-}36)$$

$$N = \frac{\beta_e l}{2\pi} = \frac{u_o l}{2\pi\omega} \qquad (10.1\text{-}70)$$

$$b = \frac{\beta_n - \beta_e}{\beta_e C} \qquad (10.1\text{-}40)$$

$$d = \frac{\alpha}{\beta_e C} \qquad (10.1\text{-}41)$$

$$QC = \frac{\omega_q{}^2}{4C^2\omega^2} \qquad (10.1\text{-}37)$$

$K_n$ is defined as in Equation (10.1-19). $\delta_1$, $\delta_2$, and $\delta_3$ are determined as the three roots of the cubic equation:

$$\delta^2 = \frac{1}{(-b + jd + j\delta)} - 4QC \qquad (10.1\text{-}42)$$

Next, $A_1$ and $A_2$ are calculated from the equations:

$$A_1 = 20 \log \left| \frac{\delta_1^2}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)} \right| \text{db} \qquad (10.1\text{-}62)$$

and

$$A_2 = 20 \log \left| \frac{\delta_1^2 + 4QC}{\delta_1^2} \right| \text{db} \qquad (10.1\text{-}74)$$
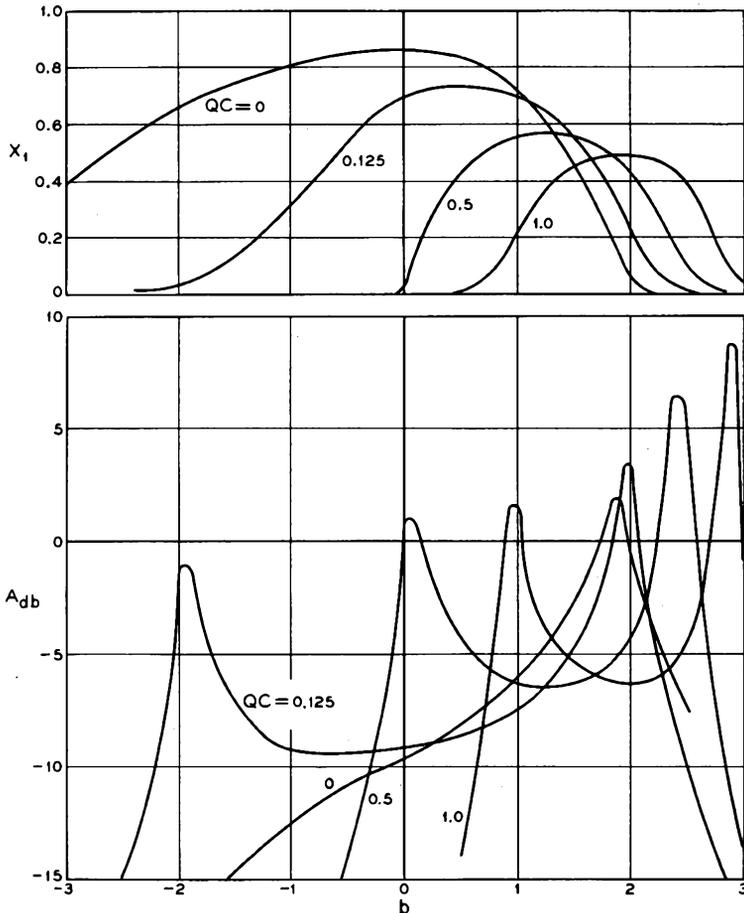


FIG. 10.1-3 Graphs useful in calculating the gain of a traveling-wave amplifier. $d = 0.025$. $C = 0$. Curves for non-zero values of $C$ are given in Reference 10.4. (Courtesy of *Transactions IRE*)

Finally, the gain is computed from the equation:

$$\text{gain}_{db} = A_1 + A_2 + BCN \qquad (10.1\text{-}71)$$

where

$$B = 54.6x_1 \qquad (10.1\text{-}69)$$

Considerable effort can be saved by using charts which give $x_1$ and $A = A_1 + A_2$ directly in terms of $b$, $d$, and $QC$.[14] Such a chart is shown in Figure 10.1-3, where $x_1$ and $A$ are plotted vs. $b$ for $d = 0.025$ and various values of $QC$. By these methods one can calculate the gain of a traveling-wave amplifier and the variation with frequency.

## 10.2  High-Power Tubes

The theory developed in the previous section enables one to calculate the performance of a traveling-wave amplifier once the properties of the beam and the slow-wave structure are known. We must know two things about the slow-wave structure — the Brillouin diagram and the strength of coupling to the electron beam. The latter quantity is measured in terms of the
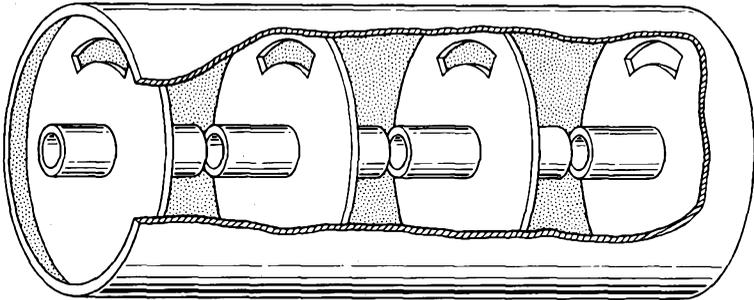


FIG. 10.2-1 High-power slow-wave structure consisting of a cascade of re-entrant cavity resonators, with mutual inductive coupling obtained through apertures in their common walls.

beam-coupling impedance $K_n$. In this section we examine some of the techniques used to evaluate these quantities for a typical high-power structure.

### (a)  The Brillouin Diagram

The slow-wave structure of Figure 10-1(c) is well suited for high-power interaction. The large metal surfaces in this structure serve the dual pur-

---

[14]Reference 10.4. Curves for $C = 0.2$ are in error. Corrected curves are given in Reference 10.5, Figure 4(c).

pose of keeping rf ohmic losses small and also of providing means for
removing the heat due to beam interception on the re-entrant parts of the
cavities. A simplified version of this structure is presented in Figure 10.2-1.
In this structure, the coupling between cavities is obtained by means of
coupling apertures placed in the inductive portions of the cavities. These
apertures allow portions of the magnetic flux of one cavity to link the
adjacent cavities, so as to provide mutual coupling between cavities.

An equivalent circuit for this structure is shown in Figure 10.2-2. $L_1$ and
$C_1$ are the inductance and capacitance of each re-entrant cavity, as dis-
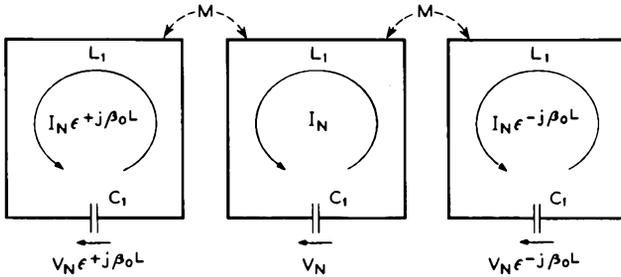


FIG. 10.2-2 Equivalent circuit for the slow-wave structure of Figure 10.2-1.

discussed in Chapter 8. The inductance is shown as a single turn. $M$ is the
mutual inductance between cavities; obviously it becomes larger as one in-
creases the size of the coupling aperture. Loop currents are shown in each of
the cavities. From Floquet's Theorem we know that, if the current in the
$N^{th}$ cavity is $I_N$, the currents in the $N-1$ and $N+1$ cavities are given by

$$I_N \epsilon^{+j\beta_o L}$$

and

$$I_N \epsilon^{-j\beta_o L}$$

respectively.

From Kirchhoff's voltage law, we can write the loop equation for the $N^{th}$
loop as

$$\left(j\omega L_1 + \frac{1}{j\omega C_1}\right)I_N - j\omega M I_N \epsilon^{j\beta_o L} - j\omega M I_N \epsilon^{-j\beta_o L} = 0 \qquad (10.2\text{-}1)$$

This may be simplified to yield

$$\cos \beta_o L = \frac{1}{k_1}\left(1 - \frac{\omega_c^2}{\omega^2}\right) \qquad (10.2\text{-}2)$$

where $k_1 = 2M/L_1$ is the coupling coefficient, and $\omega_c = 1/\sqrt{L_1 C_1}$ is the

radian resonant frequency of the cavity. The upper and lower cutoff frequencies are obtained from Equation (10.2-2) as

$$\omega_2 = \frac{\omega_c}{\sqrt{1 - k_1}} \quad \text{at} \quad \beta_o L = 0$$

and

$$\omega_1 = \frac{\omega_c}{\sqrt{1 + k_1}} \quad \text{at} \quad \beta_o L = \pi \qquad (10.2\text{-}3)$$

For $k_1 = 0$, the Brillouin diagram is a horizontal line given by $\omega = \omega_c$. Increasing the coupling increases the width of the passband.

The Brillouin diagram is shown in Figure 10.2-3 for the fundamental and $-1$ space harmonics for a value of $k_1$ equal to 0.4, as obtained from Equation (10.2-2). The fundamental is a backward wave, since its group and
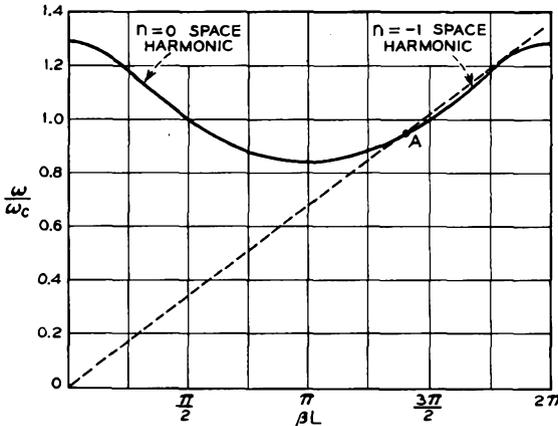


Fig. 10.2-3 Brillouin diagram for the slow-wave structure of Figure 10.2-1, for $k_1 = 0.4$. A voltage line is shown corresponding to typical interaction with the $-1$ space harmonic for forward wave interaction.

phase velocities are in opposite directions. In the traveling-wave amplifier a forward wave is necessary for amplification, since the power on the circuit travels in the same direction as the electron beam. The $-1$ space harmonic is such a wave. The slope of the line joining the origin to point $A$ gives the phase velocity of the space harmonic at the point $A$. Since this line is nearly tangent to the $-1$ space-harmonic curve, the phase velocity is nearly constant over a moderately wide frequency range. Thus, with the dc beam velocity equal to this phase velocity, synchronous interaction is obtainable over a considerable frequency range.

### (b)  The Beam Interaction Impedance

In calculating the beam interaction impedance, we shall need an expression for the total power flow in the circuit. This could be obtained directly from the equivalent circuit using network theory. However, it is easier to obtain the power from the group velocity and the average stored energy per period, as in Section 8.7(d). The average stored energy is equal to the peak electric stored energy,

$$W_L = \tfrac{1}{2}C_1 \,|\,V_N\,|^2 \qquad (10.2\text{-}4)$$

where $V_N$ is the voltage across the $N^{\text{th}}$ gap. The group velocity of the fundamental is obtained by means of a differentiation of Equation (10.2-2) as

$$v_g = \frac{d\omega}{d\beta_o} = -\frac{k_1 L}{2}\frac{\omega^3}{\omega_c^2}\sin\beta_o L \qquad (10.2\text{-}5)$$

All the space harmonics have the same magnitude of group velocity at the same frequency. From Equation (8.7-49) the power flow in the forward direction is thus given by

$$P = \frac{k_1\omega^3 C_1}{4\omega_c^2}\,|\,V_N\,|^2 \sin\beta_o L \qquad (10.2\text{-}6)$$

In order to evaluate the impedance by Equation (10.1-19) we need an expression for the amplitude of the space harmonic. This is obtained by solving Maxwell's Equations in the region of the beam.

Consider Figure 10.2-4. A series of concentric cylinders of inside radii $a$ uniformly spaced along the $z$ axis is shown. The periodic spacing is $L$, and the gap between cylinders is $\delta$. The voltage across gap $N$ is $V_N$, and we assume that the electric field across the gap at $r = a$ is uniform and equal to $V_N/\delta$.
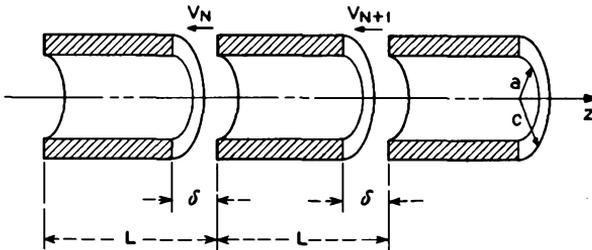


Fig. 10.2-4 Identical, concentric cylinders uniformly spaced along the $z$ axis. The gap voltages obey Floquet's Theorem. The internal fields are analyzed in terms of space harmonics.

We confine our attention to the region $r \leq a$. The general solution for the electric field in this region is given by Equation (8.7-16), for which the $z$ component is

$$E_z(x,y,z) = \sum_{-\infty}^{\infty} E_{zn}(x,y)\epsilon^{-j\beta_n z} \qquad (10.2\text{-}7)$$

As shown in Section 8.7, each space harmonic will by itself satisfy the wave Equation (8.1-26). The $z$ component of the wave equation for the $n^{\text{th}}$ space harmonic can be written as

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)E_{zn} - (\beta_n{}^2 - k^2)E_{zn} = 0 \qquad (10.2\text{-}8)$$

similar to Equation (8.7-27). Since the structure in Figure 10.2-4 has cylindrical symmetry about the $z$ axis, it will be more convenient to write the above equation in cylindrical coordinates. The two-dimensional Laplacian can be written as

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} \qquad (10.2\text{-}9)$$

in polar coordinates.[16] Since the boundary conditions in Figure 10.2-4 have no functional dependence on $\theta$, we look for space harmonics which also are independent of $\theta$. The last term in Equation (10.2-9) is thus set equal to zero. Equation (10.2-8) becomes

$$\left[\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} - \gamma_n{}^2\right]E_{zn} = 0 \qquad (10.2\text{-}10)$$

where $\gamma_n{}^2 = \beta_n{}^2 - k^2$ and is positive for slow waves. $k$ is equal to $\omega/c$.

Equations (10.2-10) may be written in a form recognizable as Bessel's Equation,

$$\left[\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \tau_n{}^2\right]E_{zn} = 0 \qquad (10.2\text{-}11)$$

with $\tau_n{}^2 = -\gamma_n{}^2$. The solution of this equation is well known:

$$E_{zn} = B_n J_o(\tau_n r) \qquad (10.2\text{-}12)$$

where $J_o$ is the Bessel function of the first kind of order 0, and $B_n$ is an arbitrary constant. Thus, the solution of Equation (10.2-10) is given by

$$E_{zn} = B_n J_o(j\gamma_n r) \qquad (10.2\text{-}13)$$

This modified Bessel function has a real value despite the fact that the

---

[16]Reference 10.2, p. 328. This may also be derived using the relations of Appendix XII together with the definition of the Laplacian of a scalar, $\nabla^2\Phi = \nabla \cdot \nabla\Phi$.

argument is a pure imaginary quantity. This function is of sufficient importance that it is given its own symbol,

$$J_o(j\gamma_n r) \equiv I_o(\gamma_n r) \qquad (10.2\text{-}14)$$

It is plotted in Figure 10.2-5. From this figure we see that the axial electric field is weakest on the axis and strongest at the circuit. This behavior was also obtained in the slow-wave structure of Figure 8.7-4. In fact, this is a
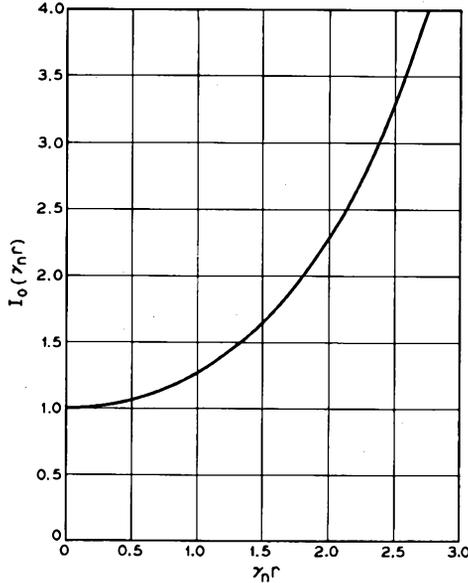


FIG. 10.2-5 Modified Bessel function which gives the radial decay of the axial component of a cylindrically symmetric space harmonic.

general characteristic of slow-wave structures; the axial field is strongest at the circuit. Most traveling-wave amplifiers have a value of $\gamma_n r$ at the circuit in the range of 1 to 2 for the synchronous space harmonic. Thus, the field on the axis may be only half that at the circuit.

Equation (10.2-7) may thus be written as

$$E_z(r,z) = \sum_{-\infty}^{\infty} B_n I_o(\gamma_n r)\epsilon^{-j\beta_n z} \qquad (10.2\text{-}15)$$

The space-harmonic amplitudes $B_n$ are evaluated by imposing the boundary conditions at $r = a$. We take $z = 0$ to correspond to the center of the $N^{\text{th}}$ gap. If we confine our attention to the unit cell centered at $z = 0$, we have

from Equation (10.2-15):

$$\sum_{-\infty}^{\infty} B_n I_o(\gamma_n a) \epsilon^{-j\beta_n z} = \begin{cases} \dfrac{V_N}{\delta} \text{ for } |z| < \dfrac{\delta}{2} \\[2ex] 0 \text{ for } \dfrac{\delta}{2} < |z| < \dfrac{L}{2} \end{cases} \qquad (10.2\text{-}16)$$

Multiplying both sides of this equation by $\epsilon^{+j\beta_m z}$ and integrating over the range $|z| \leq L/2$, we obtain

$$B_m L I_o(\gamma_m a) = \frac{V_N}{\delta} \frac{\sin \beta_m \dfrac{\delta}{2}}{\dfrac{\beta_m}{2}} \qquad (10.2\text{-}17)$$

where use has been made of Equation (8.7-35). The axial electric field is thus given by the expression

$$E_z(r,z) = \frac{V_N}{L} \sum_{-\infty}^{\infty} \frac{\sin \beta_n \dfrac{\delta}{2}}{\beta_n \dfrac{\delta}{2}} \frac{I_o(\gamma_n r)}{I_o(\gamma_n a)} \epsilon^{-j\beta_n z} \qquad (10.2\text{-}18)$$

The impedance for the $n^{\text{th}}$ space harmonic is obtained by introducing Equation (10.2-6) into Equation (10.1-19) together with the space-harmonic amplitude obtained from Equation (10.2-18). One obtains

$$K_n = M_{1(n)}{}^2 M_{2(n)}{}^2 \frac{2\omega_c{}^2}{(\beta_n L)^2 k_1 \omega^3 C_1 \sin \beta_o L} \qquad (10.2\text{-}19)$$

where

$$M_{1(n)} = \frac{\sin \beta_n \dfrac{\delta}{2}}{\beta_n \dfrac{\delta}{2}} \qquad (10.2\text{-}20)$$

is the gap factor for the $n^{\text{th}}$ space harmonic, and

$$M_{2(n)}{}^2 = \frac{1}{S I_o{}^2(\gamma_n a)} \int I_o{}^2(\gamma_n r) dS \qquad (10.2\text{-}21)$$

is the impedance reduction factor obtained by integrating the radial variation of the electric field over the cross section of the beam. Both factors are less than or equal to unity.

The impedance for the $-1$ space harmonic plotted in Figure 10.2-3 is obtained from Equation (10.2-19) by setting $n$ equal to $-1$. The result is plotted in Figure 10.2-6 for $k_1$ and $\delta/L$ both equal to 0.4. The impedance
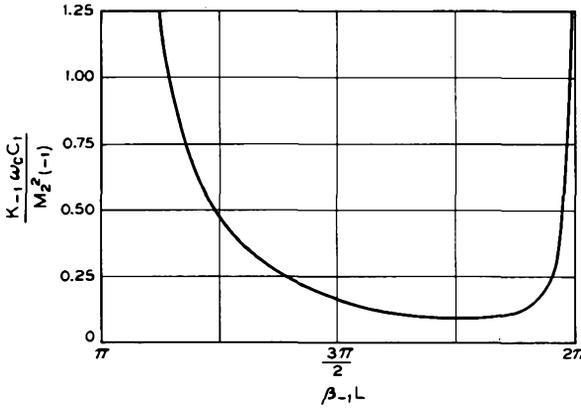
FIG. 10.2-6 Normalized impedance for the $-1$ space harmonic of the slow-wave structure of Figure 10.2-1. $k_1 = \delta/L = 0.4$.

is seen to become infinite at both band edges since the group velocity goes to zero at these frequencies.

### (c) A Numerical Example

As an example we calculate $K_{-1}$ at band center for a typical circuit. Assume that the cylinders shown in Figure 10.2-4 have the following dimensions:

$$a = 2 \text{ mm}$$

$$c = 2.5 \text{ mm}$$

$$\delta = 2 \text{ mm}$$

$$L = 5 \text{ mm}$$

The structure is designed for a center frequency of 6000 Mc with $k_1 = 0.4$. Assume an electron beam of very small diameter on the axis; under this condition,

$$M_{2(-1)}^2 = \frac{1}{I_o^2(\gamma_{-1}a)}$$

Using the relation $\gamma_{-1}^2 = \beta_{-1}^2 - k^2$, we obtain

$$\gamma_{-1}a = 0.992\beta_{-1}a = 0.992\frac{a}{L}\beta_{-1}L = 0.992(0.4)\frac{3\pi}{2} = 1.87$$

From Figure 10.2-5,

$$M_{2(-1)}^2 = \frac{1}{I_o^2(1.87)} = 0.230$$

In calculating the capacitance between cylinders we must make an allowance for fringing effects. It has been found empirically that the capacitance is about four times that given by the simple parallel plate formula for the dimensions given. Thus,

$$C_1 \cong 4\frac{\varepsilon_o 2\pi a(c - a)}{\delta} \cong 0.111 \text{ pf}$$

The ordinate in Figure 10.2-6 is read as 0.1635 at band center. Thus,

$$K_{-1} = 0.1635\frac{M_{2(-1)}{}^2}{\omega_c C_1} = 0.1635\frac{0.230}{(12\pi 10^9)(0.111 \times 10^{-12})} = 9.0 \text{ ohms}$$

This value of impedance is typical of circuits of this type.

With the above information it is necessary only to specify the beam current and structure length to complete the description of the amplifier. The beam voltage is determined by the synchronism condition, $\beta_e = \beta_{-1}$, at some frequency within the band.

Operation as a forward-wave amplifier is also possible in higher-order space harmonics, such as the $-2$, $-3$, $-4$, etc. However, from Equation (10.2-19), it is seen that the impedance falls off for these higher-space harmonics, because $\beta_n L$ is larger and $M_{1(n)}$ and $M_{2(n)}$ are smaller. Therefore, the lowest possible order forward space harmonic is always used so as to obtain maximum interaction. This explains why the structure of Figure 10-1(c) is superior to that of Figure 10-1(b); for the same voltage, the latter corresponds to higher space-harmonic interaction, under the usual condition $\omega_q/\omega \ll 1$.

### (d)   Description of a Practical Power Traveling-Wave Tube

Figure 10.2-7 shows the construction of the Bell Telephone Laboratories type M4040 traveling-wave amplifier.[16] A photograph of the tube is shown in Figure 10.2-8. This tube was designed to be the ground transmitter tube in the *Telstar* experimental satellite communications system. It uses a slow-wave structure similar to the one in Figure 10.2-1, except that the inductive coupling holes appear alternately at the top and bottom of the discs separating the cavities. This is done primarily to prevent inductive coupling between cavities which are not directly adjacent.

The slow-wave structure actually consists of two separate sections, nearly equal in length, placed end to end. Each section has its own rf input and output connections. The input rf signal is introduced onto the first section close to the electron gun. This signal is amplified by 17 db and then coupled out into an external sever termination. The electron bunches

---

[16]Reference 10.6.

PUMPING
TUBULATION

ELECTRON
GUN

LOSSY CERAMICS

RF
OUTPUT

COLLECTOR

HIGH VOLTAGE
TERMINALS

RF
INPUT

SOLENOID

CERAMIC

NON-FERROUS METAL

MAGNETIC METAL
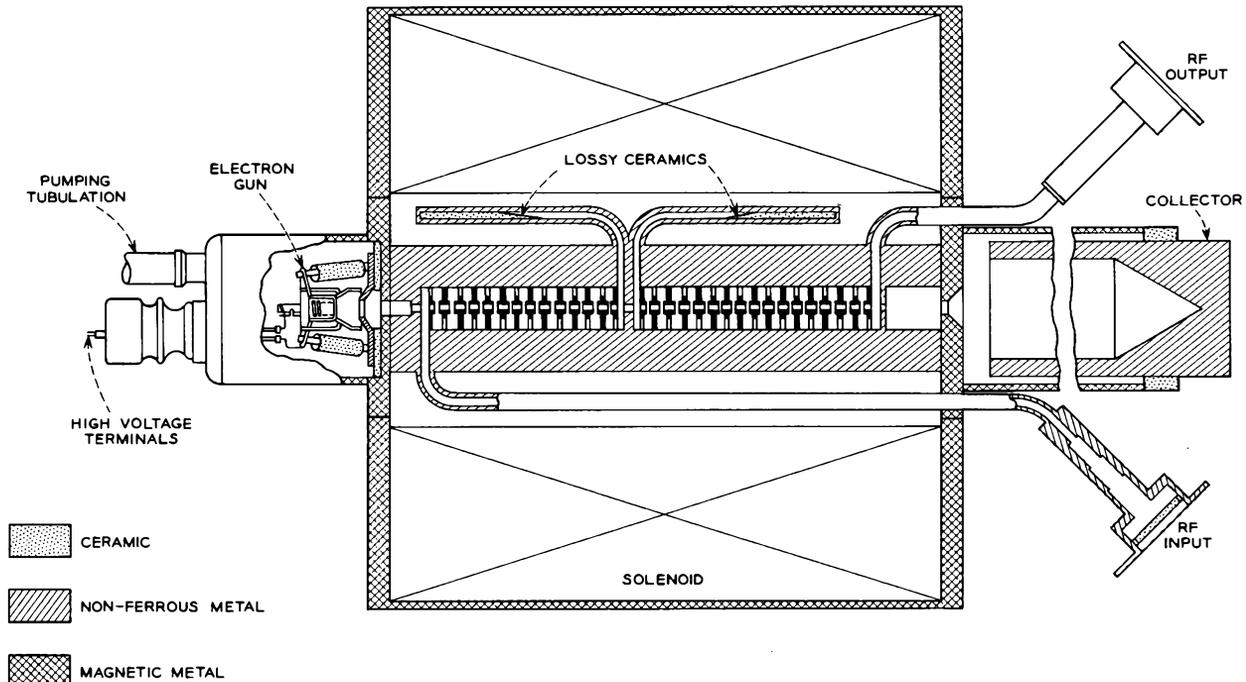
FIG. 10.2-7 The Bell Telephone Laboratories M4040 traveling-wave amplifier, a 2.6-kw, 6000-Mc tube.

in the beam induce currents into the second slow-wave structure section, and another 21 db of gain is produced. The power dissipated in the external sever termination results in an overall gain reduction of approximately 6 db, so that the net gain is 32 db. Without the beam present, the first and second sections of the slow-wave structure are completely isolated from each other. A slow-wave circuit which is terminated at some point near the middle for waves traveling in either direction is called a severed circuit.
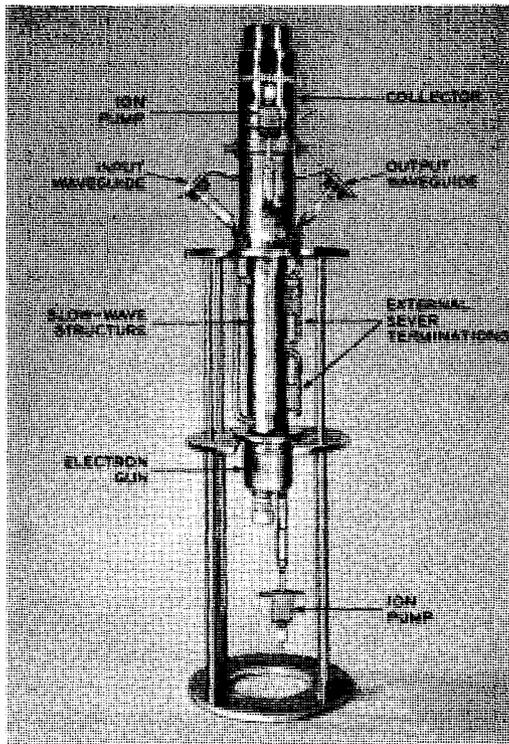


FIG. 10.2-8  The Bell Telephone Laboratories M4040. The overall height is 117 cm.

A severed circuit is necessary in a traveling-wave tube to prevent oscillations. Since it is impossible to match perfectly the input and output waveguides to the slow-wave structure over the entire operating frequency range, small rf reflections are invariably present at both ends of the structure. Without the sever, the wave reflected from the output connection travels back to the input with little attenuation and produces another small reflection. This provides a mechanism for positive feedback, and oscil-

lations result if the gain between the positions of these two reflections is large. Thus, severing the circuit keeps the gain per section low enough to prevent such oscillations.

The M4040 tube has the characteristics given in Table 10.2-1. It is operated in a solenoid which furnishes a uniform magnetic field of 730 oersteds.

TABLE 10.2-1. BTL M4040 OPERATING CHARACTERISTICS

Center frequency, Mc..........................6390
Beam voltage, kv..............................17
Beam current, amps...........................1.04
Saturation power output, kw...................2.6
Electronic efficiency at saturation, %..........15
Small-signal gain, db..........................32
Small-signal bandwidth (1 db), Mc.............780

The tube plus solenoid weigh 230 pounds. Only $\frac{1}{2}$ per cent of the beam current is intercepted on the slow-wave circuit for small rf power levels; with output powers near saturation the circuit interception increases to about $3\frac{1}{2}$ per cent. The tube is operated with the collector at the same voltage as the slow-wave structure, so that approximately 14.5 to 17 kw of power is dissipated in the water-cooled collector, depending on the rf signal level. The length of the slow-wave circuit is 26 cm, and its inside diameter is 4.2 mm. The beam interaction impedance is 13 ohms. The electron gun is a convergent gun somewhat similar to that in Figure 4.5-1(a) Its perveance is $0.466 \times 10^{-6}$ amp/volts$^{3/2}$. It produces an electron beam of 2.3 mm diameter.

The ion pumps shown in the figure are electronic vacuum pumps which help to maintain an extremely high vacuum in the tube under all conditions of operation. The tube is constructed entirely of metal and ceramic. The slow-wave structure is made of copper because of the excellent electrical and thermal conductivity of this metal.

If we compare this tube with the VA-849 klystron amplifier, whose characteristics are given in Table 9.4-1, we note that the traveling-wave tube has much greater bandwidth and considerably less electronic efficiency. Thus, a choice between these two types in any particular application would depend largely upon the relative importance of bandwidth and electronic efficiency. It should be noted that the overall efficiency of each tube can be greater than the electronic efficiency if the collector voltage is lowered below the beam-interaction-space voltage.

## 10.3   Helix Slow-Wave Circuits

The most common slow-wave structure used in traveling-wave amplifiers is the helix. Helix-type, slow-wave circuits permit large bandwidth and

high gain per unit length.  However, for rf output powers greater than
100 watts or so, a helix-type, slow-wave circuit would become excessively
heated by beam interception and rf losses.  Consequently, helix-type,
slow-wave structures are used only at lower power levels.

In order to analyze the operation of the helix traveling-wave amplifier,
we require information about the Brillouin diagram and the beam-coupling
impedance.  This information is obtainable by field analysis as in Chapter 8.
However, the field analysis of the helix is a complicated mathematical
problem; in fact, the exact solution for a helix of finite wire diameter has
not been derived.  The discussion presented here will be somewhat simpli-
fied, and we shall mention some of the more accurate results.

## (a)  The Brillouin Diagram for a Helix-Type, Slow-Wave Circuit

The electric field lines surrounding a helix are shown in Figure 10.3-1.
This field pattern moves to the right at a phase velocity corresponding to
the fundamental space harmonic.  Unlike the helix of Figure 8.7-2(a), this
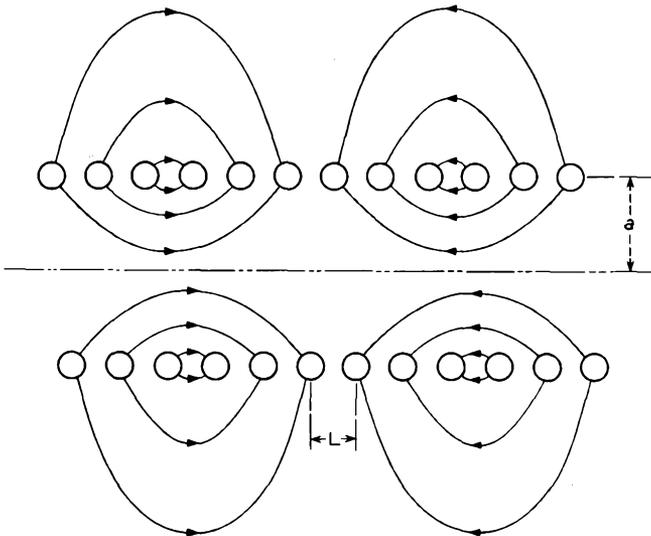


Fig. 10.3-1 The electric field lines surrounding a helix at a particular instant.
The pattern travels at a phase velocity corresponding to the fundamental
space harmonic.

helix does not have a surrounding metal tube.  The helix in Figure 10.3-1 is
shown unsupported.  In practice, it would be supported by small dielectric
rods or a surrounding dielectric cylinder, either of which would perturb
the field pattern.  However, we shall neglect this effect.

The period of the helix is called the pitch and is denoted by the symbol $L$. The mean radius of the helix is $a$. In addition, we define the pitch angle $\psi$ by the equation

$$\tan \psi = \frac{L}{2\pi a} \qquad (10.3\text{-}1)$$

The wave equation for the $z$ component of electric field is obtained from Equations (10.2-8) and (10.2-9) as

$$\left[ \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} - \gamma_n{}^2 \right] E_{zn} = 0 \qquad (10.3\text{-}2)$$

In the case of the helix the field quantities are functions of $\theta$, the angular coordinate, since the boundary conditions vary with $\theta$. The space harmonics necessary to satisfy the helix boundary conditions are of the form[17]

$$E_{zn} = R_n(r)\epsilon^{\pm jn\theta}\epsilon^{-j\beta_n z} \qquad (10.3\text{-}3)$$

where $R_n(r)$ is a function of only the radial coordinate for each space harmonic. The plus-or-minus choice for the $\theta$ dependence is dictated by the winding sense of the helix. Substituting this assumed solution into Equation (10.3-2), we obtain

$$\left[ \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} - \gamma_n{}^2 - \frac{n^2}{r^2} \right] R_n(r) = 0 \qquad (10.3\text{-}4)$$

If we put $\gamma_n{}^2 = -\tau_n{}^2$, we recognize this as Bessel's Equation with the general solution

$$R_n(r) = b_n J_n(\tau_n r) + c_n H_n{}^{(1)}(\tau_n r) \qquad (10.3\text{-}5)$$

where $J_n(\tau_n r)$ is the Bessel function of order $n$ and $H_n{}^{(1)}(\tau_n r)$ is the Hankel function of order $n$. Since $\tau_n = j\gamma_n$, the argument in the above functions is a purely imaginary number. Nevertheless, the values of the functions are either purely real or purely imaginary. The following modified Bessel functions are defined as to have purely real values:[18]

$$I_n(\gamma_n r) = j^{-n} J_n(j\gamma_n r)$$

$$K_n(\gamma_n r) = \frac{\pi}{2} j^{n+1} H_n{}^{(1)}(j\gamma_n r) \qquad (10.3\text{-}6)$$

$I_o(\gamma_o r)$, $I_1(\gamma_1 r)$, $K_o(\gamma_o r)$, and $K_1(\gamma_1 r)$ are plotted in Figure 10.3-2. The $I_n(\gamma_n r)$ functions go to infinity as $r$ goes to infinity. The $K_n(\gamma_n r)$ functions

---

[17]Reference 10c, pp. 46, 47.

[18]In order to avoid confusion with the beam interaction impedance, the Bessel function $K_n(\gamma_n r)$ will always be written showing its functional dependence. Tables of the modified Bessel functions are given in Reference 9.1, pp. 224–243.

go to infinity as $r$ goes to zero. Thus, $R_n(r)$ is given by $I_n(\gamma_n r)$ inside the helix and $K_n(\gamma_n r)$ outside the helix, in order that the fields remain finite everywhere. Each space harmonic has its maximum amplitude adjacent to
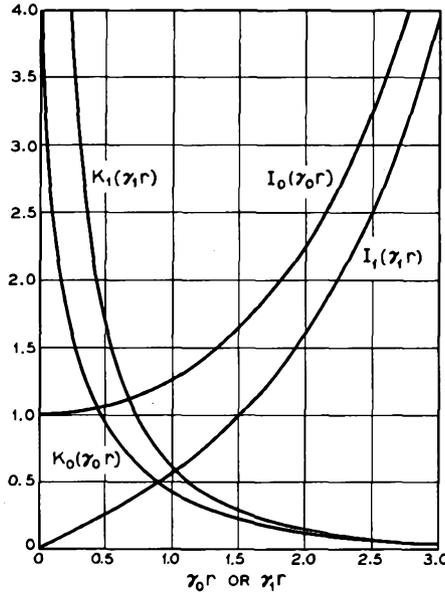


FIG. 10.3-2 Modified Bessel functions. $I_o(\gamma_o r)$ and $I_1(\gamma_1 r)$ apply to radial variations of fields within the helix, and $K_o(\gamma_o r)$ and $K_1(\gamma_1 r)$ apply to fields outside the helix.

the helix and decays radially away from the helix, both inside and outside the helix.

Thus we have the following expressions for the $z$ component of electric field. Inside the helix,

$$E_z = \sum_n B_n I_n(\gamma_n r)\epsilon^{\pm jn\theta}\epsilon^{-j\beta_n z} \qquad (10.3\text{-}7)$$

Outside the helix,

$$E_z = \sum_n C_n K_n(\gamma_n r)\epsilon^{\pm jn\theta}\epsilon^{-j\beta_n z} \qquad (10.3\text{-}8)$$

The amplitudes of the space harmonics are obtainable from the boundary conditions at the helix surface. This procedure has been carried out for a helix whose wires are thin tapes,[19] but not for round wires. After a rather

---

[19]References 10.7, 10.8.

lengthy, approximate analysis, one obtains expressions for the Brillouin diagram and beam interaction impedance. A similar analysis has been carried out for the sheath helix.[20] The sheath helix is a mathematical model consisting of a continuous, cylindrical sheath of currents, all flowing around the cylinder at the pitch angle. We shall not go through these analyses, but rather we shall cite some of the results.

First, it is found that the $\omega$-$\beta$ curve of the fundamental space harmonic of the tape helix is very nearly given by the $\omega$-$\beta$ curve of the fundamental
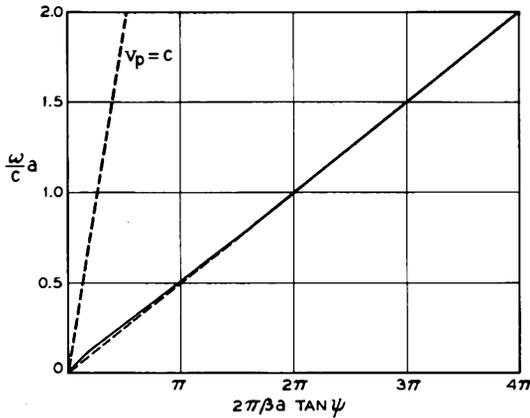


FIG. 10.3-3 $\omega$-$\beta$ curve for the fundamental mode of a sheath helix. The velocity-of-light line is shown for $\tan \psi = 0.125$.

mode of the sheath helix. The fundamental mode of the sheath helix is presented in Figure 10.3-3. It is approximately given by a line of constant slope from the origin corresponding to the phase velocity

$$v_p = \frac{\omega}{\beta} = c \tan \psi \qquad (10.3\text{-}9)$$

This value of phase velocity corresponds approximately to that of a wave traveling around the cylinder at the pitch angle with the velocity of light, which is the value physical intuition would lead us to expect. (Note that $\beta$ here is equal to $2\pi/\lambda_z$, where $\lambda_z$ is the wavelength of the field pattern shown in Figure 10.3-1.) Let us approximate the actual $\omega$-$\beta$ curve of the sheath helix by the straight line with slope given by Equation (10.3-9). The wave may be made as slow as desired by choosing the pitch angle small enough.

---

[20]Reference 10a, Chapter 3.

Let us apply this sheath helix result to derive the Brillouin diagram of the tape helix. Using Equation (10.3-1) we note that the abscissa in Figure 10.3-3 may be written as $\beta L$ for the tape helix. Taking the straight line in Figure 10.3-3 as the fundamental space harmonic for the tape helix, we draw in additional space harmonics as in Section 8.7, obtaining Figure 10.3-4.
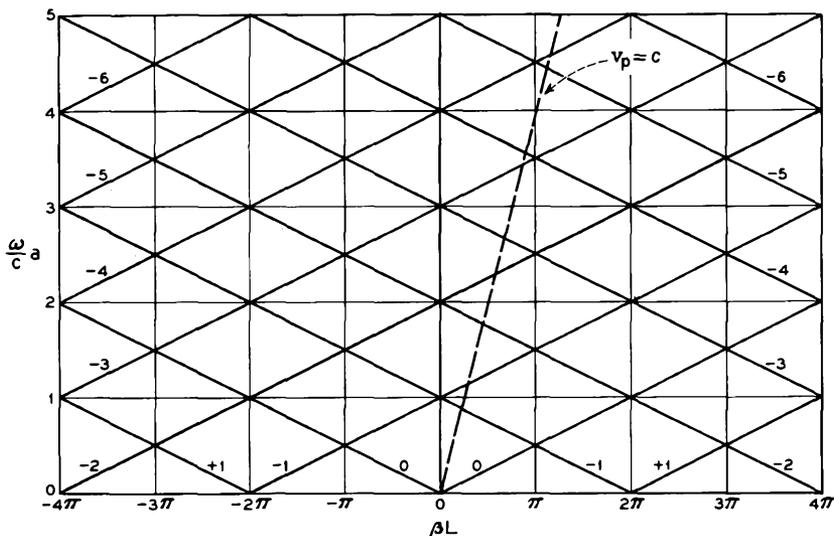


Fig. 10.3-4 Space harmonics corresponding to the straight-line approximation of the fundamental mode of a sheath helix. The branches are numbered as to the order of the space harmonic. The velocity-of-light line is shown for $\tan \psi = 0.125$.

In order to complete the Brillouin diagram for the tape helix we must include the so-called forbidden regions. The forbidden regions are regions in the Brillouin diagram for which electromagnetic energy is radiated away. In these regions the helix may be used as a helical antenna. Thus, the regions are forbidden in the sense that lossless propagation along the helix is not possible. The forbidden regions for the tape helix are shown shaded in Figure 10.3-5.

We can explain the existence of the forbidden region centered about $\theta L = 0$ in the following manner. Consider the form which the space-harmonic fields take exterior to the helix in this region. This region is characterized by the fact that all phase velocities are greater than the velocity of light. Thus, in this region,

$$\beta_n{}^2 < k^2 \qquad (10.3\text{-}10)$$
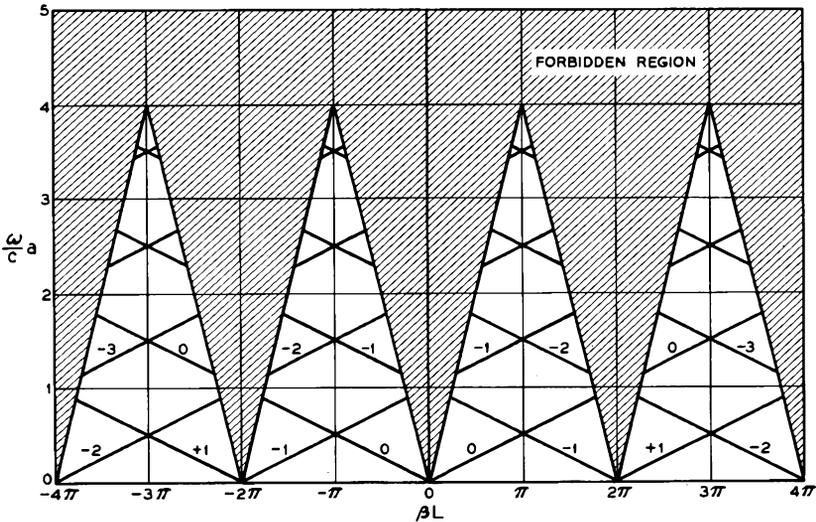
FIG. 10.3-5 Approximate Brillouin diagram for the tape helix, with $\tan \psi = 0.125$. In this approximation, changing the pitch angle changes only the lower extent of the forbidden regions. The branches are numbered as to the order of the space harmonic.

where $k = \omega/c$. This means that $\gamma_n = \sqrt{\beta_n{}^2 - k^2}$ is imaginary. If we put $\gamma_n = j\tau_n$, we obtain the following expression for $E_z$ external to the helix for the $n^{\text{th}}$ space harmonic,

$$E_{zn} = D_n H_n{}^{(1)}(-\tau_n r)\epsilon^{\pm jn\theta}\epsilon^{-j\beta_n z} \qquad (10.3\text{-}11)$$

from Equations (10.3-6) and (10.3-8). For large radii the Hankel function has the asymptotic representation:[21]

$$H_n{}^{(1)}(-\tau_n r) \underset{r \to \infty}{=} \sqrt{\frac{2}{\pi \tau_n r}} \epsilon^{-j(\tau_n r + \frac{n\pi}{2} + \frac{3\pi}{4})} \qquad (10.3\text{-}12)$$

This radial variation corresponds to that of a cylindrical wave propagating radially to infinity. Thus for $\beta_n{}^2 < k^2$ the helix radiates energy, and by definition this is a forbidden region.

The other forbidden regions in Figure 10.3-5 are explained as follows. At any operating frequency all the space harmonics must exist in order that the boundary conditions be completely satisfied. Thus, if a portion of a space harmonic lies in the forbidden region described above, all the other space harmonics within the same frequency range are also forbidden.

These forbidden regions are responsible for the following important characteristics of the helix. First, the helix will not propagate above a

---

[21]Reference 10.2, p. 159.

certain frequency ($\omega a/c \cong 3.6$ for the conditions of Figure 10.3-5). Second, frequency stop bands are produced over ranges of frequencies approximately centered about $\omega a/c = 1, 2, 3$, etc.

## (b)   The Beam Interaction Impedance for a Helix-Type, Slow-Wave Circuit

Next, we derive an approximate expression for the beam interaction impedance for the fundamental space harmonic of the helix. We do this by finding approximate expressions for the group velocity and stored energy per unit length. For the straight-line approximation to the Brillouin diagram, the group velocity is simply

$$v_g = v_p = \frac{\omega}{\beta_o} \tag{10.3-13}$$

We shall make use of the approximation $\gamma_o = \beta_o$, since $u_o \ll c$ for most helix tubes.

From a field plot as in Figure 10.3-1, we conclude that the $\theta$-component of electric field is small compared with the $r$ and $z$ components. Thus, we set $E_\theta = 0$. As a first approximation, we assume all the stored energy is in the fundamental space harmonic. For a thin tape helix the space-harmonic amplitudes must be the same at $r = a$ both inside and outside the helix, since the boundary conditions are identical. Thus, from Equations (10.3-7) and (10.3-8), we have

$$E_{zo} = B_o I_o(\gamma_o r)\epsilon^{-j\beta_o z} \tag{10.3-14}$$

inside the helix, and

$$E_{zo} = B_o \frac{I_o(\gamma_o a)}{K_o(\gamma_o a)} K_o(\gamma_o r)\epsilon^{-j\beta_o z} \tag{10.3-15}$$

outside the helix. Corresponding to these field components are the following radial components:

$$E_{ro} = jB_o \frac{\beta_o}{\gamma_o} I_1(\gamma_o r)\epsilon^{-j\beta_o z} \tag{10.3-16}$$

inside the helix and

$$E_{ro} = -jB_o \frac{\beta_o}{\gamma_o} \frac{I_o(\gamma_o a)}{K_o(\gamma_o a)} K_1(\gamma_o r)\epsilon^{-j\beta_o z} \tag{10.3-17}$$

outside the helix. The latter relationships may be verified using the divergence relation, Equation (6) of Appendix XII.

The average stored energy per unit length is equal to the peak electric

stored energy, given by[22]

$$W_{avg} = \frac{\varepsilon_o}{2} \int_0^\infty |E_{zo}|^2 + |E_{ro}|^2] 2\pi r dr$$

$$= \pi\varepsilon_o |B_o|^2 \int_0^a [I_o^2(\gamma_o r) + I_1^2(\gamma_o r)] r dr$$

$$+ \pi\varepsilon_o |B_o|^2 \frac{I_o^2(\gamma_o a)}{K_o^2(\gamma_o a)} \int_a^\infty [K_o^2(\gamma_o r) + K_1^2(\gamma_o r)] r dr$$

$$= \frac{\pi a \varepsilon_o |B_o|^2}{\gamma_o} I_o^2(\gamma_o a) \left[ \frac{1}{\gamma_o a I_o(\gamma_o a) K_o(\gamma_o a)} \right] \qquad (10.3\text{-}18)$$

The last term in the bracket is approximately equal to 2 over the useful range of the helix, as shown in Figure 10.3-6. Thus, we have the approximate expression

$$W_{avg} = \frac{2\pi a}{\gamma_o} \varepsilon_o |B_o|^2 I_o^2(\gamma_o a) \qquad (10.3\text{-}19)$$

Equation (10.1-19) may be written for the beam-coupling impedance as

$$K_o = \frac{\int |E_{zo}|^2 dS}{2\beta_o^2 S v_g W_{avg}} \qquad (10.3\text{-}20)$$

where the area integral is taken over the cross section of the beam. Introducing Equations (10.3-13), (10.3-14), and (10.3-19) we obtain

$$K_o = \frac{\int |B_o|^2 I_o^2(\gamma_o r) dS}{2S \frac{\beta_o}{\gamma_o} \omega 2\pi a \varepsilon_o |B_o|^2 I_o^2(\gamma_o a)}$$

$$= M_{2(0)}^2 \frac{30}{ka} \text{ ohms} \qquad (10.3\text{-}21)$$

where $M_{2(0)}^2$ is defined as in Equation (10.2-21). $M_{2(0)}^2$ varies between unity for a thin hollow beam grazing the helix and $1/I_o^2(\gamma_o a)$ for a thin solid beam on the axis. The expression given by Equation (10.3-21) for the impedance neglects the energy stored in the higher-order space harmonics. It can be shown[23] that the inclusion
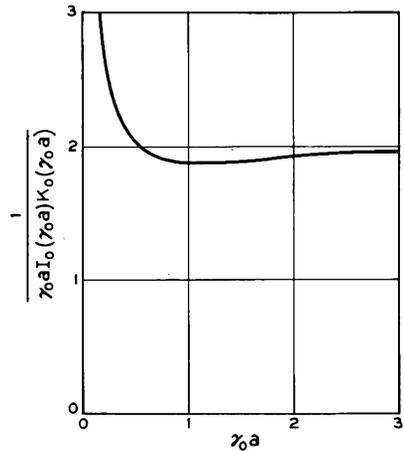


FIG. 10.3-6 Function of modified Bessel functions which is approximately equal to 2 over the useful range of the helix.

---

[22]Integrals of Bessel functions and useful identities are given in Reference 9.1, pp. 144–146. These relations may be written in terms of the modified Bessel functions using Equations (10.3-6).

[23]Reference 10.8.

of this additional energy reduces the impedance to approximately half that given by Equation (10.3-21). Thus, we have the useful approximate formula

$$K_o \cong M_{2(0)}^2 \frac{15}{ka} \text{ohms} \tag{10.3-22}$$

for the beam-coupling impedance of a tape helix where $k = \omega/c$. Although the foregoing relations are derived for a tape helix, they are also approximately valid for helices with round wires.

A typical helix tube may be designed for $ka = 0.1$ and $\gamma_o a = 1.5$. For a thin solid beam on the axis, Equation (10.3-22) gives a value of 55.3 ohms for $K_o$. This is considerably larger than values obtained with the high-power slow-wave structures discussed in the previous section. In comparing the Brillouin diagrams we see that the helix is capable of synchronism over much larger bandwidths than the high-power structures. Thus, the helix is superior in all respects except power handling capability.

(c)  *An Example of a Traveling-Wave Tube with a Helix-Type Slow-Wave Circuit*

Figure 10.3-7 shows the construction of the Western Electric 444A traveling-wave tube.[24] The slow-wave circuit consists of a molybdenum wire which is wound in a helix and glazed to three ceramic support rods. The helix assembly, consisting of the helix and support rods, slides inside a glass envelope of precise inside diameter. The electron gun is similar to that shown in Figure 4.5-1(a). An oxide-coated cathode is used.

Input and output connections to the helix are made by means of waveguides. The helix is "stretched" at each end and joined onto a cylindrical tubing which protrudes into the waveguide. This geometry provides a broadband rf match between the waveguide and the helix.

The magnetic circuit consists of two Alnico V magnets and two soft steel pole pieces. The circuit provides a nearly uniform axial magnetic focusing field of 600 oersteds. A number of permalloy "field straightener" discs are mounted perpendicular to the axis of the tube. Since these discs act as equipotential planes with respect to the magnetic field, they force the magnetic field to be axially symmetric with respect to the axis of the tube. The magnetic circuit is surrounded by an external magnetic shield (not shown in the figure) which reduces the leakage field outside the shield to a negligible value. The total weight of the magnetic circuit and shield is 85 pounds.

---

[24]Reference 10.9; the Bell Telephone Laboratories' 1789 tube is the prototype for the Western Electric 444A.
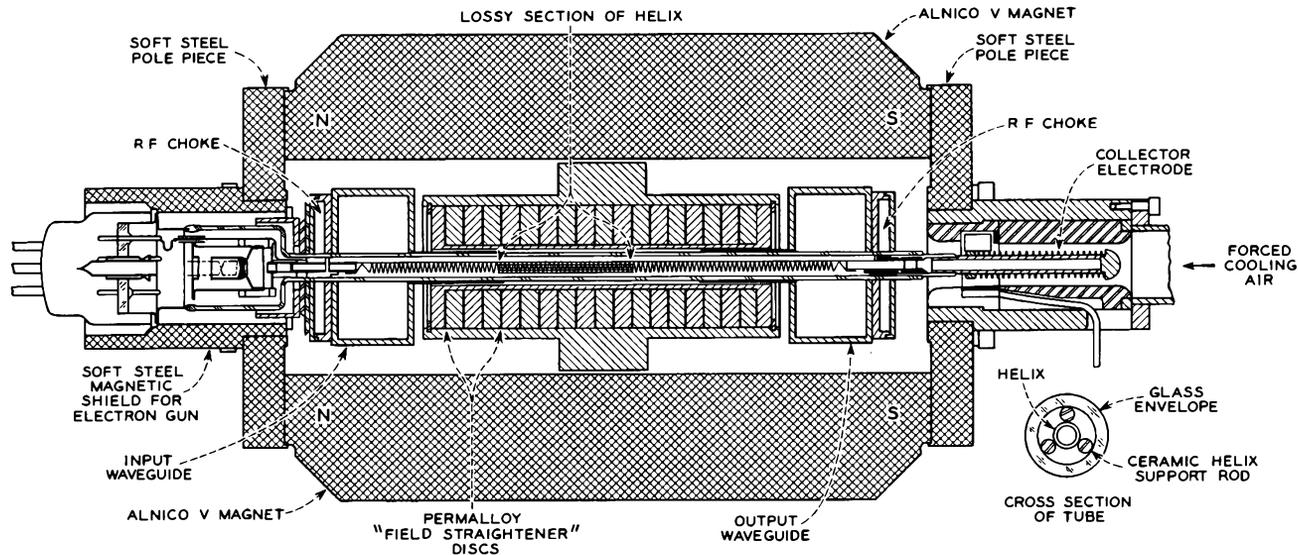
FIG. 10.3-7 The 444A traveling-wave tube and its magnetic focusing circuit. The input and output connections to the slow-wave circuit are made by means of waveguides. The axial magnetic focusing circuit is provided by two Alnico V magnets. (The polarity of the magnets does not matter provided both are magnetized in the same direction.) The thin permalloy "field straightener" discs are separated by thick nonmagnetic metal spacers. The tube can be removed from the magnetic circuit and replaced when it fails. The overall length of the tube is 31 cm.

Some additional facts about the tube are summarized in Table 10.3-1. The tube is used to provide rf amplification in a radio relay system over a band

TABLE 10.3-1. SOME FACTS PERTAINING TO THE 444A
TRAVELING-WAVE TUBE

*Helix*

Mean diameter, $2a$, mm.........................2.25
Wire diameter, mm............................0.25
Pitch, $L$, mm..................................0.75
Total turns..................................187
Length of helix, cm............................14

*Beam*

Voltage, volts.................................2400
Current, milliamps............................40
Perveance, amps/volts$^{3/2}$........................$0.3 \times 10^{-6}$
Beam power, watts............................96

*RF Interaction*

Signal frequency, Mc at midband.................6175
$ka = \omega a/c$...................................0.148
Axial wavelength, mm.........................4.7*
Total number of wavelengths on helix.............30
C...........................................0.058
$QC$...........................................0.29
Operating power output, watts...................5
Saturation power output, watts..................12
Power gain at low signal levels, db...............32–36
Power gain at 5 watts output, db.................31–35
Electronic efficiency at saturation, %.............12.6

*6.3 turns of helix.

of frequencies 500 Mc wide and centered at 6175 Mc. Over this band of frequencies the power gain is flat to within 0.7 db. The 3-db bandwidth of the tube is approximately 2400 Mc and is limited primarily by the bandwidth of the transitions between the waveguide and helix.

Near the center of the helix, the helix and support rods are coated with carbon. This provides about 70 db of attenuation on the slow-wave circuit and prevents reflected waves from the output end of the circuit from causing oscillations. This forms a severed circuit as in Section 10.2(d).

The anode of the electron gun is operated at about 200 volts higher potential than the helix to prevent positive ions formed by the beam in the region of the helix from draining toward the cathode and bombarding the cathode. The air-cooled, copper collector is normally operated at only half the helix voltage, and hence half the voltage corresponding to the electron
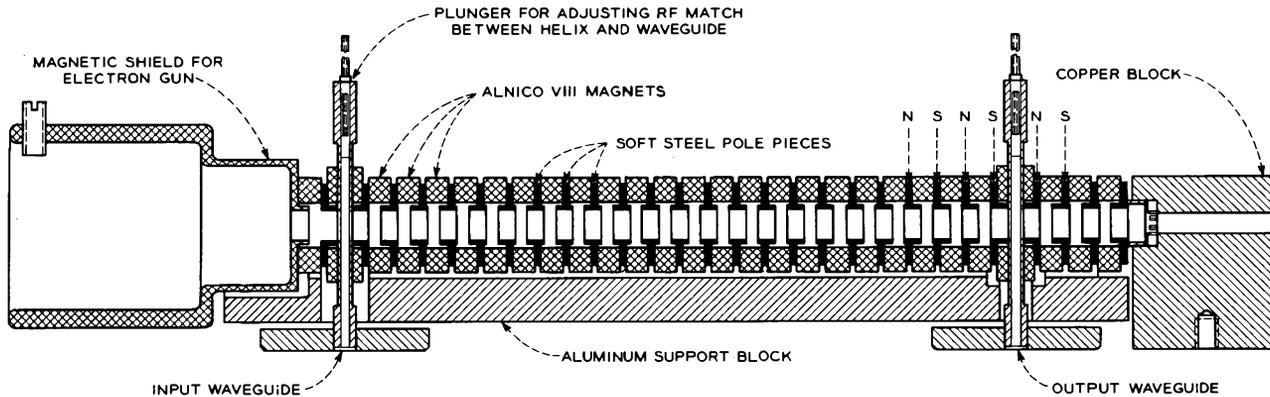
Fig. 10.3-8 A periodic permanent-magnetic (PPM) focusing system for a tube similar to the 444A. The overall length is 33 cm. The annular magnets are magnetized in the axial direction. They are stacked so that like poles of any two adjacent magnets face each other.

beam potential. Thus the overall efficiency of conversion of dc to rf power is twice the electronic efficiency.

At the time of writing, 67 per cent of an initial group of 212 444A tubes is surviving after two and one-half years of service in the radio relay system, and extrapolation of the data giving per cent survival vs. life indicates that half the tubes may survive between four and five years.

Figure 10.3-8 shows a cross-sectional view of a periodic magnetic focusing (PPM) circuit for a tube similar to the 444A. The circuit uses Alnico VIII magnets which give a peak axial magnetic field of 1000 oersteds. The leakage fields from this magnetic circuit are relatively small because of the short axial length of the individual magnets, and no external shielding is needed. The magnetic circuit is held together with an epoxy compound which is not shown in the figure. The total weight of the magnetic circuit is 8 pounds.[25] Waveguides of reduced height are used to couple rf power to the helix.

The broad bandwidths attainable with traveling-wave amplifiers make them ideally suited for broadband communication and radar systems. They are also useful as laboratory amplifiers. Bandwidths of 10 per cent in high-power tubes and 30 to 50 per cent in helix tubes are common. High gain tubes of 50 db or more are available; in this case, the circuit has more than one lossy section to prevent oscillations. Electronic efficiencies are good, but not as good as in the klystron amplifier. The positive feedback possible on the slow-wave circuit requires careful design to prevent oscillations, especially during the rise and fall times in pulsed operation, where the beam voltage is pulsed on and off.[26] High regulation and low ripple are required in the helix voltage supply to prevent undesirable phase-shift variations.

## PROBLEMS

10.1 The figure shows a typical connection of dc power supplies to a traveling-wave amplifier. Assume that the helix does not intercept any of the primary electron beam; neither does the anode of the electron gun.

(a) Which power supply provides the power that becomes useful rf power output?

(b) The bunched electron beam induces ac currents in the helix. If the helix were somewhat lossy, these currents would dissipate energy in the form of heat in the helix. Which power supply provides this energy?

[25]Other PPM stacks for similar tubes have weighed less than a pound, using platinum-cobalt magnets.

[26]As the voltage is increased toward the operating point, the electron velocity passes through points of synchronism with the phase velocity of the slow-wave structure where the impedance is very large (such as the point corresponding to $\beta_{-1}L = 2\pi$ in Figure 10.2-6). At such a point the danger of oscillation is greatest.
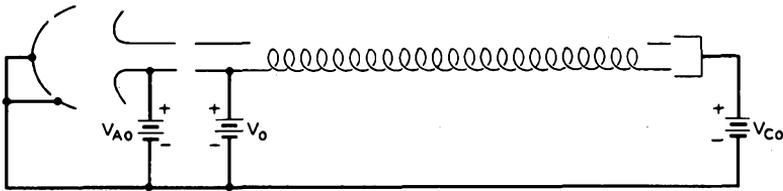
Problem 10.1

(c) If the beam current is $I_o$, what is the power supplied to the tube by the dc power supplies?

(d) The efficiency of the tube is defined as the ratio of maximum possible rf power output to the dc power input. Can the efficiency of the tube be raised by decreasing the potential $V_{Co}$? Assume that the exchange of secondary electrons between the collector and helix can be neglected.

(e) If all the electrons are to be collected by the collector, how small can the voltage $V_{Co}$ be made? Give only a qualitative answer.

10.2 The space-charge parameter $QC$ is the product of a dimensionless number $Q$ and $C$, the small-signal gain parameter. Show that $Q$ is independent of the beam current, as long as the beam current density is uniform over the beam cross section and the beam diameter is constant.

10.3 Using Figure 10.1-3 calculate the gain of a traveling-wave amplifier using the fundamental space harmonic under the following conditions:

$$f = 9 \text{ Gc}$$

$$K_o = 80 \text{ ohms}$$

$$QC = 0.50$$

$$I_o = 10 \text{ amps}$$

$$V_o = 25 \text{ kv}$$

$$d = 0.025$$

$$\beta_o = 725 \text{ radians/m}$$

Assume that a circuit sever reduces the over-all gain by 6 db and results in an effective interaction length $l$ of 7.62 cm. Neglect relativistic effects.        *Ans.*: 33.3 db.

10.4 The traveling-wave amplifier theory developed in this chapter has assumed an infinite magnetic field so that transverse motion of the electrons is negligible. In a practical tube the magnetic field is finite. When a practical tube is operated with rf power output well below the maximum (or saturated) value, the percentage of the electron beam intercepted on the slow-wave structure is negligible. When the rf power input is increased so that the output power approaches saturation, the beam interception on the slow-wave structure often becomes 2 to 5 per cent of the total beam current. Explain qualitatively the reason for this increase.

10.5 Show that the apex of the triangle bounding the forbidden region for a helix is given by $\omega a/c = \pi a/L$.

10.6 Show that the beam–coupling impedance $K_o$ for a helix-type traveling-wave amplifier utilizing the fundamental space harmonic is given by

$$K_o = [I_o{}^2(\gamma_o b) - I_1{}^2(\gamma_o b)]K_o(0)$$

where $b$ is the radius of the electron beam which is centered on the axis and $K_o(0)$ is the impedance for an infinitesimally thin beam on the axis. The following integral will be useful:

$$\int_0^r r I_o{}^2(\gamma r) dr = \frac{r^2}{2}[I_o{}^2(\gamma r) - I_1{}^2(\gamma r)]$$

10.7 Using the result of the previous problem and Equation (10.3-22), calculate the impedance of a traveling-wave amplifier with $ka = 0.2$ and $\gamma_o a = 1.0$. The diameter of the electron beam is half the helix diameter.

10.8 When positive ions are present in an electron beam, the beam diameter tends to shrink, since the electron space-charge forces of repulsion are neutralized. When the collector voltage of a traveling-wave amplifier is depressed below that of the slow-wave structure, the ions are drained out of the beam into the collector and the beam diameter increases. What will be the effect on the gain of the amplifier when the collector voltage is depressed, assuming that the amount of electrons intercepted by the slow-wave structure is negligible?

10.9 A traveling-wave amplifier, with voltages applied as in Problem 10.1, operates under the conditions $b = QC = d = 0$, so that the gain is given by Equation (10.1-75). Amplitude modulation of the rf output may be obtained by varying the anode voltage $V_{Ao}$, with the other voltages and rf power input held constant.
  (a) Show that the percentage change in rf power output is related to the percentage change in anode voltage by the expression:

$$\frac{dP_{out}}{P_{out}} = 5.45CN \frac{dV_{Ao}}{V_{Ao}}$$

  (b) Which voltage should be varied to produce phase modulation on the rf output with a minimum of amplitude modulation?

10.10 The slow-wave structure of a traveling-wave amplifier is severed perfectly at a point sufficiently far from the input that only the growing wave is of importance. (Perfect severing implies that circuit waves from either direction are absorbed without reflection and also that the sever is so short that the beam convection current and velocity are unchanged before and after the sever.) Find the loss in over-all gain of the device in db due to the severing. Assume small $C$, and $b = QC = d = 0$. *Hint:* Rewrite and solve Equations (10.1-53) for the new initial values of electric field, velocity, and convection current just beyond the sever. The electric field is zero; the velocity and convection current are continuous from just before the sever. Find the ratio of the magnitudes of the growing wave $E_{zr1}$ just before and just after the sever.                    *Ans.*: 3.52db.

## REFERENCES

Three general references to the material of this chapter are:
10a. J. R. Pierce, *Traveling-Wave Tubes*, D. Van Nostrand Co., Inc., Princeton, N. J., 1950.

10b. A. H. W. Beck, *Space-Charge Waves and Slow Electromagnetic Waves*, Pergamon Press, Inc., New York, 1958.
10c. D. A. Watkins, *Topics in Electromagnetic Theory*, John Wiley and Sons, Inc., New York, 1958.

Other references covering specific items are:

10.1 J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill Book Co., Inc., New York, pp. 131–132, 1941.
10.2 F. B. Hildebrand, *Advanced Calculus for Engineers*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1953.
10.3 W. H. Louisell, "Approximate Analytic Expressions for TWT Propagation Constants," *Trans. IRE* **ED-5,** 257–259, October, 1958.
10.4 C. K. Birdsall and G. R. Brewer, "Traveling-Wave-Tube Characteristics for Finite Values of C," *Trans. IRE* **ED-1,** pp. 1–11, August, 1954.
10.5 G. R. Brewer and C. K. Birdsall, "Traveling-Wave Tube Propagation Constants," *Trans. IRE* **ED-4,** pp. 140–144, April, 1957.
10.6 R. J. Collier, G. D. Helm, J. P. Laico, and K. M. Striny, "The Ground Station High-Power Traveling-Wave Tube," *Bell System Tech. J.* **42,** 1829–1861, July, 1963.
10.7 S. Sensiper, "Electromagnetic Wave Propagation on Helical Conductors," *Tech. Report No. 194,* MIT Research Lab. of Electronics, Cambridge, Mass., May, 1951.
10.8 P. K. Tien, "Traveling-Wave Tube Helix Impedance," *Proc. IRE* **41,** 1617–1623, November, 1953.
10.9 J. P. Laico, H. L. McDowell, and C. R. Moster, "Medium-Power Traveling-Wave Tube for 6000 Mc Radio Relay," *Bell System Tech. J.* **35,** 1285–1346, November, 1956.

# Chapter 11

## BACKWARD-WAVE OSCILLATORS
## AND AMPLIFIERS

Let us consider the operation of the traveling-wave tube shown in Figure 11-1. This tube uses a helix as the slow-wave structure with input and output connections to coaxial lines. In contrast to the forward-wave amplifier, this tube has an rf signal impressed onto the helix at the collector end, with
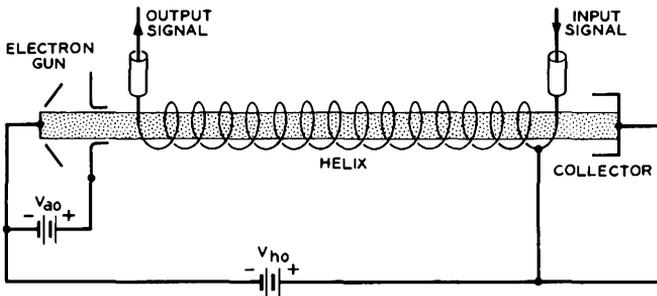


FIG. 11-1 A traveling-wave tube in operation as a backward-wave amplifier. A separate power supply connected to the anode permits beam current control independent of the helix voltage.

the signal output at the gun end. Separate power supplies are provided for the anode and the helix. This provides a means for adjusting the beam current independent of the helix voltage. The beam is assumed to be confined by a strong axial magnetic field.

The helix voltage $V_{ho}$ is adjusted to a value less than that normally used for forward-wave amplification. Assume $V_{ho}$ corresponds to an electron velocity $u_o$ such that a plot of $ka$ vs. $\omega L/u_o = \beta_e L$ gives the line $OA$

in Figure 11-2. At the point of inter-
section of this line with the $-1$ space
harmonic, the space-harmonic phase
velocity is equal to the electron veloc-
ity, and some sort of synchronous
interaction is to be expected. Since
the group velocity for this space
harmonic is opposite to the direction
of electron travel, we would not ex-
pect the interaction to be the same
as that described in the preceding
chapter.

Let us consider an electron on the
outer edge of the electron beam at
such a radius that it nearly grazes
the helix. Alternatively, we could
consider an electron in a thin hollow
beam of essentially the helix diam-
eter. We shall assume that the



FIG. 11-2 Brillouin diagram for a tape
helix. The helix radius and pitch are
given by the symbols $a$ and $L$, re-
spectively. The slope of the line $OA$ is
proportional to the square root of the
helix voltage.

helix is wound from a thin flat tape of metal. The electron sees rf axial
electric field due to the helix while passing the gap between adjacent
turns of the tape, and zero axial field when passing adjacent to a tape.
We may consider these gaps as points of interaction between the beam
and the helix.

The helix-beam coupled system effectively comprises a system of feedback
loops as indicated in Figure 11-3. $\theta_1$ is the total phase shift around a loop



FIG. 11-3 Electrons adjacent to a tape helix interacting with the fields in the helix
gaps. $\theta_1$, $\theta_2$, and $\theta_3$ denote loop phase shifts for one, two, and three periods,
respectively.

encompassing one period $L$ of the helix, $\theta_2$ two periods, etc., for a frequency
corresponding to the intersection of line $OA$ with the $-1$ space harmonic in.
Figure 11-2. The circuit is propagating a wave to the left with a funda-
mental phase shift $\beta_o L$ per period at this frequency, due to the impressed
signal. Since $\beta_e L$ is the phase change in the rf signal which occurs when the
beam travels a distance $L$, we have

$$\theta_1 = \beta_e L + \beta_o L$$
$$\theta_2 = 2(\beta_e L + \beta_o L)$$
$$\theta_3 = 3(\beta_e L + \beta_o L), \text{ etc.} \tag{11-1}$$

From Figure 11-2 we observe that $\beta_{-1}L$ and $\beta_o L$ are related numerically by

$$\beta_{-1}L = 2\pi - \beta_o L \tag{11-2}$$

Under the condition of synchronism, this must be equal to $\beta_e L$,

$$\beta_e L = 2\pi - \beta_o L \tag{11-3}$$

Introducing this expression into Equations (11-1), we obtain

$$\theta_1 = 2\pi$$
$$\theta_2 = 4\pi$$
$$\theta_3 = 6\pi, \text{ etc.} \tag{11-4}$$

The total loop phase shift for each of the feedback loops is an integral multiple of $2\pi$ radians. Hence, the tube will oscillate provided that the gain per feedback loop is sufficient. If the gain is not sufficient for oscillation, the tube acts as a narrow band regenerative amplifier. If the tube is used as an oscillator, there is no need for an impressed signal at the collector end of the helix. Oscillations are generated in the usual manner; that is, an infinitesimal noise signal at the proper frequency builds up because of the positive feedback until a stable operating point is reached.

If the tube of Figure 11-1 is to be used as an oscillator, the rf input signal is replaced by a passive termination, and operation proceeds in the following manner. The beam current is increased from zero by increasing the anode voltage $V_{ao}$, with $V_{ho}$ held constant. As the beam current increases, the gain per feedback loop also increases. Finally, a point is reached where the tube breaks into oscillation. The dc beam current corresponding to this point is termed the starting current, $I_{ST}$.

Since there are many feedback loops, it is not necessary that each one have a loop gain equal to unity for oscillations to be produced. In fact, we should expect the starting current to vary inversely with the number of feedback loops or the length of the helix. We shall find that oscillations are produced when the product of the gain per wavelength along the helix times the number of circuit wavelengths exceeds some critical value. In terms of the traveling-wave tube parameters introduced in the preceding chapter, this product is proportional to $CN$.

A traveling-wave tube operated in this manner is known as a backward-wave oscillator. The rf output power is taken from the electron gun end of the helix. In external physical appearance it is very similar to a traveling-wave amplifier; the helix is usually shorter and larger in diameter, and there is no circuit sever or loss pattern.

The frequency of oscillation is tuned electronically by varying the helix voltage. This is easily seen by referring to Figure 11-2. Changing the beam voltage changes the frequency corresponding to the intersection of the line $OA$ with the $-1$ space harmonic. Extremely wide electronic tuning ranges are obtainable; typically, the ratio of the highest to the lowest frequency is two to one.

With the beam current adjusted to a value below the starting current, the tube is usable as a backward-wave amplifier. RF power introduced at the collector end of the helix is amplified and delivered out the gun end. Because of the regenerative nature of the amplification process, extremely narrow fractional bandwidths are obtained, of the order of 0.5 per cent. The frequency of amplification is tuned electronically by varying the helix voltage. In systems requiring a highly selective, electronically tunable amplifier the backward-wave amplifier is unexcelled.

In the following sections we shall examine the quantitative aspects of these devices. Both oscillation and amplification modes of behavior are derived from the same basic theory, which is closely related to the theory of the preceding chapter.

## 11.1 Theory of Backward-Wave Interaction

The theory of backward-wave interaction closely parallels that given for the traveling-wave amplifier in Section 10.1. An electronic equation and a circuit equation are derived individually and then solved simultaneously to determine allowed values for the "hot" propagation constants.

### (a) The Electronic Equation

This equation expresses the electron motion induced by a space harmonic of the circuit field traveling synchronously with the electrons. It is given by Equation (10.1-18) for backward waves as well as for forward waves. It should be clear that synchronism refers to the equality of the electron velocity and the space-harmonic *phase velocity;* no condition is placed on the space-harmonic *group velocity.*

### (b) The Circuit Equation

This equation expresses the manner in which currents are induced into the slow-wave structure by the beam convection currents and the way in which these induced currents propagate and combine.

This equation is derived with reference to Figure 10.1-1. In the case of backward-wave interaction, the arrows above the circuit refer to the direction of the *group velocity,* the direction of energy propagation, for each

of the waves shown. In the traveling-wave amplifier there was no ambiguity since the phase and group velocities were in the same direction. The convection current segment in the beam induces power into the circuit, giving rise to incremental waves having group velocities directed away from the point of induction.

With this interpretation in mind, the development of Section 10.1(b) follows exactly for backward-wave interaction up to and including Equation (10.1-27). We shall proceed with the development from this point.

Let $\Gamma_o = -\alpha + j\beta_n$ be the complex cold circuit propagation constant for a synchronous backward-wave space harmonic with positive phase velocity and negative group velocity. The $-1$ space harmonic shown in Figure 11-2 is of this type. Such a wave has a $z$ dependence of the form

$$\epsilon^{-\Gamma_o z} = \epsilon^{\alpha z}\epsilon^{-j\beta_n z} \qquad (11.1\text{-}1)$$

with $\alpha$ and $\beta_n$ both positive. We note that the wave amplitude is attenuated in the minus $z$ direction corresponding to the direction of power flow.

In order to include the possibility of a backward-wave amplifier we assume that energy is introduced onto the circuit from an external signal source at the collector end. The total space-harmonic field at an arbitrary point ($z = a$) on the circuit is then given by the superposition of three contributions as follows:

A. The power coming from the external source at the collector end of the circuit ($z = l$),

$$E_{znA}(a) = E_{zno}\epsilon^{\Gamma_o(l-a)} \qquad (11.1\text{-}2)$$

where $E_{zno}$ is the value of the space harmonic at $z = l$ corresponding to this power.

B. The superposition of the incremental waves $dE_{zn+}$ arriving at $z = a$ from the left are given by

$$E_{znB}(a) = \int_0^a \epsilon^{\Gamma_o(a-z)}dE_{zn+} \qquad (11.1\text{-}3)$$

C. The superposition of the incremental waves $dE_{zn-}$ arriving at $z = a$ from the right are given by

$$E_{znC}(a) = \int_a^l \epsilon^{\Gamma_o(z-a)}dE_{zn-} \qquad (11.1\text{-}4)$$

In all cases we note that the waves advance in phase and decay in amplitude from the point of origin, characteristic of backward waves.

The total field at $z = a$ is obtained by summing the above three contribu-

tions and making use of Equation (10.1-27):

$$E_{zn}(a) = E_{zno}\epsilon^{\Gamma_o(l-a)} - \frac{1}{2}\beta_n^2 K_n \int_0^a i\epsilon^{\Gamma_o(a-z)}dz - \frac{1}{2}\beta_n^2 K_n \int_a^l i\epsilon^{\Gamma_o(z-a)}dz$$

$$(11.1-5)$$

where $i$ is the ac convection current in the electron beam. The variable of integration is replaced by $\tau$, and $a$ is replaced by $z$, a variable point, obtaining

$$E_{zn}(z) = E_{zno}\epsilon^{-\Gamma_o(z-l)} - \frac{1}{2}\beta_n^2 K_n \int_0^z i(\tau)\epsilon^{\Gamma_o(z-\tau)}d\tau$$

$$- \frac{1}{2}\beta_n^2 K_n \int_z^l i(\tau)\epsilon^{-\Gamma_o(z-\tau)}d\tau \quad (11.1-6)$$

This equation can be differentiated twice to obtain

$$\Gamma^2 E_{zn} = \Gamma_o^2 E_{zno}\epsilon^{-\Gamma_o(z-l)} - \frac{1}{2}\Gamma_o^2\beta_n^2 K_n \int_0^z i(\tau)\epsilon^{\Gamma_o(z-\tau)}d\tau$$

$$- \frac{1}{2}\Gamma_o^2\beta_n^2 K_n \int_z^l i(\tau)\epsilon^{-\Gamma_o(z-\tau)}d\tau - \Gamma_o\beta_n^2 K_n i \quad (11.1-7)$$

where it has been assumed that the resultant space harmonic has a "hot" $z$ dependence of the form

$$\epsilon^{-\Gamma z}$$

The last two equations are combined to yield the *circuit equation* for backward-wave interaction:

$$(\Gamma^2 - \Gamma_o^2)E_{zn} = -\Gamma_o\beta_n^2 K_n i \qquad (11.1-8)$$

This differs from the corresponding forward-wave amplifier equation, Equation (10.1-34), only in the sign of the term containing the impedance.

### (c)  Solutions for Cumulative Interaction

Allowed values of the "hot" propagation constant $\Gamma$ are determined from a simultaneous solution of the electronic and circuit equations.  One obtains

$$(\Gamma^2 - \Gamma_o^2)\left[(\Gamma - j\beta_e)^2 + \frac{\omega_q^2}{u_o^2}\right] = -\frac{j\beta_e\beta_n^2\Gamma_o K_n I_o}{2V_o} \qquad (11.1-9)$$

This equation is simplified by introducing certain parameters.  $C$ and $QC$ are defined as in Equations (10.1-36) and (10.1-37).  $b$ and $d$ are defined by

$$\Gamma_o \equiv j\beta_e(1 + Cb + jCd) \qquad (11.1-10)$$

and $\delta$ is defined by Equation (10.1-39).  Equations (10.1-40) and (10.1-41)

are correct as written for backward-wave interaction. Equation (11.1-9) simplifies to

$$\delta^2 = \frac{1}{b + jd - j\delta} - 4QC \qquad (11.1\text{-}11)$$

Equation (11.1-11) determines three allowed values for the "hot" propagation constant in terms of the various operating parameters of the tube. The various fields and beam quantities may thus be written as a superposition of three waves, as in Equations (10.1-53). The boundary conditions at $z = 0$ for these various quantities are the same as in the forward-wave amplifier, as indicated in Equations (10.1-60). It should be noted that $E_{zn}(0)$ corresponds to the rf output power for backward-wave interaction, and we have not so far determined its value explicitly. Nonetheless, we can solve Equation (10.1-60) to determine the values of $E_{zT1}$, $E_{zT2}$, and $E_{zT3}$ in terms of $E_{zn}(0)$. Thus, we obtain Equation (10.1-61) and similar expressions for $E_{zT2}$ and $E_{zT3}$ by permuting the subscripts.

It will be most convenient to write all of our equations in terms of the waves corresponding to the space-harmonic field due to the circuit. Combining Equations (10.1-61) and (10.1-73), we obtain

$$\frac{E_{zn1}}{E_{zn}(0)} = \frac{\delta_1{}^2 + 4QC}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)} \qquad (11.1\text{-}12)$$

with expressions for $E_{zn2}$ and $E_{zn3}$ obtained by permuting subscripts. The total space-harmonic field at any axial position is given by

$$E_{zn}(z) = E_{zn1}\epsilon^{-\Gamma_1 z} + E_{zn2}\epsilon^{-\Gamma_2 z} + E_{zn3}\epsilon^{-\Gamma_3 z} \qquad (11.1\text{-}13)$$

Using Equation (11.1-12) and similar expressions for $E_{zn2}$ and $E_{zn3}$, we obtain at $z = l$:

$$\frac{E_{zn}(l)}{E_{zn}(0)} = \frac{\delta_1{}^2 + 4QC}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)}\epsilon^{-\Gamma_1 l} + \frac{\delta_2{}^2 + 4QC}{(\delta_2 - \delta_3)(\delta_2 - \delta_1)}\epsilon^{-\Gamma_2 l}$$

$$+ \frac{\delta_3{}^2 + 4QC}{(\delta_3 - \delta_1)(\delta_3 - \delta_2)}\epsilon^{-\Gamma_3 l} \qquad (11.1\text{-}14)$$

The definitions given by Equations (10.1-39) and (10.1-70) are substituted into this equation, obtaining

$$\frac{E_{zn}(l)}{E_{zn}(0)}\epsilon^{j2\pi N} = \frac{\delta_1{}^2 + 4QC}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)}\epsilon^{2\pi\delta_1 CN} + \frac{\delta_2{}^2 + 4QC}{(\delta_2 - \delta_3)(\delta_2 - \delta_1)}\epsilon^{2\pi\delta_2 CN}$$

$$+ \frac{\delta_3{}^2 + 4QC}{(\delta_3 - \delta_1)(\delta_3 - \delta_2)}\epsilon^{2\pi\delta_3 CN} \qquad (11.1\text{-}15)$$

The gain of a backward-wave amplifier is given by

$$\text{gain} = 20 \log \left| \frac{E_{zn}(0)}{E_{zn}(l)} \right| \text{db} \qquad (11.1\text{-}16)$$

which may be calculated directly from Equation (11.1-15). The gain is seen to be a function of $QC$ and $CN$ directly and of $b$, $d$, and $QC$ indirectly, since the latter quantities determine the values of the $\delta$'s. We shall further consider the backward-wave amplifier in a later section.

As the value of the right-hand side of Equation (11.1-15) approaches zero, the backward-wave amplifier gain approaches infinity. At this point the tube will oscillate, since a finite value of $E_{zn}(0)$ is obtainable for a vanishingly small value of $E_{zn}(l)$. Thus, the threshold conditions for oscillation are given by the solution of the equation

$$\frac{\delta_1{}^2 + 4QC}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)}\epsilon^{2\pi\delta_1 CN} + \frac{\delta_2{}^2 + 4QC}{(\delta_2 - \delta_3)(\delta_2 - \delta_1)}\epsilon^{2\pi\delta_2 CN}$$
$$+ \frac{\delta_3{}^2 + 4QC}{(\delta_3 - \delta_1)(\delta_3 - \delta_2)}\epsilon^{2\pi\delta_3 CN} = 0 \quad (11.1\text{-}17)$$

Since this equation has real and imaginary parts, there are in reality two simultaneous equations to be solved. Thus, two conditions are determined. One condition is $CN$, a quantity which we may interpret loosely as the product of gain per wavelength times the number of circuit wavelengths. This interpretation is based upon assuming an analogy to forward-wave amplifier theory, whereas in reality Equation (11.1-15) shows that backward-wave amplifier gain is given by an extremely complicated expression, not simply proportional to $CN$. The other condition determined by Equation (11.1-17) is $b$, the degree of synchronism between the beam and the space harmonic. This condition is analogous to the phase condition in oscillator theory. For a given helix voltage, this condition determines the frequency of oscillation precisely.

Solutions for the backward-wave oscillator starting conditions are not easily obtained. This is due to the fact that the values of the $\delta$'s are by Equation (11.1-11) functions of $b$ and $C$, which are unknowns. Thus, Equations (11.1-11) and (11.1-17) must be solved simultaneously, a task accomplished most readily with an analogue computer.

Let us examine the nature of the solutions in two simple cases. First, we assume zero space charge $(QC = 0)$ and zero circuit loss $(d = 0)$. The start-oscillation conditions are from computer calculations:[1]

$$CN = 0.314$$
$$b = 1.522 \qquad (11.1\text{-}18)$$

For a given helix voltage on a particular tube, the first condition in effect determines the beam current above which oscillations are produced. The second condition determines the exact frequency of oscillation. The cor-

[1]Reference 11.1.

responding values of $\delta$ are

$$\delta_1 = 0.725 + j0.150$$

$$\delta_2 = -0.725 + j0.150$$

$$\delta_3 = 0.083 - j1.796 \qquad (11.1\text{-}19)$$

Phasors representing the three waves are shown in Figure 11.1-1. These phasors are derived from Equation (11.1-12) and the appropriate values of $\delta$.



(a) COLLECTOR END
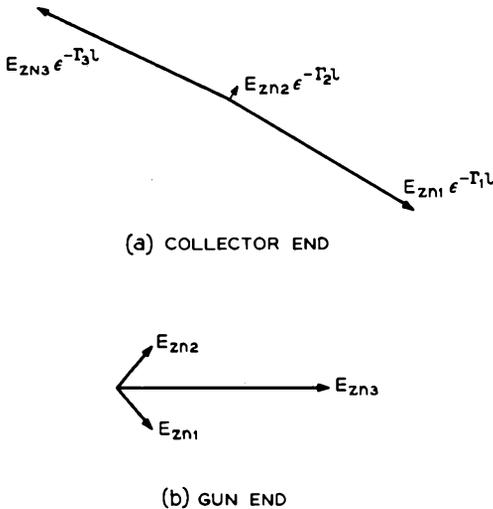
(b) GUN END

FIG. 11.1-1 Phasor diagrams for the three waves in a backward-wave oscillator with $QC = d = 0$. (a) Waves at the collector end of the helix; these waves add to zero. (b) Waves at the gun end of the helix; these waves add to produce the output signal.

At the collector end of the circuit, the three phasors add to zero. As these waves propagate from the collector end to the gun end, they experience differential phase shift and different degrees of amplitude change by Equations (11.1-19) so as to obtain a non-zero resultant field at the gun end. The only wave that grows in this direction is $E_{zn2}$, and we see from Figure 11.1-1 that this contributes only slightly to the total field at the gun end. Thus, backward-wave interaction is seen to be principally an interference effect between various waves rather than a growing-wave phenomenon as in the forward-wave amplifier. The conditions in a backward-wave amplifier are similar to those shown in Figure 11.1-1; in this case, the three waves at the collector end do not quite add to zero.

As a second example, we consider the solutions for large space charge $(QC > 0.25)$ and zero circuit loss $(d = 0)$. Under these conditions it can be shown[2] that the start-oscillation conditions are

$$CN \cong \tfrac{1}{2}(QC)^{1/4}$$

$$b \cong \sqrt{4QC} \qquad (11.1\text{-}20)$$

and the corresponding values of $\delta$ are

$$\delta_1 \cong -j\sqrt{4QC}[1 - \tfrac{1}{4}(QC)^{-3/4}]$$

$$\delta_2 \cong j\sqrt{4QC}$$

$$\delta_3 \cong -j\sqrt{4QC}[1 + \tfrac{1}{4}(QC)^{-3/4}] \qquad (11.1\text{-}21)$$

$\delta_2$ corresponds to the fast space-charge wave, as in Equation (10.1-51). This wave is negligible in amplitude compared with the other two waves. The phasor diagrams for the three waves are shown in Figure 11.1-2 for $QC = 1$. We see that for large space charge we may think of the backward-wave oscillator circuit field as given by the interference of two waves. At the collector end they are 180 degrees out of phase, whereas differential phase shift causes them to be in phase at the gun end.

Accurate computations have been made of the start-oscillation conditions over the useful ranges of space charge and circuit loss.[3] These results are presented in Figures 11.1-3 and 11.1-4. These curves are used as follows. For a given helix voltage the ratio $Q/N$ is calculated (both $Q$ and $N$ are independent of beam current)[4]. $CN$ for start oscillations is obtained from Figure 11.1-3. Then $b$ is calculated from Figure 11.1-4 and the known value of $CN$. The beam current can be calculated from the value of $CN$ and the known values of beam-coupling impedance, beam diameter, etc. This procedure gives the start-



(a) COLLECTOR END
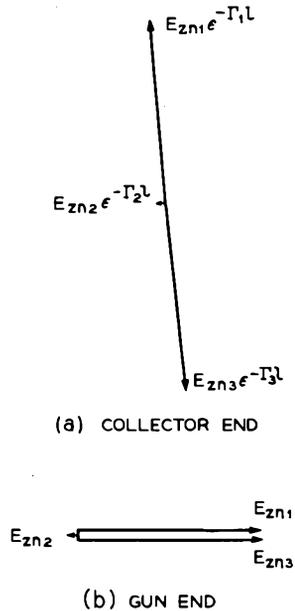
(b) GUN END

Fig. 11.1-2 Phasor diagrams for the three waves at the two ends of the helix of a backward-wave oscillator with $QC = 1$, $d = 0$. (a) Collector end. (b) Gun end.

[2] Reference 11.1. Equations (11.1-21) were not correct as originally published.
[3] *Ibid.*
[4] See Problem 10.2.
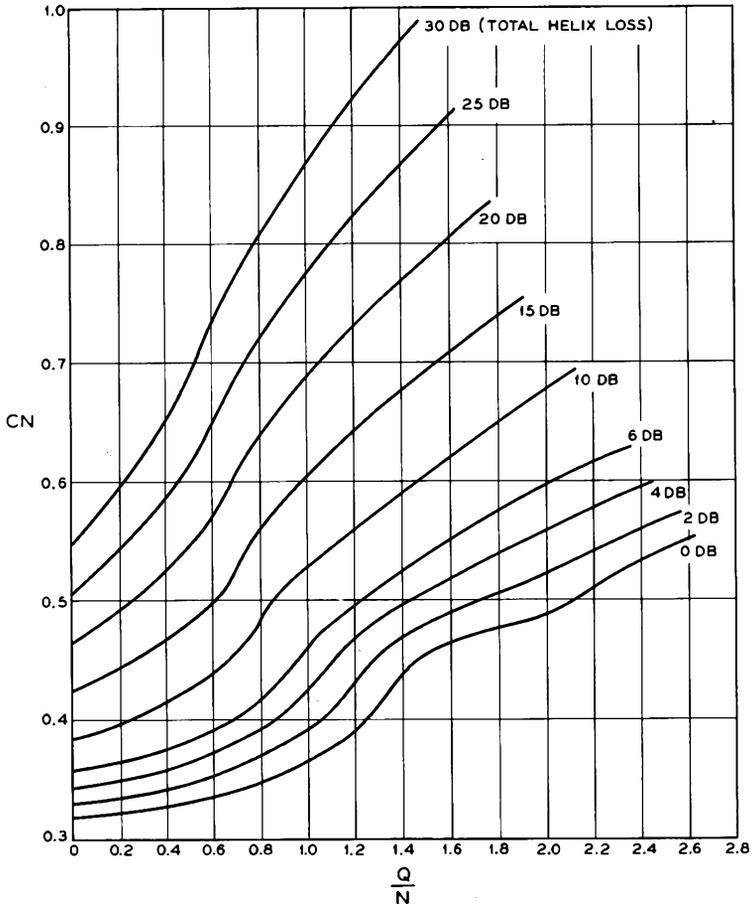
F$_{IG}$. 11.1-3 Oscillations are produced in a backward-wave oscillator for values of *CN* equal to or greater than the value given above. Its value is a function of the space-charge parameter and the total circuit loss (Reference 11.1). (Courtesy of *Proceedings IRE*)

ing current at one frequency; it is repeated to determine the starting conditions at other frequencies.

The above theory and results apply only at the threshold of oscillation. The equations were derived on the basis that all rf beam perturbations are extremely small so that the various physical equations are all linear.[5] As

---

[5]This was discussed in connection with Equations (9.3-12) and (10.1-12).

the beam current is increased above the starting current, the level of oscillation is limited by nonlinear effects in the beam kinematics. The output power increases with beam current and the oscillation frequency decreases slightly; the latter effect is known as frequency pushing.
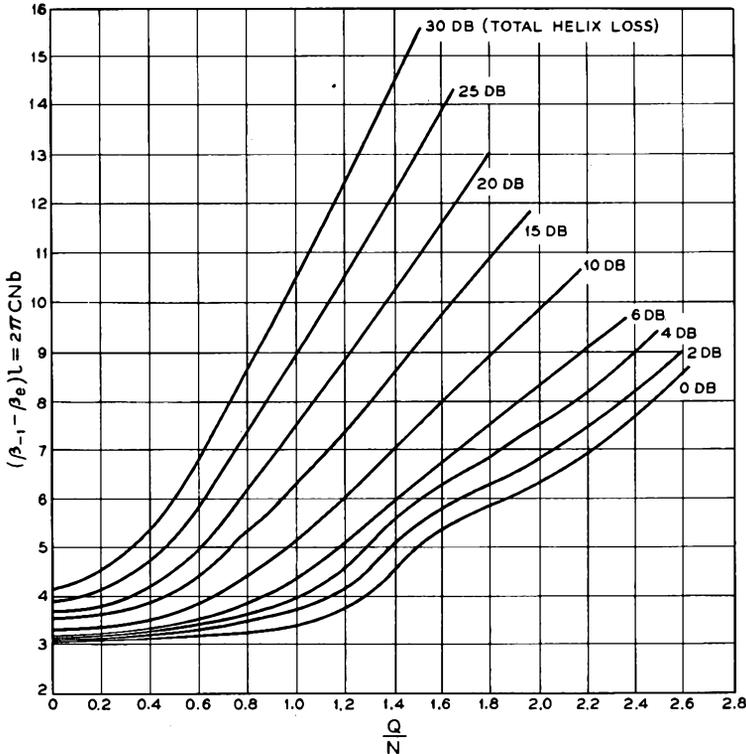


FIG. 11.1-4 Curves determining $b$ in a backward-wave oscillator at start oscillation as a function of the space-charge parameter and circuit attenuation. $b$ determines the exact frequency of oscillation, using Equation (10.1-40) (Reference 11.1). (Courtesy of *Proceedings IRE*)

Another effect that may appear at higher beam currents is a higher mode of oscillation. This mode is predicted as the next higher-order solution of the transcendental Equation (11.1-17). Since this mode requires a beam current related to the main-mode starting current by a factor of eight or greater, it is not obtained under normal operating conditions.

The rf power output cannot be predicted from the linear theory presented here. A combined theoretical-empirical analysis has been made to

determine this important quantity.[6] The results are that the maximum electronic efficiency is given by

$$\eta_e \cong 1.3C \qquad\qquad (11.1\text{-}22)$$

for small space charge $(QC < 0.5)$ and by

$$\eta_e \cong \frac{\omega_q}{\omega}$$

for large space charge $(QC > 0.5)$, where $C$ and $\omega_q$ are calculated at the operating value of beam current, which is somewhat larger than the starting current. Since $C$ and $\omega_q/\omega$ are typically quite small in these tubes, maximum electronic efficiencies are usually no more than a few per cent. These efficiencies are maximum in the sense that various imperfections in the tube (such as circuit loss and poor circuit match) cause further reduction of the output power.

## 11.2  Backward-Wave Oscillators

The most commonly used slow-wave structure for backward-wave oscillators is the helix. Interaction is with the $-1$ space harmonic as shown in Figure 11-2.

The characteristics of the Brillouin diagram for the helix have been determined in the preceding chapter. The frequency of oscillation is approximately given by the intersection of the voltage line $OA$ in Figure 11-2 with the $-1$ space harmonic. An accurate determination of the frequency requires the use of Equation (10.1-40) together with a knowledge of $b$ and $C$.

Over what frequency range will a typical tube operate? As the helix voltage is increased, the voltage line approaches coincidence with the fundamental space harmonic. This situation occurs at a frequency corresponding to $ka = 0.5$. The electrons are in synchronism with two space harmonics simultaneously. If the ends of the helix are not perfectly terminated, reflections are set up which are amplified in both directions, causing erratic behavior in the power output as a function of frequency. This situation is usually avoided by restricting operation to frequencies below this point.

At extremely low frequencies the coupling impedance $K_{-1}$ drops off rather seriously, as we shall see later. Since there is an upper limit to the beam current, $CN$ falls below the start-oscillation value, and the tube will not oscillate. Thus, operation is usually restricted to the frequency range from $ka$ equal to 0.1 or 0.2 up to $ka$ slightly below 0.5.

---

[6]Reference 11.2.

Next we shall consider an evaluation of the coupling impedance $K_{-1}$. No exact theoretical analysis has been made of this parameter. However, several rather complicated approximate analyses have been made.[7]  Follow-
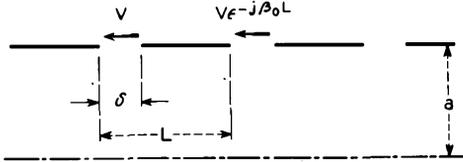


FIG. 11.2-1 Cross section of a thin tape helix obtained by the intersection with the semi-infinite plane given by $\theta = 0$.

ing the pattern set up in the previous chapter, we shall present an extremely simple approximate evaluation of $K_{-1}$ which compares favorably with the more complicated analyses over the useful frequency range of the helix.

Let us consider a thin tape helix. Figure 11.2-1 shows a cross-sectional view of the helix obtained by means of the intersection with the semi-infinite plane $\theta = 0$. By Floquet's Theorem the voltages across two successive gaps in this plane differ by the factor $\epsilon^{-j\beta_0 L}$, assuming negligible circuit attenuation. We consider, for the purpose of this analysis, the set of space harmonics with positive group velocity. Identical results are obtained for the set with negative group velocity.

From Equation (10.3-7) the axial electric field at $r = a$ and $\theta = 0$ is given by

$$E_z = \sum_n B_n I_n(\gamma_n a)\epsilon^{-j\beta_n z} \qquad (11.2\text{-}1)$$

We assume that the electric field does not vary with position within the gap. The evaluation of the coefficients $B_n$ proceeds as in Equations (10.2-15) to (10.2-17), obtaining

$$B_n L I_n(\gamma_n a) = M_{1(n)} V \qquad (11.2\text{-}2)$$

where $M_{1(n)}$ is defined by Equation (10.2-20) and $V$ is the voltage across the gap at $z = 0$. Thus, the total axial field within the helix is given by

$$E_z = \frac{V}{L}\sum_n M_{1(n)}\frac{I_n(\gamma_n r)}{I_n(\gamma_n a)}\epsilon^{\pm jn\theta}\epsilon^{-j\beta_n z} \qquad (11.2\text{-}3)$$

a summation of space harmonics.

---

[7]References 11.3, 11.4.

The impedance is evaluated from Equation (10.1-19) as

$$K_{-1} = \frac{\int |E_{z(-1)}|^2 dS}{2\beta^2_{-1}PS} = \frac{M_{1(-1)}{}^2 M_{2(-1)}{}^2 |V|^2}{2\beta_{-1}{}^2 L^2 P} \qquad (11.2\text{-}4)$$
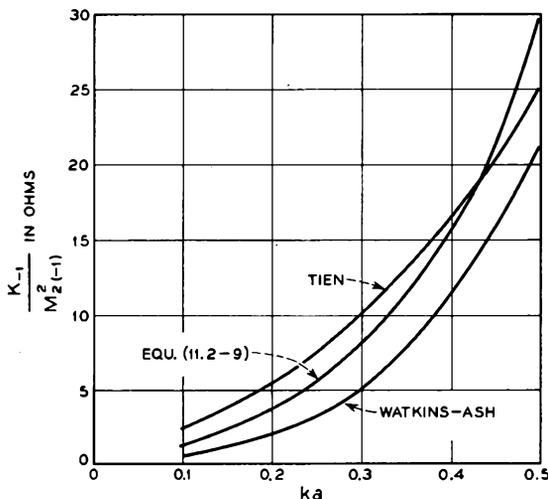


Fig. 11.2-2 Backward-wave impedance of a tape helix for $\delta/L = 1/3$. The simplified theory of this chapter is compared with the more complicated theoretical values derived by Tien (Reference 11.4) and Watkins-Ash (Reference 11.3).

where the space-harmonic amplitude is obtained from Equation (11.2-3). $M_{2(n)}{}^2$ is defined by the equation

$$M_{2(n)}{}^2 = \frac{1}{S} \int \frac{I_n{}^2(\gamma_n r)}{I_n{}^2(\gamma_n a)} dS \qquad (11.2\text{-}5)$$

where $S$ is the beam cross-sectional area, and the integral is taken over the beam cross section.

The impedance for the fundamental is given in the same manner as

$$K_o = \frac{\int |E_{zo}|^2 dS}{2\beta_o{}^2 PS} = \frac{M_{1(0)}{}^2 M_{2(0)}{}^2 |V|^2}{2\beta_o{}^2 PS} \qquad (11.2\text{-}6)$$

Combining Equations (11.2-4) and (11.2-6), we obtain

$$K_{-1} = K_o \frac{M_{1(-1)}{}^2}{M_{1(0)}{}^2} \frac{M_{2(-1)}{}^2}{M_{2(0)}{}^2} \frac{(\beta_o L)^2}{(\beta_{-1} L)^2} \qquad (11.2\text{-}7)$$

relating the beam-coupling impedances of the fundamental and $-1$ space harmonics. Since we have derived an expression for $K_o$ in the preceding chapter, this equation enables us to determine $K_{-1}$.
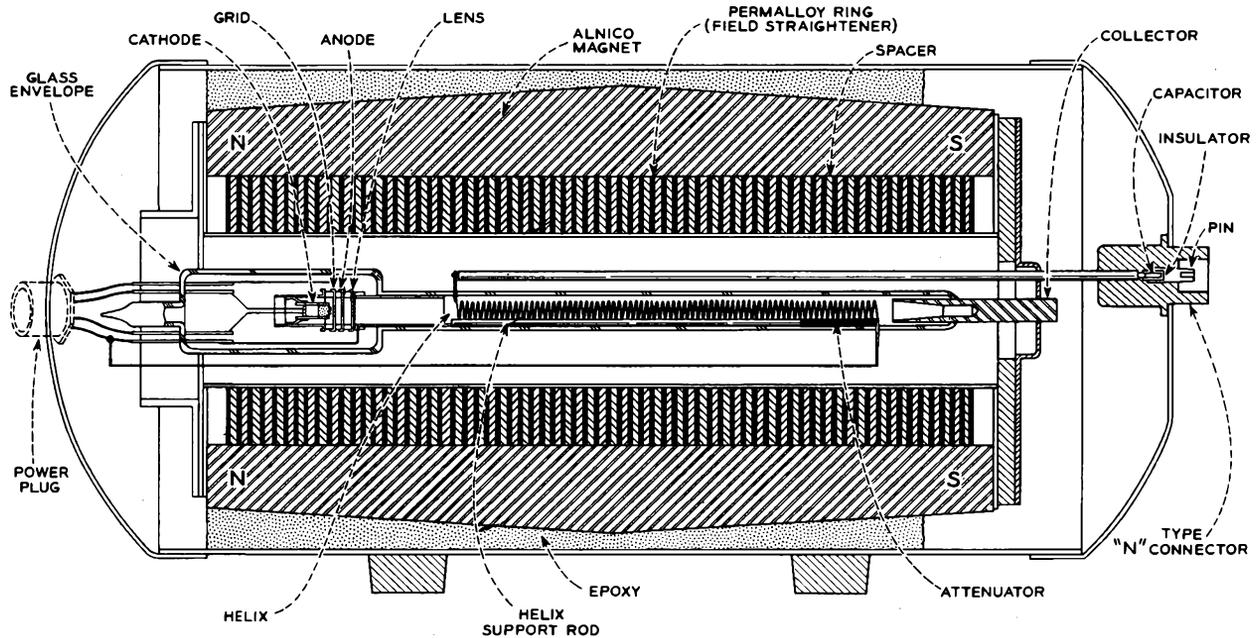
FIG. 11.2-3 SE 201 backward-wave oscillator. Since this tube is operated with the cathode at ground potential, an insulator is provided in the rf output connector to isolate the high voltage on the helix. (Courtesy of Stewart Engineering Co.)

From Figure 11-2 we have the following relationships:

$$\beta_o L = 2\pi ka$$

$$\beta_{-1} L = 2\pi(1 - ka) \tag{11.2-8}$$

These relations are inserted into Equation (11.2-7) together with Equation (10.3-22) for $K_o$ to obtain

$$K_{-1} = M_{2(-1)}{}^2 \frac{15(ka)^3}{(1 - ka)^4} \frac{\sin^2 \dfrac{\pi\delta}{L}(1 - ka)}{\sin^2 \dfrac{\pi\delta}{L}ka} \text{---ohms} \tag{11.2-9}$$

This is plotted in Figure 11.2-2 as a function of $ka$ for $\delta/L = \frac{1}{3}$. Also plotted are the results of two other analyses.[8] A comparison of these three results tends to substantiate the approximate analysis we have used.

The radial variation of the $-1$ space-harmonic field is proportional to the function $I_1(\gamma_1 r)$, plotted in Figure 10.3-2. This function goes to zero on the axis, so electrons there do not interact with the circuit. Consequent-



(a)

(b)
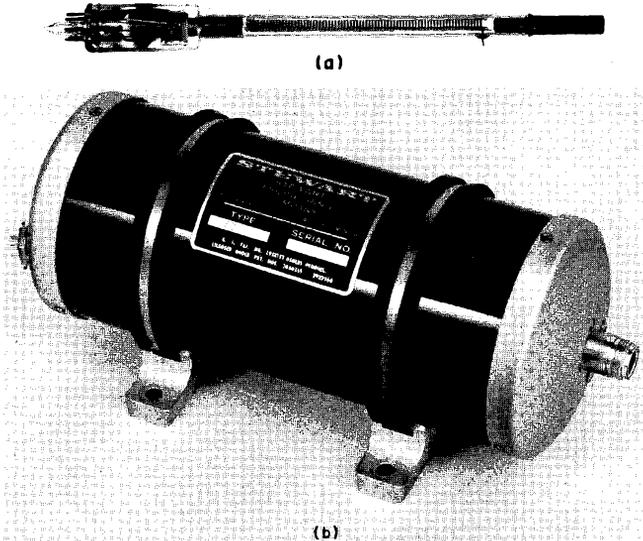
FIG. 11.2-4 SE 201 backward-wave oscillator. (a) Vacuum tube alone, without rf output coupling and focusing magnet. (b) Complete package of tube in permanent magnet. The package is 22 cm long and 11 cm in diameter. (Courtesy of Stewart Engineering Co.)

[8] *Ibid.*

ly, backward-wave oscillators are often built using hollow beams instead of solid ones.[9]

As an example of a practical backward-wave oscillator, let us consider the Stewart Engineering Company SE 201, shown in Figure 11.2-3. Photographs of the tube are shown in Figure 11.2-4. This tube delivers a minimum of 10 mw of power electronically tunable over the frequency range 7 to 12.4 Gc. The power output and helix voltage as a function of frequency are shown in Figure 11.2-5 for a beam current of 7 milliamps. The electron
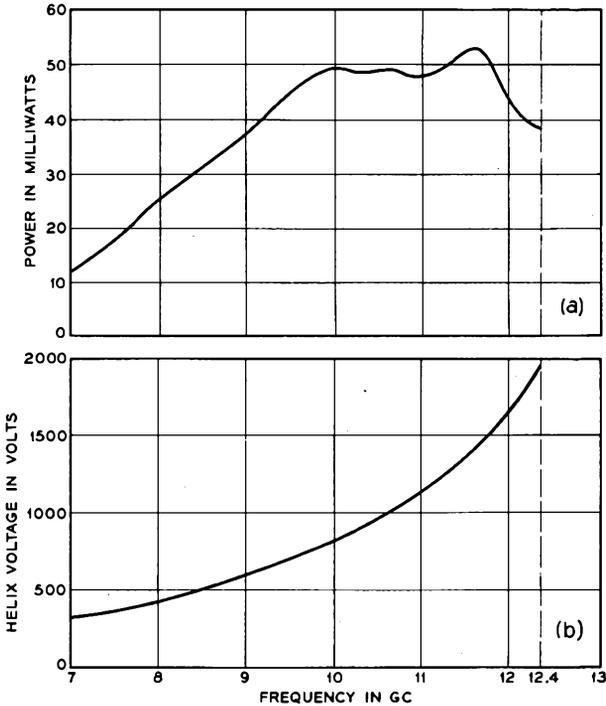


Fig. 11.2-5 Electrical characteristics of the SE 201 for a constant beam current of 7 ma. (a) Power output vs. frequency. (b) Tuning curve, helix voltage vs. frequency. (Courtesy of Stewart Engineering Co.)

beam is hollow, with an outer diameter of 2.9 mm and an inner diameter of 2.4 mm. The power may be varied by means of a control grid in the electron gun, which varies the beam current. Amplitude modulation is obtained in

---

[9]Space-charge reduction factors for hollow beams are given in Reference 9.3.

this manner.  The collector is cooled by heat conduction to the envelope of the package.

The helix is a molybdenum tape, 0.127 mm thick and 0.508 mm wide. The tape is wound into a helix with an inside diameter of 3.2 mm and a pitch of 1.06 mm.  At the highest operating frequency, $ka = 0.416$.

The tube is packaged in a permanent magnet; the complete package weighs 11 pounds.  The electron beam is focused using confined flow as described in Section 3.4(b); that is, the magnetic field is relatively uniform from the cathode to the collector.  The permalloy rings help to provide a uniform magnetic field as in the tube of Figure 10.3-7.  The gun end of the helix is connected to the center conductor of a coaxial cable for rf output power by means of a pin through the glass envelope of the tube.

A lossy material is applied at the collector end of the helix so as to provide an internal rf termination.  The pin connection at this end of the helix is for connection to the helix dc power supply.

An important characteristic of oscillators is the relative strength of the desired output signal as compared with all other spurious frequencies.  The desired signal in this tube is at least 60 db larger than the total power in all spurious signals.  Backward-wave oscillators in general produce extremely clean output signals.

The chief disadvantage of the backward-wave oscillator is its low electronic efficiency.  At the highest frequency at which the SE 201 operates the tube has an electronic efficiency of only 0.3 per cent.  On the other hand, the backward-wave oscillator provides a wider electronic tuning range than any other microwave tube.  The backward-wave oscillator has been built at higher frequencies than any other microwave tube; power outputs of a few milliwatts have been obtained at 500 Gc.

## 11.3   Backward-Wave Amplifiers

The backward-wave oscillator may be used as a backward-wave amplifier, provided that rf coupling is furnished at the collector end of the helix for application of the input signal.  The beam current is adjusted to a value below the current needed to start oscillation.  The amplified output signal is taken from the gun end of the helix.

The helix voltage is adjusted as indicated in Figure 11-2.  Amplification is obtained at the frequency of synchronism with the $-1$ space harmonic. As the signal frequency is varied from the synchronous value, the space-harmonic phase velocity departs from the beam velocity much faster than in a forward-wave amplifier.  As a result, backward-wave amplifiers have much narrower fractional bandwidths, typically 0.1 to 1 per cent.

The voltage gain of a backward-wave amplifier is given by the reciprocal

of the right-hand side of Equation (11.1-15). For a particular tube, at the beam voltage and frequency corresponding to synchronism, the right-hand side of Equation (11.1-15) is a function of the beam current, $F(I_o)$. At start oscillation it is equal to zero,

$$F(I_{ST}) = 0 \tag{11.3-1}$$

We can find an approximate expression for the gain as a function of beam current by expanding $F(I_o)$ in a Taylor series about the starting current. That is,

$$F(I_o) \cong F(I_{ST}) + \left[\frac{\partial F}{\partial I_o}\right]_{I_{ST}} (I_o - I_{ST}) \tag{11.3-2}$$

In taking the partial derivative, we make the simplifying assumption that the total change in $F$ with current is due to the change in $C$ in each exponent. Thus, we neglect the variations in the $\delta$'s and $QC$. Using this approximation, one obtains:

$$F(I_o) = \frac{I_o - I_{ST}}{I_{ST}} \frac{2\pi CN}{3I_{ST}} \left[ \frac{\delta_1(\delta_1{}^2 + 4QC)}{(\delta_1 - \delta_2)(\delta_1 - \delta_3)} \epsilon^{2\pi\delta_1 CN} + \frac{\delta_2(\delta_2{}^2 + 4QC)}{(\delta_2 - \delta_3)(\delta_2 - \delta_1)} \epsilon^{2\pi\delta_2 CN} \right.$$
$$\left. + \frac{\delta_3(\delta_3{}^2 + 4QC)}{(\delta_3 - \delta_1)(\delta_3 - \delta_2)} \epsilon^{2\pi\delta_3 CN} \right] \tag{11.3-3}$$

where all the parameters are evaluated at $I_{ST}$. The gain may thus be written as

$$\text{gain} = 20 \log \frac{1}{|F(I_o)|} \text{db} = 20 \log \frac{k'}{1 - \dfrac{I_o}{I_{ST}}} \text{db} \tag{11.3-4}$$

where $k'$ is a function of the helix loss and the space-charge parameter, as defined by the last two equations. Values of $k'$ are given in Table 11.3-1 for zero helix loss and various values of $QC$. Maximum gain is obtained for $QC$ in the neighborhood of 0.5. Equation (11.3-4) is plotted in Figure 11.3-1 for $QC = 0$.

TABLE 11.3-1

| $QC$ | $k'$ |
|---|---|
| 0 | 1.01 |
| 0.25 | 1.22 |
| 0.50 | 2.03 |
| 0.75 | 1.71 |
| 1.00 | 1.83 |
| 1.50 | 1.68 |

Examination of Figure 11.3-1 reveals that appreciable gain is achieved only for beam currents extremely close to the starting current value. In addition, slight variations in the beam current produce large fluctuations in
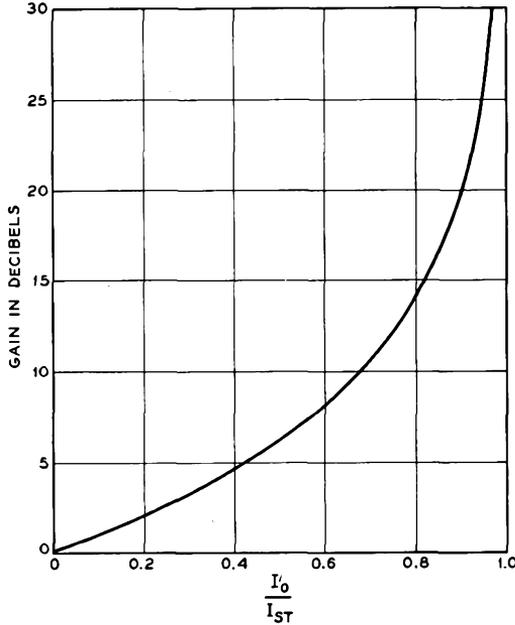


FIG. 11.3-1 Theoretical gain of a single helix backward-wave amplifier for $QC = 0$. The value of gain is the maximum value at the center of the amplification band. Appreciable amounts of gain are obtained only for beam currents just below the starting current, $I_{ST}$.

the gain at high-gain levels. This characteristic makes gain stability a serious problem. This is a consequence of operating so close to the point of oscillation. Since $I_{ST}$ generally increases with frequency, the gain at constant beam current generally decreases with frequency.

Another disadvantage of this device is the lack of discrimination against signals outside the amplifying passband. All signals may propagate along the helix from the input to the output without any attenuation other than the normal helix attenuation. This latter attenuation is kept as small as possible so as to obtain maximum gain at the operating frequency. In contrast, the circuit sever in the forward-wave amplifier provides 60 db or more of attenuation to signals outside the amplifying band.

These disadvantages are eliminated in the device shown in Figure 11.3-2,

the cascade backward-wave amplifier.[10]  This device is evolved from the
single-helix backward-wave amplifier by using two helices of equal length.
Each helix has one end terminated as shown.  This procedure is analogous
to the transition from a two-cavity to a three-cavity klystron amplifier.

The amplifier functions in the following manner.  The first helix acts as a
single-helix backward-wave amplifier, with the amplified circuit power be-
ing dissipated in a termination at the gun end of the helix.  The modulation
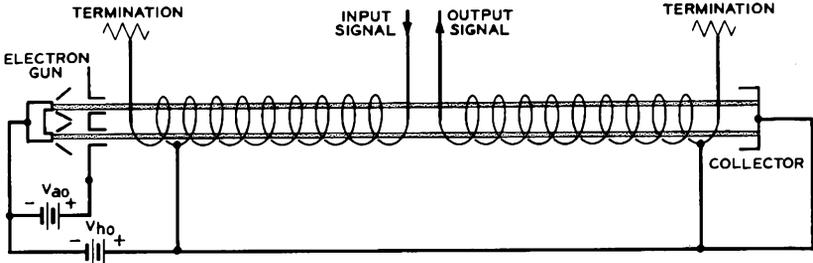


FIG. 11.3-2 The cascade backward-wave amplifier.

produced on the beam by the first helix is carried by space-charge waves to
the second helix, where backward-wave interaction produces additional
gain.  The output signal is removed from this helix as shown in Figure
11.3-2.  Because of the physical separation between the two helices, the
input and output ports of the tube are effectively isolated for frequencies
outside of the amplification band.

It may at first appear that the gain of the cascade amplifier is merely
twice the gain in db obtained on the first helix.  However, this is not the
case; the gain of the second helix is considerably larger than that of the first
helix.  This is due to the fact that the beam is premodulated upon entering
the second helix, resulting in enhanced interaction and larger overall gain.
As an example, for a beam current of one-half the starting current for either
helix, the gain in the first helix is 6 db, whereas the gain in the second helix
is 15 db.  These theoretical values are for $QC = d = 0$.  In this case, an
overall gain of 21 db is obtained for a beam current considerably below the
starting current.

Experimental curves for the gain of a particular cascade backward-wave
amplifier are shown in Figure 11.3-3.[11]  The variation of gain with beam
current is much less severe than in the case of a single-helix tube, for equal
overall gains.  One drawback of the backward-wave amplifier is evident

[10]Reference 11.5.
[11]*Ibid.*
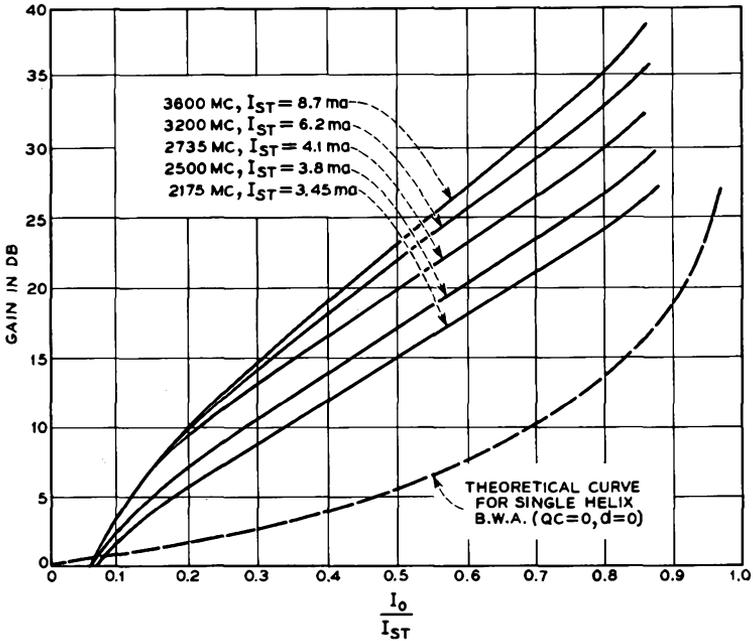
FIG. 11.3-3 Experimental measurements of gain in a cascade backward-wave amplifier. Appreciable values of gain are obtained for beam currents considerably below the starting current. The helix voltage is adjusted for maximum gain at each frequency (Reference 11.5). (Courtesy of *Proceedings IRE*)

from these curves. The maximum gain varies considerably with frequency for a constant beam current. For a beam current of 3 ma, the gain varies from 17 to 27 db for the frequencies indicated in the figure.[12] Constancy of maximum gain can be achieved by varying the beam current in conjunction with the helix voltage.

The cascade backward-wave amplifier is thus seen to overcome the most serious drawbacks of the single helix version. As a result, we have a practical device for those applications requiring a narrow-band amplifier, voltage tunable over a wide frequency range.

### PROBLEMS

11.1 An idealized, lossless slow-wave structure has a Brillouin diagram whose fundamental or zero-order space harmonic is a straight line between the points

---

[12]Thus at 2175 Mc the starting current is 3.45 ma. For a beam current of 3.0 ma, $I_o/I_{ST} = 3.0/3.45 = 0.87$, and the corresponding gain is 27 db. On the other hand, at 3600 Mc a beam current of 3.0 ma gives $I_o/I_{ST} = 3.0/8.7 = 0.34$. The corresponding gain is 17 db.

$\beta L = 0$, $\omega = \omega_o/2$ and $\beta L = \pi$, $\omega = \omega_o$. The axial component of the electric field is given by the expression

$$E_z(x,y,z) = \frac{\sqrt{P}}{10L}\left(\frac{\mu_o}{\epsilon_o}\right)^{1/4}\sum_{-\infty}^{\infty}\frac{\epsilon^{-j\beta_n z}}{n^2 + 1}$$

where

$$\beta_n = \beta_o + \frac{2\pi n}{L}$$

and $P$ is the total power flow for a propagating mode.

(a) Sketch the Brillouin diagram over the range $-2\pi \leq \beta L \leq 2\pi$. Indicate typical points of operation for traveling-wave amplifiers and backward-wave oscillators.

(b) Derive expressions for the beam-coupling impedances for fundamental and backward-wave interaction.

(c) In operation as a forward-wave amplifier the beam voltage is pulsed up to give synchronous interaction with the fundamental space harmonic. During the rise time of this voltage pulse, the electron velocity is synchronous with the phase velocity of the backward wave, and there is the possibility of backward-wave oscillations. Derive an expression for $CN$ for the backward wave as a function of the variables $\omega$ and $\beta_e$, assuming $\beta_{-1} = \beta_e$ (small $C$). The ratio of $I_o$ to $V_o^{3/2}$ remains constant during the rise time of the voltage.

(d) What is the maximum gain of the traveling-wave amplifier at the frequency $\omega = \frac{3}{4}\omega_o$, limited by the criterion that backward-wave oscillations are not produced during the rise time of the beam voltage pulse? Assume small $C$, and $QC = d = 0$. $b = 0$ for the amplifier. The circuit is perfectly matched at both ends.                    *Ans.*: 11.9 db.

11.2 Show that $M_{2(-1)}{}^2$ in Equation (11.2-9) for an annular beam in a helix is given by

$$M_{2(-1)}{}^2 = \frac{b^2[I_1{}^2(\gamma_{-1}b) - I_o(\gamma_{-1}b)I_2(\gamma_{-1}b)] - c^2[I_1{}^2(\gamma_{-1}c) - I_o(\gamma_{-1}c)I_2(\gamma_{-1}c)]}{(b^2 - c^2)I_1{}^2(\gamma_{-1}a)}$$

where $a$ is the helix radius, $b$ is the outer beam radius, and $c$ is the inner beam radius.

The following expressions will be helpful:

$$I_{-1}(\alpha r) = -I_1(\alpha r)$$

$$\int rI_1{}^2(\alpha r)dr = \frac{r^2}{2}[I_1{}^2(\alpha r) - I_o(\alpha r)I_2(\alpha r)]$$

11.3 A tape helix backward-wave oscillator is to be designed to oscillate at 9 Gc. Find the helix length and beam current necessary for oscillations to start at a beam voltage of 2500 volts, using Figures 11.1-3 and 11.1-4, assuming negligible helix loss. Assume the electron beam is a thin annular beam just grazing the helix, so that $M_{2(-1)} = 1$; assume $\delta/L = \frac{1}{3}$ so that the helix impedance is given by Equation

(11.2-9) as plotted in Figure 11.2-2. The following parameters are also given:

$$\text{Helix radius } a = 2.54 \text{ mm}$$
$$\cot \psi = 10$$
$$\text{Space-charge reduction factor } R = 0.4$$
$$\text{Beam cross-sectional area} = 4 \text{ mm}^2$$

*Ans.:* 2.06 cm, 49.6 ma.

11.4 Calculate the power output for the backward-wave oscillator of the previous problem at beam currents of 60 ma, 100ma, and 200 ma.

11.5 For large values of the space-charge parameter $QC$, the incremental propagation constants of a backward-wave oscillator are given by Equations (11.1-21).

(a) Show that the magnitudes of the convection current and the $z$ component of the electric field due to the circuit at the threshold of oscillation are given as functions of $z$ by the equations:

$$|\, i(z) \,| = |\, i(l) \,| \sin \tfrac{1}{2}\beta_e C(QC)^{-1/4}z$$

$$|\, E_{zn}(z) \,| = |\, E_{zn}(0) \,| \cos \tfrac{1}{2}\beta_e C(QC)^{-1/4}z$$

(b) From the results of part (a) show that the starting value of $CN$ is given by

$$CN = \tfrac{1}{2}(QC)^{1/4}$$

(c) If the beam current is made much larger than the value necessary to start oscillations, the backward-wave oscillator will oscillate in a higher-order mode. Assuming that this higher order mode occurs at the same frequency and beam voltage as the fundamental mode of oscillation, show from the results of part (a) that it occurs at a beam current equal to 81 times the starting value for the fundamental mode of oscillation.

## REFERENCES

Three general references on backward-wave interaction are:

11a. H. Heffner, "Analysis of the Backward-Wave Traveling-Wave Tube," *Proc. IRE* **42**, 930–937, June, 1954.

11b. A. H. W. Beck, *Space-Charge Waves and Slow Electromagnetic Waves*, Pergamon Press, Inc., New York, pp. 241–255, 1958.

11c. R. Kompfner and N. T. Williams, "Backward-Wave Tubes," *Proc. IRE* **41**, 1602–1611, November, 1953.

References covering specific items discussed in the chapter are:

11.1 H. R. Johnson, "Backward-Wave Oscillators," *Proc. IRE* **43**, 684–697, June, 1955.

11.2 R. W. Grow, and D. A. Watkins, "Backward-Wave Oscillator Efficiency," *Proc. IRE* **43**, 848–856, July, 1955.

11.3 D. A. Watkins, and E. A. Ash, "The Helix as a Backward-Wave Circuit Structure," *J. Appl. Phys.* **25**, 782–790, June, 1954.

11.4 P. K. Tien, "Bifilar Helix for Backward-Wave Oscillators," *Proc. IRE* **42**, pp. 1137–1143, July, 1954.

11.5 M. R. Currie and J. R. Whinnery, "The Cascade Backward-Wave Amplifier: A High-Gain Voltage-Tuned Filter for Microwaves," *Proc. IRE* **43**, 1617–1631, November, 1955.

# Chapter 12

## CROSSED-FIELD DEVICES

In the microwave devices we have described in previous chapters, the dc magnetic field has been used merely as a means of confining the electron beam to a given diameter. The beam focusing using the dc magnetic field and the rf interaction process have been considered as two independent problems. In crossed-field devices, on the other hand, the dc magnetic field plays a direct role in the rf interaction process. We shall find that amplifier gain, oscillator starting conditions, and power output are all functions of the dc magnetic field in crossed-field devices.

All crossed-field devices have one property in common. In the rf interaction space, there exists both a dc magnetic field and a dc electric field, perpendicular to each other. As an introduction to these devices, let us study the motion of a single electron in such a field configuration.

Figure 1.2-2 in Chapter 1 shows a rectangular coordinate system in a region of crossed electric and magnetic fields. The electric field intensity $E$ is directed in the negative $y$ direction, and the magnetic flux density $B$ is directed in the negative $z$ direction. We wish to study the motion of an electron which leaves the origin at time $t = 0$ with an initial velocity given by the three components $u_{ox}$, $u_{oy}$, and zero. Since there is no force in the $z$ direction, the motion of the electron is constrained to the $x$-$y$ plane.

The electron motion is given by the solution of Equations (1.2-10). The solutions for the instantaneous velocity and position are

$$u_x = \frac{dx}{dt} = \frac{E}{B} - \left(\frac{E}{B} - u_{oz}\right)\cos \omega_c t + u_{oy} \sin \omega_c t \qquad (12\text{-}1)$$

$$u_y = \frac{dy}{dt} = \left(\frac{E}{B} - u_{oz}\right)\sin \omega_c t + u_{oy} \cos \omega_c t \qquad (12\text{-}2)$$

and

$$x = \frac{E}{B}t - \frac{1}{\omega_c}\left(\frac{E}{B} - u_{oz}\right)\sin \omega_c t + \frac{u_{oy}}{\omega_c}(1 - \cos \omega_c t) \qquad (12\text{-}3)$$
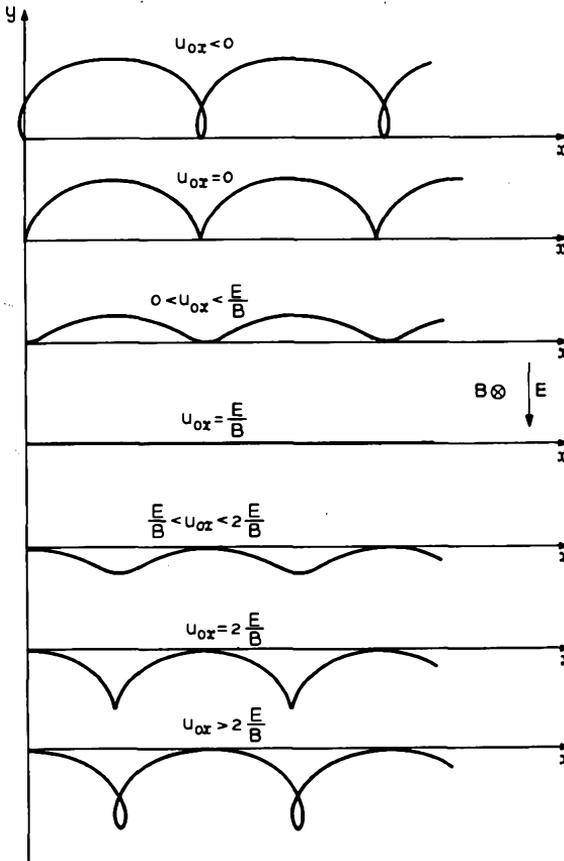
FIG. 12-1 Electron motion in a region of uniform, crossed dc electric and magnetic fields. Trajectories are shown for electrons with a zero $y$ component of velocity at $x = 0$ and with various values for the $x$ component.

$$y = \frac{1}{\omega_c}\left(\frac{E}{B} - u_{ox}\right)(1 - \cos \omega_c t) + \frac{u_{oy}}{\omega_c} \sin \omega_c t \qquad (12\text{-}4)$$

where

$$\omega_c = \frac{eB}{m}$$

is the cyclotron frequency in radians per second. The electron trajectories determined by these equations are shown in Figure 12-1 for $u_{oy} = 0$ and various values of $u_{ox}$.

Solutions for non-zero values of $u_{oy}$ are similar; the trajectories are obtained from Figure 12-1 by translating the $x$-$y$ coordinate system so that the origin falls at other points on the trajectories. In this case, a value of $u_{ox}$ given in the figure is interpreted as the $x$ component of velocity along the trajectory for which the $y$ component is zero.

The solutions exhibit several interesting properties.

1. The motion is cycloidal except for the case $u_{ox} = E/B$, $u_{oy} = 0$, for which the electron moves in a straight line. In this particular case, the force due to the electric field is exactly balanced by the force due to the magnetic field.

2. The frequency of the cycloidal motion is given by the cyclotron frequency and is thus a function only of the magnetic field and the charge-to-mass ratio for an electron.

3. The average drift velocity in the $x$ direction is given by $E/B$, completely independent of the initial velocities $u_{ox}$ and $u_{oy}$.

Item number 3 is of particular interest. An approximate picture of the electron motion in a region of crossed electric and magnetic fields can be obtained by neglecting the cyclic motion of the electron and considering only the average motion. This average motion is perpendicular to both $E$ and $B$ and has a magnitude $E/B$. It is a function only of the fields at the location of the electron and does not depend on the electron's past history. Thus, the electron tends to follow an equipotential line of the electric field.

The basic elements of a crossed-field device are shown in Figure 12-2.[1]
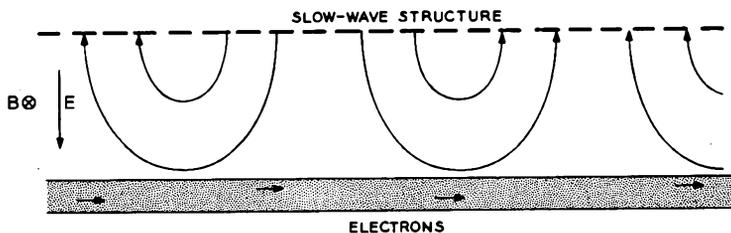


FIG. 12-2 Basic elements of a crossed-field device. Rf electric field lines due to a traveling-wave on the slow-wave structure are pictured at a particular instant of time.

Electrons are traveling parallel to a slow-wave structure. There is a dc electric field $E$ directed as shown away from the slow-wave circuit and a dc magnetic field $B$ directed into the page. The dc motion of the electrons is

[1]This linear model is useful for analysis. Practical crossed-field devices are usually constructed in a cylindrical configuration, as we shall see later.

assumed to be as in Figure 12-1 for $u_{oz} = E/B$. The slow-wave structure is assumed to be propagating a wave in the direction of the electron flow with a phase velocity equal to $E/B$. Thus, some form of synchronous inter-action is evidently possible between the electrons and the electromagnetic wave.

We can view the interaction from a frame of reference traveling to the right with a velocity $E/B$. In this reference frame both the electromagnetic wave and the electrons appear motionless until we include the effect of the rf electric field on the electron motion.

The trajectories which result when we include the effect of the rf electric field are shown in Figure 12-3. These trajectories are derived using the
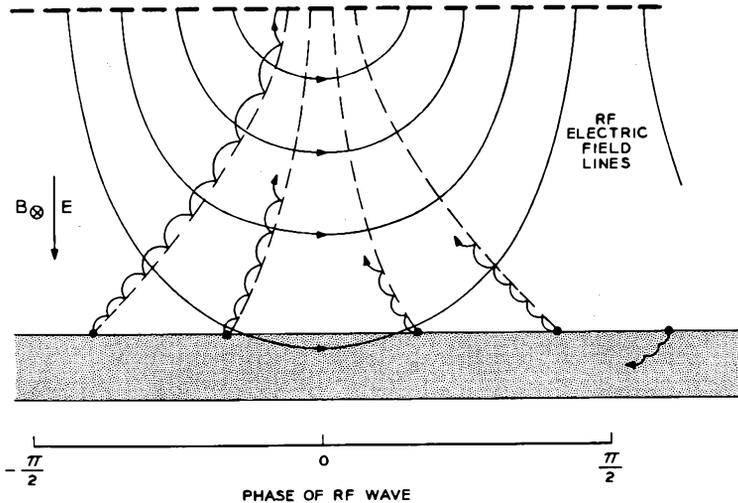


Fig. 12-3 Electron trajectories in a reference frame synchronous with the rf travel-ing wave. Space-charge forces are neglected. The average drift motion of the electrons is seen to be along lines orthogonal to the electric field lines.

following reasoning. In the reference frame the rf electric field appears stationary. Thus, we may consider the electron motion to be determined by static fields, the electric field having spatial dependence but no time dependence. If we assume that the electric field's spatial dependence is slow and continuous, we may use the solutions previously derived for a constant field, which are illustrated in Figure 12-1. Thus, the motion has a cycloidal appearance with a net drift of the electron along a path orthogonal to the electric field lines. The cycloidal amplitude and period vary mono-tonically corresponding to the monotonically increasing strength of the

electric field as the slow-wave structure is approached. For simplicity, space-charge forces are neglected.

The average kinetic energy of the electron at any position is determined from the average drift velocity, given by

$$u_{avg} = \frac{E_T}{B} \tag{12-5}$$

where $E_T$ is equal to the sum of the dc electric field and the rf electric field at the position under consideration. Since the rf field is strongest at the slow-wave structure, and decays monotonically away from it, the electron continuously gains average kinetic energy as it drifts toward the slow-wave structure. On the other hand, the electron loses dc potential energy of an amount given by the product: $eE \times$ (displacement of the electron in the direction of the slow-wave circuit). *This dc potential energy loss is greater than the increase in kinetic energy of the electron; the balance of the energy is delivered to the rf field by the mechanism of induced currents.* That is, each electron continuously induces a current in the slow-wave circuit; the power delivered by each electron to the circuit is proportional to the dot product of the total electron velocity and the rf electric field.

Electrons in phases given by $\pi/2 < |\theta| < \frac{3}{2}\pi$ (see Figure 12-3) move away from the slow-wave structure rather than toward it. These electrons extract a net amount of energy from the rf field. However, since the rf electric field is much weaker for these electrons than for the group which moves toward the circuit, there is still a large net transfer of energy to the rf field. In a magnetron oscillator the electrons which move away from the slow-wave circuit return to the cathode and serve the useful purpose of heating it so that less heater power is required. They also liberate secondary electrons which constitute a large percentage of the total emission.

Approximate trajectories which neglect the cycloidal motion are indicated by dotted lines in Figure 12-3. These trajectories are constructed by drawing lines orthogonal to the $E$-field lines. The average velocity in the reference frame at any point is $1/B$ times the rf electric field at that point.

One important characteristic of crossed-field devices should be noted at this time. As the electrons interact with the rf field, they remain in synchronism with it until they are finally collected on the slow-wave structure. Crossed-field devices are said to convert potential energy into rf energy, since the electron's kinetic energy in the direction of propagation of the rf wave is essentially unchanged. This is in contrast to the klystron and the traveling-wave tube, wherein rf energy is derived from the longitudinal kinetic energy of the electrons. The fact that the electrons remain locked to (or phase focused by) the rf wave, even at extremely large signal levels, leads to very high electronic efficiencies in crossed-field devices. Efficiencies of 70 per cent and higher have been obtained.

In the following sections we shall examine three classes of crossed-field devices, the magnetron, the crossed-field amplifier, and the M-carcinotron. Because of the inherent complexity of the two-dimensional electron motion in crossed fields, we shall restrict ourselves to somewhat idealized models of the physical phenomena. This idealized approach is also necessitated by the fact that there is not complete agreement among individuals working in the field as to the correct approach to some of the basic theoretical problems.

## 12.1   Magnetrons

A schematic illustration of the functional parts of a magnetron oscillator is shown in Figure 12.1-1. A cylindrical cathode is surrounded by an
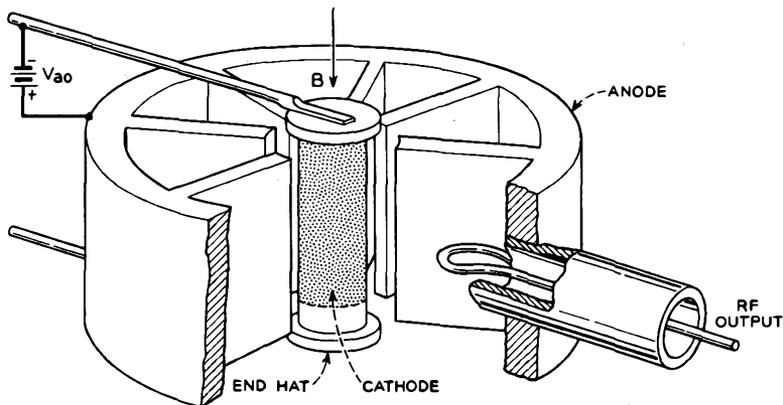


FIG. 12.1-1  A magnetron oscillator.

anode consisting of a re-entrant slow-wave structure. A uniform magnetic field, parallel to the cylindrical axis, fills the region between cathode and anode.

Electrons emitted from the cathode would flow radially to the anode were it not for the magnetic field, which causes the electron trajectories to be bent. In the crossed-field geometry of Figure 1.2-2, the maximum distance an electron can penetrate toward the anode is obtained from Equation (12-4). We see that this distance is inversely proportional to the strength of the magnetic field. Neglecting the effect of the rf field, a sufficiently large magnetic field causes most of the electrons to turn back toward the cathode before reaching the anode. This results in a dense, turbulent sheath of electrons rotating about the cathode.

The slow-wave structure propagates an electromagnetic wave around its circumference with the average angular speed of the electrons. This pro-

duces an interaction of the type described in connection with Figure 12-3. Electrons flow to the anode in "spokes" which are synchronous with the rf wave. See Figure 12.1-2. The re-entrant nature of the slow-wave structure provides a built-in feedback mechanism which permits oscillations to
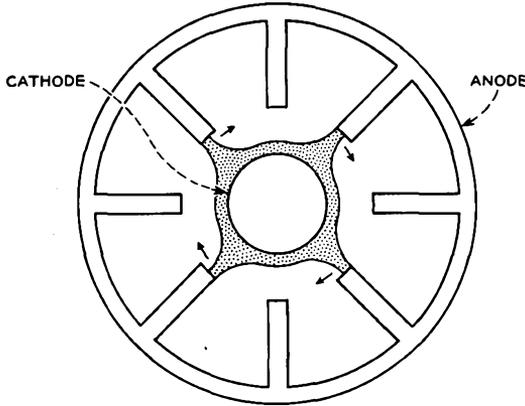


FIG. 12.1-2 Approximate shape of the space-charge configuration in the interelectrode space of a magnetron. This configuration rotates, giving the appearance of the spokes and hub of a wheel.

occur. One period of the slow-wave structure bounded by two vanes is referred to as a resonator. One of the resonators of the slow-wave structure is coupled inductively to a loop formed from the center conductor of a coaxial cable. The coaxial cable in turn delivers the rf output signal to the load.

The end hats on the cathode structure are at cathode potential but nonemitting. They serve to prevent the electrons from spreading axially out of the interaction space.

The slow-wave structure in the magnetron of Figure 12.1-1 is essentially a folded version of the structure of Figure 8.7-4. The Brillouin diagram for this structure was presented in Figure 8.7-8. The portion of this diagram currently of interest is replotted in Figure 12.1-3, where $\varphi$ is the phase shift per period.

Since the circuit is closed on itself, or "re-entrant," oscillations are possible at frequencies for which the total phase shift around the circuit is an integral multiple of $2\pi$ radians. Thus, if $N$ is the total number of periods of the slow-wave structure, oscillations can occur at the frequencies corresponding to

$$\varphi = \frac{2\pi n}{N} \qquad (12.1\text{-}1)$$
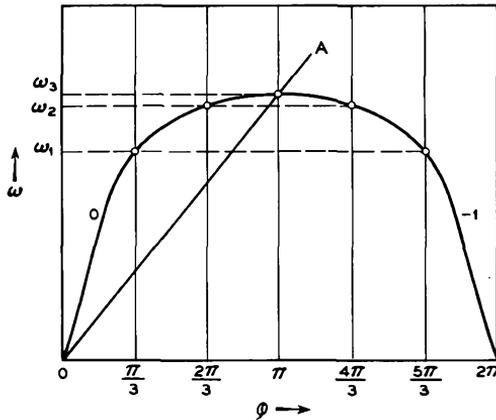
FIG. 12.1-3 Brillouin diagram for the slow-wave structure of a magnetron. Points at which oscillations are possible are indicated for a re-entrant structure 6 periods long.

where $n$ is an arbitrary integer. These values of phase shift and the corresponding frequencies are indicated by points in Figure 12.1-3 for $N = 6$.

In order for oscillations to be produced at these points, two conditions must be fulfilled. First, the anode voltage $V_{ao}$ must be adjusted so that the average rotational velocity corresponds to the phase velocity of the slow-wave structure. Second, the beam-coupling impedance must be sufficiently large. The operating points shown in Figure 12.1-3 are termed modes of oscillation, with the mode identified by stating the phase shift per cavity.

Although oscillations are theoretically possible in any of the modes given by Equation (12.1-1), the magnetron is always designed for operation in the $\pi$ mode because this allows the strongest possible interaction. Thus, in Figure 12.1-3, the magnetron is designed to oscillate at a radian frequency $\omega_3$.

Magnetrons are usually operated with the voltage $V_{ao}$ pulsed on and off, so as to produce short pulses of rf signal. As the voltage is increased toward the value corresponding to the $\pi$ mode, it must necessarily pass through the points of synchronism with other modes ($4\pi/3$ and $5\pi/3$ modes in Figure 12.1-3). Since spurious oscillations in these modes are undesirable, it is necessary to design the slow-wave structure carefully to avoid them. Either the beam-coupling impedance must be too low for oscillations, or else the losses associated with these modes must be large enough to prevent oscillation buildup in the time available.

The magnetron as described thus far oscillates at only one frequency, the frequency corresponding to the $\pi$ mode of the slow-wave structure.

Such fixed-frequency magnetrons have very limited application; some form of tunability is highly desirable. Tuning of the magnetron is accomplished by means of a mechanical perturbation of the slow-wave structure. This causes a change in the Brillouin diagram with a consequent change in the $\pi$ mode frequency $\omega_3$. One way of tuning the structure of Figure 12.1-1 is to insert a uniform array of metal pins into the inductive regions of the resonators, one pin in each resonator. This causes a reduction in the inductance of each resonator so that the frequency $\omega_3$ is increased, the increase being proportional to the depth of penetration. Practical schemes for electronic tuning of magnetrons have not been developed at the time of writing.

Oscillation in the $\pi$ mode is unique in that the electrons are actually synchronous with two space harmonics simultaneously. These space harmonics are: First, the fundamental space harmonic corresponding to power flow in the direction of the electron flow, and second, the $-1$ space harmonic corresponding to power flow in the opposite direction. Since the circuit is cut off at this frequency, the net power flow in either direction is zero. Thus the two power flows described above are equal and opposite. Space-harmonic analysis reveals that the two space harmonics described above are excited to equal amplitude at $\pi$ phase shift per period, for equal power flow in the two directions. Thus, the summation of the two synchronous waves and their space harmonics corresponds to a standing wave as in any microwave cavity at resonance. Interaction phenomena is thus conveniently described in terms of the cavity $Q$'s — the unloaded $Q$, external $Q$, and loaded $Q$ — as defined in Chapter 9 for the klystron.

In the case of the $\pi$ mode, the synchronous electromagnetic wave described in connection with Figures 12-2 and 12-3 is interpreted as the superposition of the two synchronous space harmonics described above.

### (a)   The Hull Cutoff Condition

We are ready to consider some of the quantitative aspects of magnetron interaction. First, we shall consider a model of the electron motion which includes space-charge forces but excludes all rf fields. This is known as the static magnetron problem. Instead of the re-entrant cylindrical structure used in most magnetron oscillators, we shall consider the linear magnetron structure shown in Figure 12.1-4. This model exhibits all of the essential characteristics of the cylindrical version, without the attendant mathematical complexity. A uniform dc magnetic field $B$ is directed into the page. The solution we wish to demonstrate is one in which the electrons all move in straight lines from left to right with velocity $E/B$. We have already noted that this type of motion is possible for an individual elec-

tron, so that the problem that remains is that of finding the potential and charge distribution in the interelectrode space when many electrons are present. Because of the planar geometry, we must assume that when the electrons reach the right-hand end of the model of Figure 12.1-4, they disappear and reappear at the left-hand end.
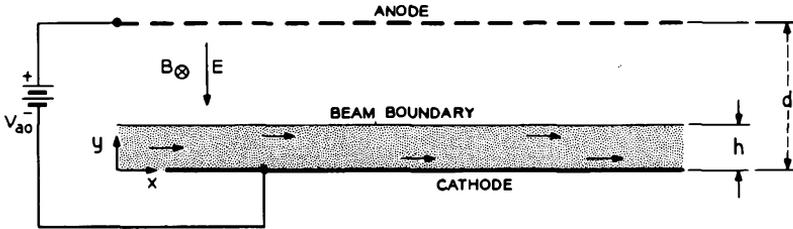


FIG. 12.1-4  Linear model of a magnetron. Rf fields may be precluded by replacing the slow-wave structure by a uniform conducting plane.

The dc voltage $V_{ao}$ applied between the slow-wave structure (anode) and the cathode causes a dc electric field $E(y)$ directed away from the anode.

Using the coordinate axes shown in Figure 12.1-4, the electron motion is assumed to be such that all electrons flow to the right with a velocity given by

$$u_x = \frac{E(y)}{B} = \frac{1}{B}\frac{dV}{dy}$$

$$u_y = u_z = 0 \tag{12.1-2}$$

where $V$ is the electrostatic potential, a function of $y$. The dc electric field $E(y)$ is a function of the $y$ coordinate within the beam because of the space charge. The beam lies within a region extending a distance $h$ from the cathode, where $h$ is known as the hub thickness.[2]

Next, we solve for the electrostatic potential within the beam. If the electrons originated at the cathode, energy conservation would indicate that

$$\tfrac{1}{2}mu_x^2 = eV \tag{12.1-3}$$

Eliminating $u_x$ between the last two equations gives us

$$\left(\frac{dV}{dy}\right)^2 = 2\frac{e}{m}B^2V \tag{12.1-4}$$

---

[2] This stream of electrons is referred to as the hub, since in the cylindrical magnetron it has the appearance of the hub of a wheel.

This differential equation may be rearranged as

$$\frac{1}{\sqrt{2\frac{e}{m}B}}\frac{dV}{\sqrt{V}} = dy \tag{12.1-5}$$

in which form it may be integrated directly, with the solution

$$V = \frac{eB^2}{2m}y^2 \tag{12.1-6}$$

for the potential within the beam. The constant of integration has been eliminated assuming $V = 0$ at $y = 0$.

The potential and electric field at the hub surface are obtained from Equation (12.1-6) as

$$V(h) = \frac{e}{2m}B^2h^2 \tag{12.1-7}$$

and

$$E_y = -\frac{dV}{dy} = -\frac{e}{m}B^2h \tag{12.1-8}$$

From Section 8.3, we know that the electric field must be continuous across the beam boundary. Therefore, Equation (12.1-8) gives the electric field in the region $h \leq y \leq d$. The potential at the anode is thus obtained from Equation (1.1-6) as

$$\begin{aligned}
V_{ao} &= -\int_0^d E_y dy \\
&= -\int_0^h E_y \, dy - \int_h^d E_y dy \\
&= V(h) + \frac{e}{m}B^2h(d - h) \\
&= \frac{e}{m}B^2h(d - h/2) \tag{12.1-9}
\end{aligned}$$

where use has been made of Equation (12.1-7). This equation allows one to calculate the hub thickness for various values of magnetic field and anode voltage.

From Poisson's Equation, Equation (1.4-9), one may calculate the charge density within the hub from Equation (12.1-6) as the constant value

$$\rho = -\frac{e\varepsilon_o}{m}B^2 \tag{12.1-10}$$

From Equation (12.1-9) we may calculate the threshold anode voltage for which the anode draws current. This occurs when the hub thickness $h$ is exactly equal to the spacing $d$. This voltage is known as the Hull cutoff voltage, given by

$$V_{ao} = \frac{1}{2}\frac{e}{m}B^2 d^2 \qquad (12.1\text{-}11)$$

This relationship is plotted in Figure 12.1-5. For operating points below the parabola of this figure, the anode theoretically will not draw current.
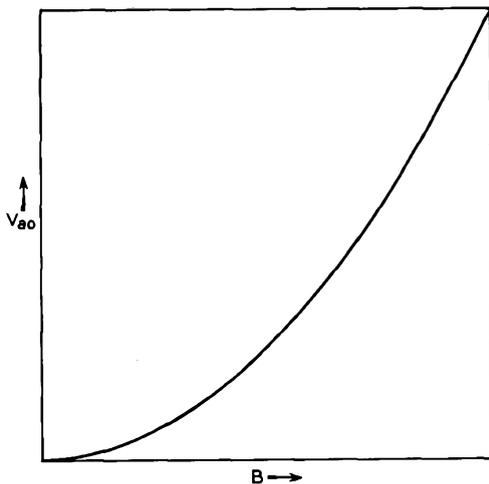


FIG. 12.1-5 Hull cutoff voltage for a magnetron. For voltages below the parabola, the anode theoretically will draw no current.

It is interesting to note that Equation (12.1-11) is a consequence of the law of conservation of energy. Thus the same result is obtained for any other assumed set of electron trajectories or distribution of charge density, provided that they are independent of the $x$ coordinate. This can be shown very easily as follows. An electron leaving the cathode with zero velocity of emission arrives at the anode with a kinetic energy given by

$$\tfrac{1}{2}m[u_x^2 + u_y^2]_{y=d} = eV_{ao} \qquad (12.1\text{-}12)$$

From a single integration of the first of Equations (1.2-10), we have

$$u_x = \frac{e}{m}By \qquad (12.1\text{-}13)$$

where use has been made of the initial condition $u_x = 0$ at $y = 0$. Com-

bining the last two equations, we obtain

$$V_{ao} = \frac{1}{2}\frac{e}{m}B^2d^2 + \frac{1}{2}\frac{m}{e}[u_y{}^2]_{y=d} \qquad (12.1\text{-}14)$$

Cutoff oocurs for $u_y = 0$ at the anode, obtaining Equation (12.1-11).

A typical experimental curve[3] of anode current as a function of anode voltage is shown in Figure 12.1-6 for a cylindrical magnetron with a constant value of magnetic field. Also shown is the corresponding curve for zero magnetic field. The latter curve exhibits the three-halves power dependence of current on voltage as given by Equation (4.1-10) up to the point where the emission becomes temperature limited. The experimental curve with a non-zero magnetic field is characterized by three segments as indicated in the figure.
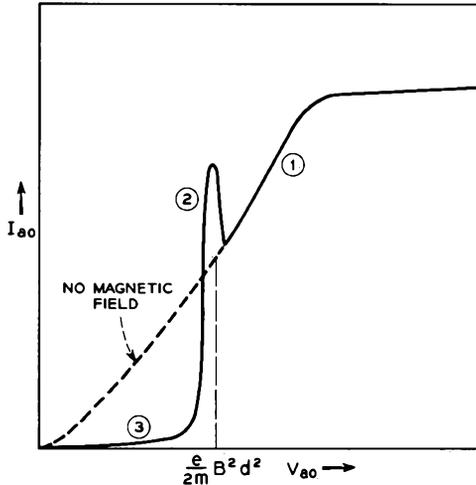


FIG. 12.1-6 Experimental variation of anode current with anode voltage for a constant magnetic field in a magnetron without an rf circuit.

Segment 1 occurs for very large values of anode voltage. Here the electric field force is so strong compared with the magnetic field force that the electrons travel nearly radially to the anode; the situation is essentially the same as for zero magnetic field.

Segment 2 occurs for voltages approximately equal to the Hull cutoff value. In this region the current increases rapidly as the voltage is raised, reaches a maximum, and then abruptly falls to the value corresponding to

[3]Reference 12.1.

Segment 1. Segment 3 shows a small, non-zero anode current for voltages below the Hull cutoff value.

The existence of anode current for voltages below the Hull cutoff value is in direct violation of the Hull cutoff theory, indicating the inadequacy of this theory. A number of hypotheses have been proposed to explain this behavior.[4] Certainly, the idealized electron flow assumed in Figure 12.1-4 is not readily attained. Perturbations of this flow give rise to space-charge fields which tend to push certain of the electrons to the anode and others back to the cathode.[5] This instability results in anode current, emission of secondary electrons from the cathode and additional heating of the cathode due to the electron back bombardment, and the generation of excess rf noise. The overall picture is complicated by the spread of electron emission velocities and electron-electron collisions in the hub. These effects are most pronounced for voltages near the Hull cutoff value. For this condition the amount of rotating space charge is the greatest, and hence the conditions for space-charge instabilities are most favorable.

In spite of the experimental contradictions to the electron flow model of Figure 12.1-4, we shall assume in subsequent discussion that the model provides a first approximation to the unperturbed electron flow in the magnetron. The presence of the electromagnetic field due to a slow-wave structure will act as a strong perturbation on this model, an effect which will be considered in the next section. We may consider the perturbations discussed in the previous paragraph to be negligible in comparison.

(b)   *The Hartree Condition*

The Hull cutoff condition (Figure 12.1-5) determines the anode voltage necessary to obtain non-zero anode current as a function of the magnetic field in the absence of electromagnetic fields. In this section we shall consider a second condition which must be imposed on $B$ and $V_{ao}$. If interaction is to take place between the electrons and the rf wave propagating along the slow-wave circuit, the electron velocity parallel to the circuit must be approximately equal to the phase velocity of the wave. This leads to a relationship between $B$ and $V_{ao}$ known as the Hartree condition.

The Hartree condition is determined as follows. Electron flow is assumed to exist as in Figure 12.1-4. The hub thickness $h$ is of the order of one fourth the cathode-anode spacing $d$. Now if a weak electromagnetic wave is present, electron flow to the anode is possible, as in Figure 12-3, provided that the electrons are in synchronism with the wave. Since the electrons

<hr />

[4]Reference 12a, Vol. 1, pp. 179-326, 359-366.

[5]This perturbed flow produces space-charge forces in the $x$ direction so that Equation (12.1-13) is no longer valid.

at the hub surface ($y = h$) have the greatest velocity, the threshold anode voltage for oscillation is determined by the condition of synchronism of these electrons with the rf wave.

The velocity of electrons at the hub surface is obtained from Equations (12.1-2) and (12.1-8) as

$$u_x(h) = \frac{e}{m} Bh \qquad (12.1\text{-}15)$$

For synchronism, this is equal to the phase velocity of the slow-wave structure,

$$\frac{\omega}{\beta} = \frac{e}{m} Bh \qquad (12.1\text{-}16)$$

where $\beta$ is the circuit propagation constant at the desired operating point. For the $\pi$ mode, $\beta$ is given by

$$\beta L = \pi \qquad (12.1\text{-}17)$$

where $L$ is the period of the circuit. At the onset of oscillations we may neglect the presence of space charge exterior to the hub. Thus, Equation (12.1-9) determines the anode voltage as a function of the hub thickness. The hub thickness may be eliminated between Equations (12.1-9) and (12.1-16), obtaining

$$V_{ao} = \frac{\omega B d}{\beta} - \frac{m}{2e} \frac{\omega^2}{\beta^2} \qquad (12.1\text{-}18)$$

This is the Hartree condition.

The Hartree voltage is plotted as a function of the magnetic field in Figure 12.1-7, together with the Hull cutoff curve. The usual |operating region of a magnetron is shown in the figure. Below this region, neither oscillation nor appreciable anode current is obtained. Above this region, anode current independent of the rf properties of the circuit would be obtained. Typically, the operating point is quite close to the Hartree line in order to avoid spurious oscillations in higher voltage modes. Thus, the hub thickness is a small fraction of the cathode-anode spacing.

Within the region of operation, raising the anode voltage causes increased anode current and rf power output. The velocity of the electrons at the surface of the hub is obtained from the relations of Section 12.1(a) as

$$u_x(h) = \frac{eBd}{m} \left[ 1 - \sqrt{1 - \frac{2mV_{ao}}{eB^2 d^2}} \right] \qquad (12.1\text{-}19)$$

As the anode voltage is raised above the Hartree value, this velocity increases above the synchronous value. Nonetheless, the forces due to the rf field tend to pull these electrons back into synchronism with the wave.
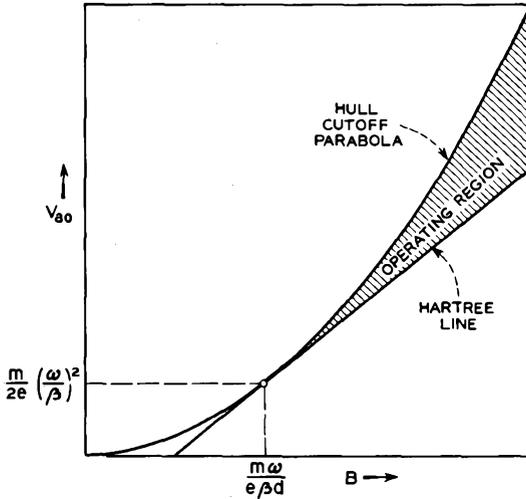
FIG. 12.1-7 Region of operation of a magnetron, between the Hartree line and the Hull cutoff parabola.

This behavior is shown in the computed trajectories of Figure 12.1-8.[6] The trajectories correspond to electrons leaving the hub surface with average velocities in the $x$ direction given by Equation (12.1-19) for three different anode voltages. Space-charge forces are neglected. Figure 12.1-8(a) shows the trajectories at the Hartree voltage; electrons leave the hub in synchronism with the wave and remain in synchronism with it up to the anode. Figure 12.1-8(b) shows the trajectories for a somewhat higher anode voltage, representing typical magnetron operation. Although the electrons start out faster than the wave, they are phase focused back by the rf field forces so that the spoke is nearly centered about zero phase. Zero phase corresponds to the maximum rf electric field and consequently to the position of strongest interaction. Figure 12.1-8(c) shows the trajectories at still a higher value of anode voltage. Once again we see the focusing action of the rf field. The noncycloidal trajectories of Figure 12.1-8 are approximate in that the initial acceleration at the hub surface is arbitrarily set to zero. The cycloidal trajectories in Figure 12.1-8(c) illustrate the trajectories resulting without this approximation.

As the level of oscillation is increased, the average center of the spoke tends to advance in phase with respect to the wave. This has the effect of an inductive loading of the circuit, and the frequency of oscillation

[6]J. Feinstein, Reference 12a, Vol. 1, pp. 554–579.
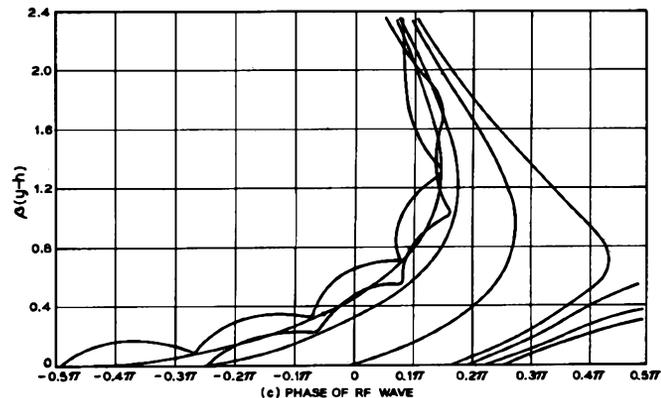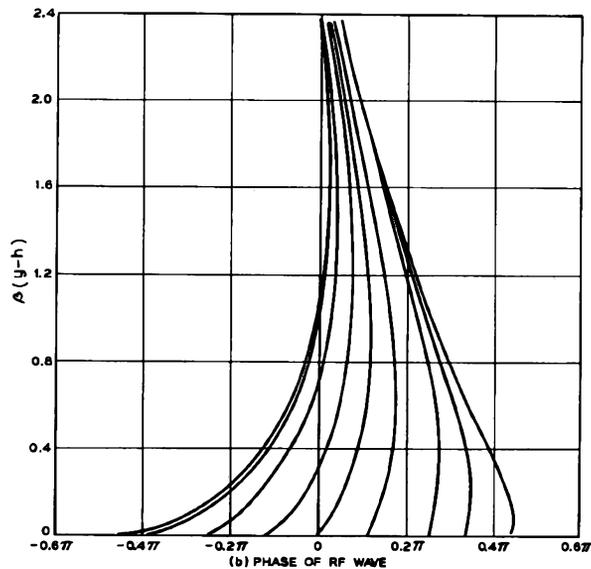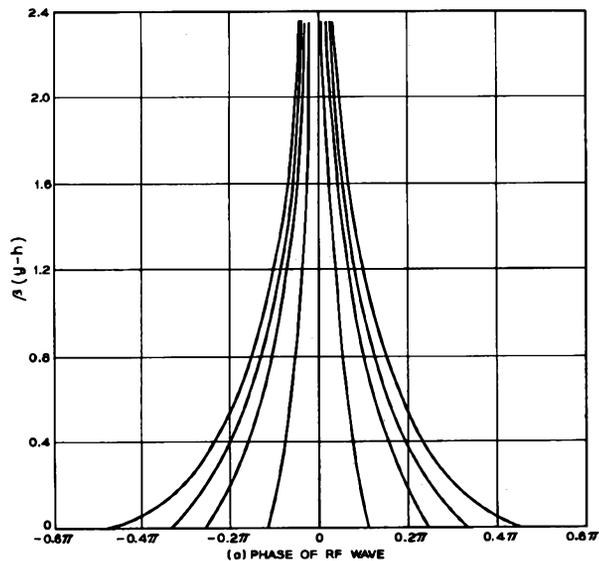
FIG. 12.1-8 Computed electron trajectories in a reference frame synchronous with the rf wave on the circuit, neglecting space-charge forces. Zero phase corresponds to a maximally retarding field, as in Figure 12-3. (a) Trajectories for operation at the Hartree voltage; the oscillation level is low. (b) Trajectories at a typical rf level of operation. (c) Trajectories at a very high rf level of operation (Reference 12a). (Courtesy of Academic Press, Inc.)

consequently increases slightly. This effect is known as frequency pushing. Its magnitude is too small to provide a practical means for electronic tuning

Further calculations of the type shown in Figure 12.1-8 indicate that there is a minimum or threshold value for the rf electric field in order to obtain a stable current spoke for each value of anode voltage greater than the Hartree value. Below this threshold the electrons do not form a spoke but instead are returned to the hub.

### (c)  Power Output and Efficiency

The efficiency of the magnetron is the ratio of the rf power output to the product of anode voltage and anode current. It may be expressed in turn as the product of the electronic efficiency $\eta_e$ and the circuit efficiency $\eta_c$. The electronic efficiency expresses the percentage of the dc or pulsed input power which is converted into rf power on the slow-wave structure. The circuit efficiency, on the other hand, determines the percentage of this rf power which is delivered to the load exterior to the tube.
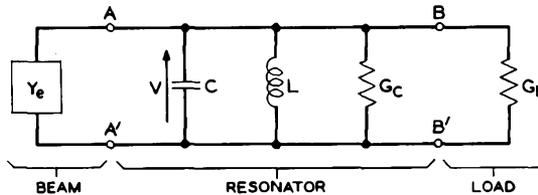


FIG. 12.1-9 Equivalent circuit for one resonator of a magnetron. The symbols are defined in the text.

The circuit efficiency may be evaluated in terms of the equivalent circuit presented in Figure 12.1-9. Each resonator of the slow-wave structure is taken to comprise a separate resonant circuit, with equivalent circuit parameters $C$, $L$, and $G_c$. Thus for vane resonators, as shown in Figure 12.1-1, $C$ is the capacitance at the vane tips, including fringing capacitance. From a knowledge of $C$ and the resonant frequency, the inductance $L$ can be determined. The resonant frequency of each individual resonator is taken as the $\pi$ mode frequency of the whole anode structure. A conductance $G_c$ is chosen so as to give the correct value of rf loss on the anode circuit per resonator. The unloaded $Q$ of the magnetron is thus

$$Q_u = \frac{\omega_o C}{G_c} \qquad (12.1\text{-}20)$$

where $\omega_o$ is the radian resonant frequency.

The rf voltage across the vane tips (terminals $AA'$) is given by the phasor $V$. $G_L$ is the equivalent load conductance per resonator chosen in such a

manner as to give the power delivered to the load per resonator. The anode circuit and load are connected at terminals $BB'$ by various components such as coupling loops, transformers, lengths of transmission line, etc., which usually introduce reactive effects and loss by their presence. For simplicity, we neglect these effects. The external $Q$ is defined by

$$Q_e = \frac{\omega_o C}{G_L} \tag{12.1-21}$$

The loaded $Q$ of the circuit is given by

$$Q_l = \frac{\omega_o C}{G_e + G_L} \tag{12.1-22}$$

and the $Q$'s are related by Equation (9.1-33). It is to be noted that the values of the $Q$'s are the same when defined for the anode structure as a whole, since they represent ratios of energy stored to energy dissipated.

The circuit efficiency is given by

$$\eta_c = \frac{G_L}{G_L + G_c} \tag{12.1-23}$$

which may be written in terms of the $Q$'s as

$$\eta_c = \frac{Q_l}{Q_e} = \frac{1}{1 + \dfrac{Q_e}{Q_u}} \tag{12.1-24}$$

Maximum circuit efficiency is obtained when the magnetron is loaded very heavily by the load, that is, for $G_L \gg G_c$. However, heavy loading makes the tube operation quite sensitive to the load, which is undesirable in some cases. Because of the reactive effects associated with the load and output coupling circuit, load changes can cause shifts in the frequency of oscillation. Therefore, the ratio of $Q_l/Q_e$ chosen is often a compromise between the conflicting requirements for high circuit efficiency and frequency stability.

The electronic admittance due to the electrons is represented by $Y_e$ in Figure 12.1-9. This admittance has a negative conductance representing the power generated and a small inductive susceptance whose variation with oscillation level produces frequency pushing. This latter effect was discussed in the previous section.

The electronic efficiency may be obtained from the following formula:

$$\eta_e = \frac{P_{\text{gen}}}{V_{ao} I_{ao}} = \frac{V_{ao} I_{ao} - P_{\text{lost}}}{V_{ao} I_{ao}} \tag{12.1-25}$$

where $P_{\text{gen}}$ is the rf power induced into the anode circuit by the electrons, $V_{ao}$ is the anode voltage, $I_{ao}$ is the anode current, and $P_{\text{lost}}$ is the power

lost in the tube other than the rf losses in the slow-wave structure and output circuit.

$P_{\text{lost}}$ consists of two main parts. The largest loss is the anode dissipation, given by the kinetic energy of the electrons striking the anode. A much smaller loss is due to the back bombardment of the cathode; electrons in unfavorable phases gain energy from the rf field and return to the cathode with an excess of kinetic energy. The back bombardment is usually 3 to 10 per cent of the input power.[7]

The kinetic energy of the electrons reaching the anode may be estimated in the following manner. From trajectory plots as in Figure 12.1-8 we conclude that it is approximately correct to assume that all the electrons arrive at the anode in zero phase of the rf wave. Thus, if $E_{\text{max}}$ is the maximum value of the $x$ component of the synchronous space harmonic of the rf electric field, the average $y$ component of the velocity of the electrons on striking the anode is approximately

$$u_y(d) = \frac{E_{\text{max}}}{B} \qquad (12.1\text{-}26)$$

using Equation (12-5). The average value of the $x$ component of velocity is approximately given by the synchronous value,

$$u_x(d) = \frac{\omega}{\beta} \qquad (12.1\text{-}27)$$

The rf power generated by the electrons is given by

$$
\begin{aligned}
P_{\text{gen}} &= I_{ao}V_{ao} - P_{\text{lost}} \\
&= I_{ao}V_{ao} - I_{ao}\frac{m}{2e}\left[u_x{}^2(d) + u_y{}^2(d)\right] \\
&= I_{ao}V_{ao} - I_{ao}\frac{m}{2e}\left[\frac{\omega^2}{\beta^2} + \frac{E_{\text{max}}{}^2}{B^2}\right] \qquad (12.1\text{-}28)
\end{aligned}
$$

where use has been made of Equations (12.1-26) and (12.1-27). The power loss associated with cathode back-bombardment is neglected for the time being.

From the equivalent circuit of Figure 12.1-9 we have the following expression for the power generated by the electrons,

$$P_{\text{gen}} = \frac{1}{2}N \mid V \mid^2 \frac{\omega C}{Q_l} \qquad (12.1\text{-}29)$$

where $N$ is the total number of resonators comprising the anode structure, and $V$ is the rf voltage appearing across the resonator gap. The amplitudes

---

[7]A. M. Clogston, Reference 12b, p. 525.

of the space harmonics may be determined by space-harmonic analysis as in the preceding chapters. The amplitude of the $x$ component of the electric field at the circuit is given by

$$E_{\max} = \frac{M_1 |V|}{L} \qquad (12.1\text{-}30)$$

where $M_1$ is the gap factor for the $\pi$ mode, computed from Equation (10.2-20), and $L$ is the center-to-center spacing of the vane tips. The power generated may thus be written as

$$P_{\text{gen}} = \frac{NL^2 \omega C}{2M_1^2 Q_l} E_{\max}^2 \qquad (12.1\text{-}31)$$

$E_{\max}^2$ may be eliminated between Equations (12.1-28) and (12.1-31) and the result solved for $P_{\text{gen}}$. One then obtains the following expression for the electronic efficiency:

$$\eta_e' = \frac{P_{\text{gen}}}{I_{ao}V_{ao}} = \frac{1 - \dfrac{m\omega^2}{2eV_{ao}\beta^2}}{1 + \dfrac{I_{ao}}{B^2}\dfrac{mM_1^2 Q_l}{eNL^2\omega C}} \qquad (12.1\text{-}32)$$

where the prime indicates that the cathode back-bombardment power loss has been neglected. Using this formula, one may evaluate the influence of the various magnetron parameters on the electronic efficiency. Since $\omega/\beta$ is approximately $E/B$, both numerator and denominator approach unity as the magnetic field is increased. Thus, magnetrons are designed to operate at as high a magnetic field as can be conveniently obtained in order to achieve high electronic efficiency.

Decreasing $Q_l$ would increase the electronic efficiency but decrease the circuit efficiency. Also, a low value of $Q_l$ makes it difficult to start the oscillations in the desired mode. The value of $Q_l$ is thus a compromise between the values for best overall efficiency and ease of starting oscillations.

In the following section we shall describe a typical magnetron, the Western Electric 7208B. The parameters of this magnetron which appear in Equation (12.1-32) are:

| | |
|---|---|
| $f = 16.5$ Gc | $Q_u = 4000$ |
| $V_{ao} = 17.5$ kv | $C = 0.15$ pf |
| $I_{ao} = 16.8$ amps | $N = 32$ |
| $B = 0.74$ webers/meter$^2$ | $L = 0.94$ mm |
| $Q_l = 1025$ | $M_1^2 = 0.762$ |

Using Equation (12.1-32) one calculates

$$\eta_e' = 64\%$$

If we assume that the cathode back-bombardment power loss is 6 per cent of the input power, we obtain the net electronic efficiency,

$$\eta_e = 58\%$$

From Equations (9.1-33) and (12.1-24), the circuit efficiency is calculated as

$$\eta_c = 74\%$$

resulting in an overall efficiency of

$$\eta = \eta_c\eta_e = 43\%$$

an excellent value of efficiency for an oscillator at such a high frequency. This corresponds to an output power of 126 kw.

### (d)   Description of the Western Electric 7208B Magnetron

The Western Electric 7208B magnetron is shown in Figure 12.1-10.[8] This tube is operated with the anode voltage pulsed on and off so that short
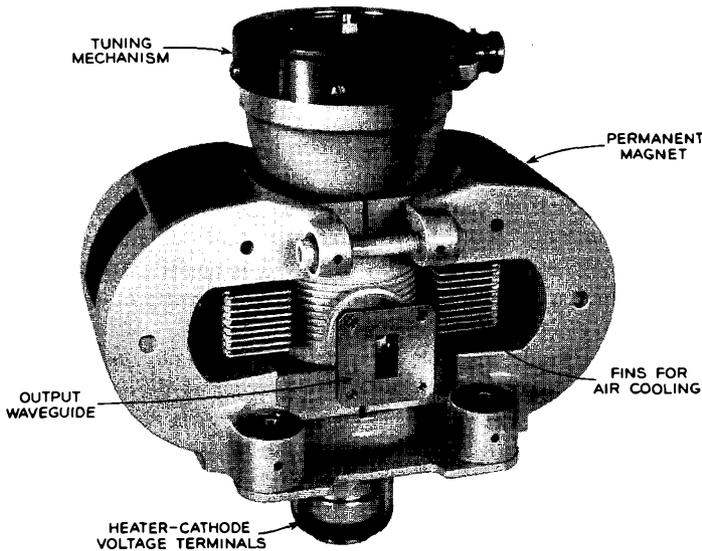


Fig. 12.1-10 Western Electric 7208B magnetron, complete with permanent magnet. The complete unit weighs 14 pounds and is 20 cm high.

pulses of rf power are produced. The rf output frequency is mechanically tunable over the range 15,500 to 17,500 Mc. Typical operating characteristics are listed in Table 12.1-1.

[8]Reference 12.2.

TABLE 12.1-1. W.E. 7208B TYPICAL OPERATING CHARACTERISTICS

Frequency, Mc . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15,500 to 17,500
Peak power output, kw . . . . . . . . . . . . . . . . . . . . . . . . 125
Peak anode voltage, kv . . . . . . . . . . . . . . . . . . . . . . . 17.5
Peak anode current, amps . . . . . . . . . . . . . . . . . . . . . 19
Current pulse duration, microsec . . . . . . . . . . . . . . . . 3
Duty cycle . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 0.001
Magnetic field, webers /m² . . . . . . . . . . . . . . . . . . . . . 0.74
Pulling figure (VSWR = 1.5:1), Mc . . . . . . . . . . . . . 6

The pulling figure is defined as the total variation in frequency when the phase of a mismatched load is varied through all values. A voltage-
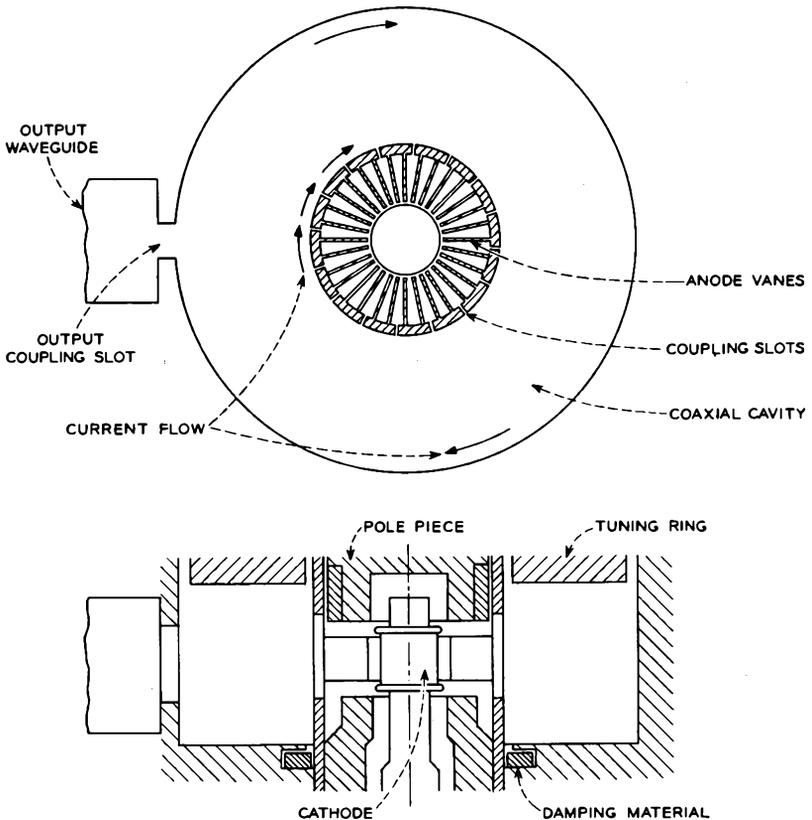


FIG. 12.1-11 Functional drawing of the Western Electric 7208B.

standing-wave ratio (VSWR) of 1.5 to 1 is equivalent to a mismatch such that 4 per cent of the output power is reflected back into the tube.

The maximum duty cycle of the 7208B is 0.001, where

$$\text{duty cycle} = \frac{\text{average power output}}{\text{peak power output}} \quad (12.1\text{-}33)$$

Another useful formula for the duty cycle is

$$\text{duty cycle} = (\text{pulse length}) \times (\text{pulse repetition frequency})$$
$$(12.1\text{-}34)$$

Most magnetrons are designed to be pulse operated. This form of operation is particularly useful in radar systems. Since the magnetron anode is grounded, pulsed operation is achieved by applying negative, rectangular-shaped voltage pulses to the cathode.

The 7208B magnetron is of a type known as a coaxial magnetron because of the particular anode slow-wave structure used. A functional drawing of the tube is shown in Figure 12.1-11. The anode consists of a cylinder with vanes extending radially toward the cathode. This anode cylinder also forms the center conductor of a coaxial resonator. Every other resonator in the vane array is coupled to the coaxial resonator by means of a slot in the common wall.

Since the vanes are approximately a quarter wavelength long in the radial direction, the impedance seen at the slot looking toward the vane tips is very small. Thus, circumferential current may flow unimpeded on the



SECT. AA

E LINES ──────▶
H LINES ────▶
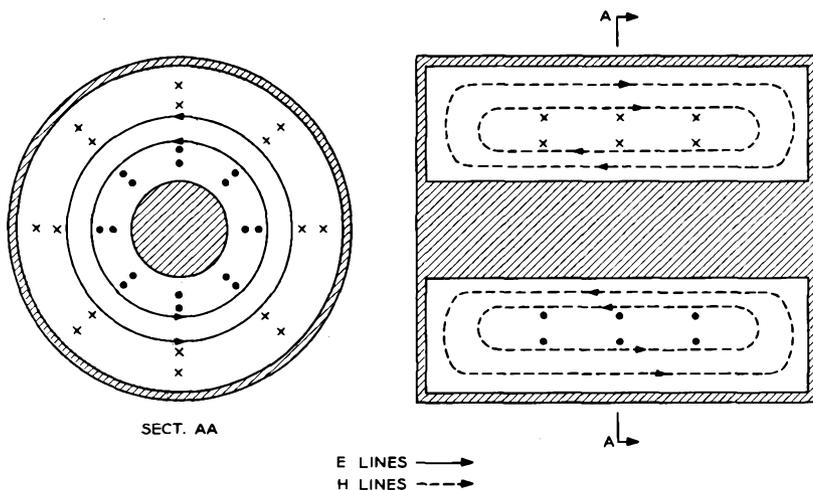
Fig. 12.1-12 Field patterns of the $TE_{011}$ mode in a coaxial resonator.

outer surface of the anode in much the same way as if it were unslotted. We may therefore consider the coaxial cavity to have the same field patterns as for the case of a solid cylindrical center conductor,

The desired mode of operation for the coaxial resonator is the $TE_{011}$ mode, whose field patterns are shown in Figure 12.1-12. Since the wall currents are perpendicular everywhere to the magnetic field lines, we see that the field patterns for the $TE_{011}$ mode correspond to circumferential current flow on all the wall surfaces. There is no current flow in either the radial or axial direction. Since there is no current flow across from the coaxial cylinders to the end plates, a gap may exist between these surfaces without disturbing the field patterns. This permits the incorporation of a tuning ring, as shown in Figure 12.1-11, so that the length of the coaxial resonator may be varied. Since the resonant frequency of the cavity increases as the cavity length is decreased, this provides a means for tuning the cavity.

The currents entering the vane array through the coupling slots have the same phase in each of the slotted resonators. Hence, the fields in each of the slotted resonators are in phase. Currents are induced in the unslotted resonators by virtue of the mutual coupling between adjacent resonators. These induced currents are 180 degrees out of phase with the currents in the slotted resonators. Hence the field components in adjacent resonators are 180 degrees out of phase. This produces the desired $\pi$-mode field configuration.

Because of the tight coupling of the vane array to the high-$Q$ coaxial cavity, the $\pi$-mode frequency is essentially that of the coaxial cavity alone. The unloaded $Q$ of the 7208B is 4000 at midband. Rf power output is coupled from the coaxial cavity to a waveguide by means of a coupling slot as shown in Figure 12.1-11. This results in a midband loaded $Q$ of approximately 1025.

Pole pieces are built right into the tube so as to concentrate the magnetic flux in the interaction space. The horseshoe-shaped permanent magnets provide a magnetic flux density of 0.74 webers/meter$^2$ in this region. The tube, complete with magnets, weighs 14 pounds and is 20 cm high.

The damping material indicated in Figure 12.1-11 damps out resonant modes of the cavity with current flow patterns across the junctions of cylindrical surfaces and end walls. As indicated above, the $TE_{011}$ mode does not fall into this category, and consequently it is not damped by this lossy material. If this damping material were not present, troublesome oscillations could occur in modes other than the desired one.

A magnetron performance chart is an instructive manner of presenting the important operating characteristics. Such a chart is shown in Figure 12.1-13 for the 7208B. Primarily, the chart is a plot of anode voltage vs.

22

200KW

175 KW

20

150KW

125KW

42 % EFF

100KW

44%
EFF

75KW

POWER OUTPUT   50KW

18     = 25KW

40% EFF

PEAK VOLTAGE IN KV

7750
GAUSS

38%EFF

7400
GAUSS

36%EFF

16

14

6700
GAUSS

12

6200
GAUSS

0
0        4        8        12       16       20       24       28
PEAK CURRENT IN AMPERES
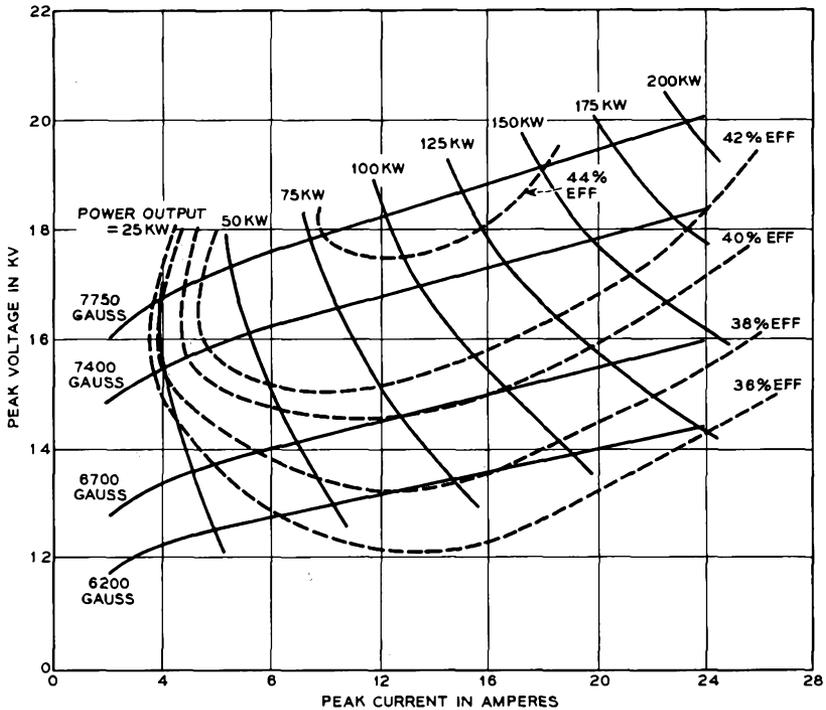
Fig. 12.1-13 Performance chart of the Western Electric 7208B at a frequency of
16.5 Gc. The usual operating magnetic field is 7400 gauss, furnished by the per-
manent magnets.

anode current at a fixed frequency, with magnetic field as a parameter.
Variation of the magnetic field is achieved by operating the tube in an
electromagnet. Contours of constant power output and constant efficiency
are also included. As the voltage is raised, the anode current and power
output increase rapidly. The power output increases almost in direct pro-
portion to the power input. The upper limit on power output is usually
reached when arcing commences in the tube due to the high electric fields.

As the anode voltage is increased, the efficiency first increases and then
decreases, for constant magnetic field. Equation (12.1-32) demonstrates
this behavior. For small anode current, the efficiency goes up as $V_{ao}$ in-
creases. Physically, this corresponds to a decreasing percentage of the total
energy associated with the synchronous velocity of the electrons. As
the anode current $I_{ao}$ becomes appreciable, the efficiency begins to fall off.
This corresponds to the increasing percentage of the electron kinetic
energy associated with velocity directed toward the anode.

Magnetrons provide a compact, high-efficiency source of rf power with an operating voltage which is relatively low in comparison with other tubes of similar power output. Because of cathode back-bombardment, anode erosion due to electron bombardment, and other factors, magnetrons tend to have shorter life than other microwave tubes. Magnetrons have a lower signal-to-noise ratio than other microwave tubes; however, for most power tube applications, this is not important. The noisiness is primarily due to the excitation of other modes of the resonant system. Unlike the reflex klystron and the backward-wave oscillator, where the frequency is voltage tunable, the frequency of the magnetron is determined primarily by the resonant frequency of the anode structure; hence, stability of the output frequency in the magnetron is attained without precise regulation of the supply voltage. On the other hand, the requirement of mechanical tuning in the magnetron prohibits the use of the magnetron in applications where extremely rapid tuning is desired.

## 12.2  Crossed-Field Amplifiers

The crossed-field amplifier[9] is very similar to the magnetron oscillator, both in internal construction and external appearance. The only important difference is that the anode slow-wave structure of the amplifier does not close on itself. Instead, the two ends of the circuit are both connected to separate external transmission lines, one for the input and one for the output rf signal.

A schematic drawing of the crossed-field amplifier is shown in Figure 12.2-1. An applied voltage and dc magnetic field are provided as in the magnetron oscillator. Spokes of current are formed by the rf electric fields. These spokes rotate



Fig. 12.2-1 Schematic drawing of a crossed-field amplifier.

in a clockwise direction around the cathode. For forward-wave inter-action, rf power enters at terminals $BB'$, flows around the circuit in a

[9]This discussion is limited to crossed-field amplifiers having a continuous cathode with a re-entrant beam. Crossed-field amplifiers have also been constructed where the electron beam is injected into the interaction region and then collected (as in Figure 12.3-1); however, the latter device has not been as widely used as the device considered here.

clockwise direction, and exits at terminals $AA'$. The current spokes travel in synchronism with the circuit wave, inducing currents in phase with those of the circuit wave, thus causing it to grow. This interaction mechanism is like that of the magnetron oscillator, except that the circuit wave grows with distance whereas it is of constant amplitude in the oscillator.

Alternatively, backward-wave interaction is obtained when the rf power is introduced at terminals $AA'$ and removed at terminals $BB'$. This is actually the mode of operation for the particular circuit shown in Figure 12.2-1.[10] This circuit consists of a two-wire line with connections made alternately to the vanes, as shown. It is a fundamental backward-wave structure.[11]
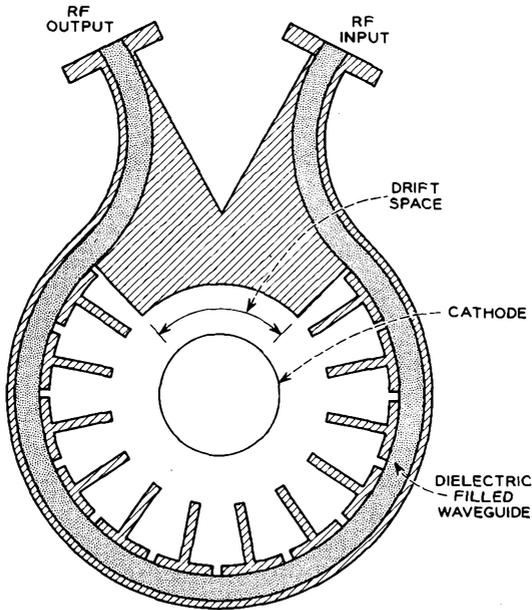


FIG. 12.2-2 A crossed-field amplifier with a drift region between the input and output ends of the rf circuit.

Although the circuit is non-re-entrant, the beam is re-entrant. As a result, the crossed-field amplifier could oscillate if the loop gain were high enough. This limits the gain to about 10 or 15 db in high-power amplifiers

[10]Reference 12.3.
[11]Slow-wave structures are characterized as either fundamental forward or fundamental backward wave structures, depending upon whether the space harmonic of greatest amplitude is a forward or a backward wave.

with a circuit completely around the circumference, as in Figure 12.2-1. The spoke moves from the output end of the circuit to the input end, phased so as to produce positive feedback and high efficiency. This phasing condition limits the bandwidth to the order of 10 to 15 per cent.

Some crossed-field amplifiers are built with the circuit encompassing only a fraction of the total circumference as illustrated in Figure 12.2-2.[12] This results in a significant drift space for the electron beam between the two ends of the circuit. In this drift space, the electron spokes are considerably dispersed by the space-charge forces, thus reducing the feedback mechanism of the electron stream. As a result, gains of the order of 15 to 20 db are possible. Since loop phasing of the electron stream is not critical, this type of crossed-field amplifier is capable of wider bandwidths. On the other hand, the reduction of positive feedback results in a lower electronic efficiency.

The anode slow-wave structure presented in Figure 12.2-2 consists of a waveguide periodically coupled by slots to each resonator of the anode vane array. The waveguide is filled with dielectric so that the phase velocity is slowed down to the synchronous velocity of interaction. Interaction is with the fundamental forward-wave space harmonic.

Crossed-field amplifiers are designed with the rf electric field at the input strong enough to form a complete current spoke of the type shown in Figure 12.1-8(a). Thus, fully formed spokes are present at all positions along the circuit. Under this condition both theoretical and experimental evidence indicates that the power induced into the circuit by each spoke and the dc current per spoke are nearly constant, independent of both the spoke location and the rf electric field at that point.

Each current spoke induces two circuit waves, one traveling toward the input, the other toward the output. However, only the waves traveling toward the output add in phase. The waves traveling toward the input tend to cancel each other so that the net power out of the input port is very small. Since each spoke induces equal power into the circuit, the circuit wave grows with a power increasing linearly with distance from input to output.

Since the circuit wave growth is linear, increasing gain from 10 db to 20 db corresponds to increasing circuit length by a factor of ten. In contrast, exponential wave growth achieves this gain increase with a doubling of the circuit length. This is another reason why crossed-field amplifiers typically exhibit low values of gain.

The total power generated in a given crossed-field amplifier is independent of the rf power input, so long as the input power exceeds the thresh-

[12]J. Feinstein and R. J. Collier, Reference 12a, Vol. 2, pp. 211–222.

old value for spoke stability at the input. The power generated can be increased only by increasing the anode voltage and current. This behavior is illustrated in Figure 12.2-3. Neglecting circuit attenuation, the
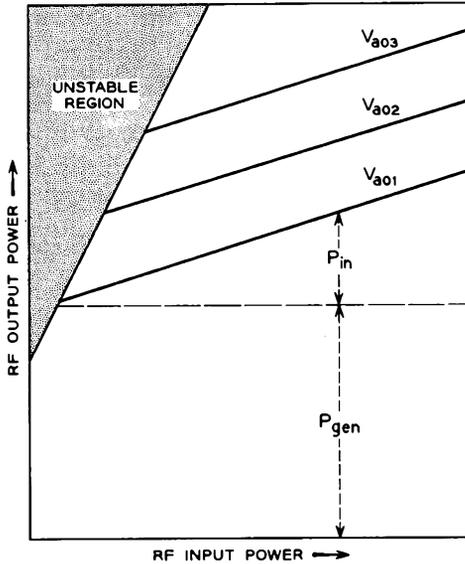


FIG. 12.2-3 Rf output power as a function of rf input power, with anode voltage as a parameter. The rf power generated $P_{gen}$ and the rf input power $P_{in}$ are indicated for the lowest anode voltage. $V_{ao1}$, $V_{ao2}$, and $V_{ao3}$ indicate successively increasing anode voltages.

output power is given by the sum of the input power and the power generated (which is constant for constant anode voltage). Thus, the amplifier gain is a function of the power input. That is,

$$\text{gain} = \frac{P_{in} + P_{gen}}{P_{in}} = 1 + \frac{P_{gen}}{P_{in}} \qquad (12.2\text{-}1)$$

Therefore, the device is not a linear amplifier but rather is termed a saturated amplifier.

If the rf drive is insufficient to form a stable current spoke at the input, no gain is produced and no power is generated at the desired frequency. Instead spurious oscillations are produced which may be attributed to the re-entrant nature of the beam and the inherent instability of electrons in crossed fields. These spurious oscillations become "locked out" when the tube is amplifying normally at the input frequency.

Since high efficiency is one of the principal attributes of the crossed-field amplifier, let us derive an expression for this quantity in terms of the various tube parameters.

The efficiency may be defined as the product of the electronic efficiency $\eta_e$ and the circuit efficiency $\eta_c$. The electronic efficiency is defined as in Equation (12.1-25). The overall efficiency is defined as

$$\eta = \frac{P_{\text{out}} - P_{\text{in}}}{V_{ao}I_{ao}} \qquad (12.2\text{-}2)$$

where $P_{\text{out}}$ and $P_{\text{in}}$ are the rf power ouput and input for the tube, respectively. It is necessary to include the latter quantity in the definition of efficiency since it is usually an appreciable percentage of the output power. For example, if the gain of the tube is 6 db, then $P_{\text{in}}$ would be 25 per cent of $P_{\text{out}}$. This sizable contribution must be subtracted from the output power in order to obtain an accurate evaluation of the efficiency of conversion of dc to rf energy. The circuit efficiency is thus obtained from Equations (12.1-25) and (12.2-2) as

$$\eta_c = \frac{\eta}{\eta_e} = \frac{P_{\text{out}} - P_{\text{in}}}{P_{\text{gen}}} \qquad (12.2\text{-}3)$$

Since the power generated per unit length is constant, the output power is given by

$$P_{\text{out}} = P_{\text{in}}\epsilon^{-2\alpha l} + \int_0^l \frac{P_{\text{gen}}}{l}\epsilon^{-2\alpha(l-x)}dx$$

$$= P_{\text{in}}\epsilon^{-2\alpha l} + \frac{P_{\text{gen}}}{2\alpha l}(1 - \epsilon^{-2\alpha l}) \qquad (12.2\text{-}4)$$

where $\alpha$ is the attenuation constant of the circuit, and $l$ is the circuit length. Introducing this expression into Equation (12.2-3) results in the following expression for the circuit efficiency:

$$\eta_c = \left[\frac{1}{2\alpha l} - \frac{P_{\text{in}}}{P_{\text{gen}}}\right](1 - \epsilon^{-2\alpha l}) \qquad (12.2\text{-}5)$$

The term $P_{\text{in}}/P_{\text{gen}}$ becomes negligible for high gain devices.

Next we calculate an expression for the electronic efficiency. Starting with Equation (12.1-25) an analysis is made similar to that for the magnetron oscillator. It is assumed that the input signal is sufficiently strong for spoke stability and that the rf power grows linearly with distance along the circuit. We further assume that the dc current per spoke is a constant, independent of the rf level at any point. The loss due to back-bombardment is not considered at this time. We assume that the circuit runs com-

pletely around the circumference, as in Figure 12.2-1. We shall show later how the result may be modified to apply to circuits that do not.

The energy lost by the electrons on striking the anode is equal to their kinetic energy and consists of two parts, corresponding to the two velocity components given by Equations (12.1-26) and (12.1-27). The velocity given by Equation (12.1-26) increases with position along the circuit since the rf field is increasing. The rf field is related to the power flow at any point by the definition of the beam-coupling impedance at the circuit,

$$ K = \frac{E_{\max}^2}{2\beta^2 P} \tag{12.2-6} $$

The value of impedance is taken at a specific point, rather than being averaged across the beam as in Equation (10.1-19). The point at which it is taken is adjacent to the anode circuit, corresponding to a maximum value of the $x$ component of the synchronous space harmonic, $E_{\max}$.

Using Equations (12.1-26) and (12.2-6), the power lost per spoke due to the motion toward the anode is given at any position by

$$ P_s = I_{so} \frac{m}{e} \frac{\beta^2 K P}{B^2} \tag{12.2-7} $$

where $I_{so}$ is the dc current per spoke, and $P$ is the power flow on the circuit at the position considered. Since the power varies linearly with position, the average loss over the circuit length is

$$ P_{s\ \text{avg}} = I_{so} \frac{m}{e} \frac{\beta^2 K}{2B^2}(P_{\text{in}} + P_{\text{out}}) \tag{12.2-8} $$

The total power loss for all the spokes is obtained using Equations (12.1-27), (12.2-1), and (12.2-8) as

$$ P_{\text{lost}} = \frac{1}{2} I_{ao} \frac{m}{e} \left(\frac{\omega}{\beta}\right)^2 + I_{ao} \frac{m}{e} \frac{\beta^2 K}{2B^2}\left(\frac{\mathcal{G}+1}{\mathcal{G}-1}\right) P_{\text{gen}} \tag{12.2-9} $$

where $\mathcal{G}$ is the power gain.

The expression for the electronic efficiency is obtained by combining Equations (12.1-25) and (12.2-9) as

$$ \eta_e' = \frac{1 - \dfrac{m\omega^2}{2e V_{ao}\beta^2}}{1 + \dfrac{I_{ao}}{B^2}\dfrac{m\beta^2 K}{2e}\left(\dfrac{\mathcal{G}+1}{\mathcal{G}-1}\right)} \tag{12.2-10} $$

where the prime indicates that loss due to back-bombardment of the cathode has not been included. Comparing this result with the corresponding result for the magnetron, Equation (12.1-32), we note a strong similar-

ity. Designing for high efficiency in both devices involves optimizing more or less the same parameters.

Equation (12.2-10) applies to a circuit which runs completely around the circumference of the tube. A rough estimate of the efficiency for a tube with a drift space can be obtained in the following manner. The total anode current is assumed to be independent of the length of the drift space. The power generated, on the other hand, must be only a fraction $F$ of the power generated using a circuit covering the whole circumference, where $F$ is the fraction of the circumference covered. Thus, the electronic efficiency is $F$ times the value for a tube with no drift space.

Let us calculate the efficiency of a typical crossed-field amplifier with the following parameters:

$$\begin{aligned}
f &= 9100 \text{ Mc} \\
V_{ao} &= 35.8 \text{ kv} \\
I_{ao} &= 18.5 \text{ amps} \\
B &= 0.40 \text{ weber/meter}^2 \\
\beta &= 1155 \text{ radians/meter} \\
K &= 120 \text{ ohms} \\
\mathcal{G} &= 10 \\
2\alpha l &= 0.230 \text{ (1 db attenuation)} \\
P_{\text{in}} &= 42 \text{ kw}
\end{aligned}$$

These parameters apply to an experimental tube assembled at Bell Telephone Laboratories. From Equation (12.2-10), we calculate the value

$$\eta_e' = 69\%$$

If we subtract 5 per cent for back-bombardment of the cathode, the net electronic efficiency becomes

$$\eta_e'' = 64\%$$

for a circuit completely around the circumference. In the particular experimental tube whose parameters are given above, the circuit covered only three fourths of the circumference ($F = \frac{3}{4}$). Therefore,

$$\eta_e = F\eta_e'' = 48\%$$

for this particular tube.

In order to calculate the circuit efficiency using Equation (12.2-5), we must first calculate $P_{\text{gen}}$ from the value of electronic efficiency obtained above. We obtain

$$P_{\text{gen}} = \eta_e I_{ao} V_{ao} = 316 \text{ kw}$$

From Equation (12.2-5), the circuit efficiency is calculated as

$$\eta_c = 87\%$$

so that the overall efficiency is

$$\eta = \eta_c\eta_e = 42\%$$

A value of 43 per cent was measured on this experimental tube. This value would be improved considerably if the length of the drift space were reduced.
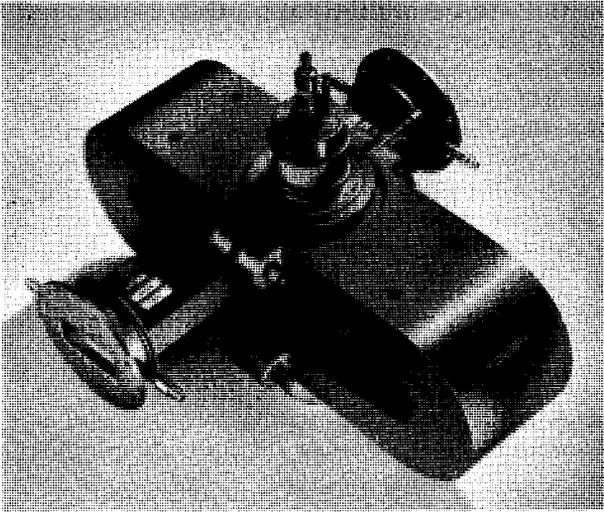


Fig. 12.2-4 Raytheon RK8129 *Amplitron*. The tube weighs 110 pounds with magnet, and it fits within a 51-cm cube. (Courtesy of Raytheon Company)

Next, we present a description of a particular crossed-field amplifier, the Raytheon RK8129 *Amplitron*,[13] pictured in Figure 12.2-4. Typical

TABLE 12.2-1. RK8129 *Amplitron* TYPICAL
OPERATING CHARACTERISTICS

Frequency, Mc.................................2900 to 3100
Peak power output, Mw............................ 3
Average power output, kw......................... 15
Peak anode voltage, kv........................... 50
Peak anode current, amps......................... 66
Peak rf input power, kw (minimum)................550
Gain, db......................................... 8
Pulse duration, microsec......................... 10
Efficiency, %.................................... 78

[13]*Amplitron* is a Raytheon trade name for a magnetron amplifier. The RK8129 is also designated the QK622.

operating characteristics of this tube are tabulated in Table 12.2-1.[14] The device is characterized by an extremely high efficiency, albeit the gain is only 8 db. Such a tube is useful as the final stage in a high-power transmitter, where over-all efficiency is the prime consideration.

High electronic efficiency is obtained through the use of an anode circuit of the type shown in Figure 12.2-1, with essentially a zero-length drift space. Since the circuit attenuation is less than 0.5 db, the circuit efficiency is also high.

Typical operating characteristics are shown as a function of frequency in Figure 12.2-5. The curves shown are for constant rf input signal, with the peak $V_{ao}I_{ao}$ product held approximately constant. The latter condition is automatically obtained when the anode voltage pulses are obtained from
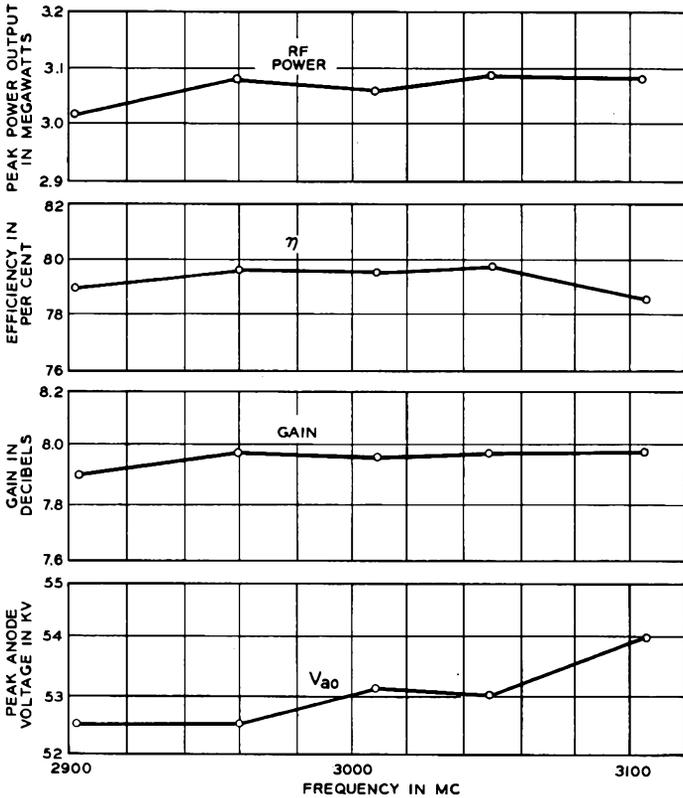


FIG. 12.2-5 RK8129 typical operating characteristics. The rf input power and peak anode current are held constant at 490 kw and 62 amps, respectively. (Courtesy of Raytheon Company)

[14]Reference 12.4.

a line-type modulator.[15]  The power output and gain are seen to vary little over the frequency band.

Tubes of this type have been operated in parallel to obtain additional power.

Proper operation of crossed-field amplifiers prohibits the application of anode voltage unless an rf signal is being applied at the input.  This prevents the generation of spurious rf output signals, which could possibly damage the tube.  These spurious signals result from the basic instabilities of electrons in crossed fields.

Because of the phase locking phenomenon in crossed-field amplifiers, the phase pushing is very small.  In the RK8129, the phase changes 0.5 degree per 1 per cent change in peak anode current.

Crossed-field amplifiers have a number of attributes.  They provide a compact, high-efficiency  amplifier with fractional bandwidths of the order of 10 per cent and relatively low-voltage operation.  The efficiency is typically higher than that of the magnetron oscillator and substantially higher than that of the traveling-wave or klystron amplifier.  When the tube is used as a power output tube in a radar installation, rf signals received by the antenna may be passed through the tube to the lower rf level input waveguide before they are separated from the common transmit-receive circuit.  This is possible because there is no sever in the anode circuit.  A simpler duplexer[16] can therefore be used.  Another attribute is the low phase pushing.  Cold cathode operation is possible, thus eliminating the necessity of a high-voltage-insulated heater supply.

On the other hand, the crossed-field amplifier has several disadvantages in comparison with klystron amplifiers and traveling-wave amplifiers.  Since it is a saturated amplifier, it cannot successfully transmit amplitude modulated signals.  The low gain requires the use of a high-power traveling-wave tube or klystron as a driver tube.  The noise level is high, as is typical of all crossed-field tubes.[17]  The fact that the electron beam must be dissipated on the anode slow-wave structure presents serious cooling problems at high power levels.

## 12.3  M-Carcinotron Oscillators

The M-carcinotron oscillator[18] is a backward-wave oscillator in which the

---

[15]An anode voltage source where the voltage pulse is obtained by discharging an artificial transmission line using a thyratron as a switch.

[16]A duplexer is a device which alternately connects the transmitter and receiver circuits to a radar antenna.

[17]Recent experiments have indicated that the noise is not inherently high in injected-beam crossed-field amplifiers (footnote 9); see Reference 12.6.

[18]Also known as an M-type backward-wave oscillator.  To avoid ambiguity, the device of Chapter 11 is termed an O-type backward-wave oscillator.

interaction between the electrons and the slow-wave structure takes place in a region of crossed dc fields.

A linear version of an M-carcinotron is shown in Figure 12.3-1. A slow-wave structure is arranged parallel to an electrode known as the sole. With
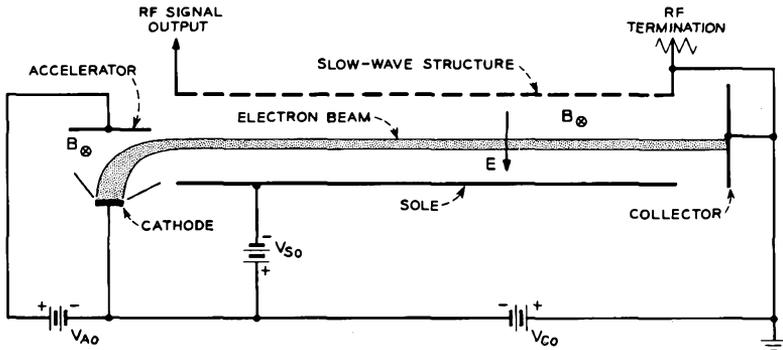


Fig. 12.3-1 Linear version of an M-carcinotron oscillator.

the slow-wave structure at ground potential, a voltage $V_{So}$ is provided to make the sole negative. This produces a dc electric field as shown. A dc magnetic field is directed into the page. As indicated in the figure, this magnetic field is also present in the electron gun region.

The electron gun is different from the electron guns in Chapter 4, because of the action of the magnetic field. The electrons are drawn from the cathode toward the accelerator by the accelerator voltage $V_{Ao}$. However, the magnetic field causes the electron trajectories to be curved through a 90-degree angle as shown. At this point the electrons leave the gun region and enter the interaction region. If the voltages are chosen properly, the electrons enter the interaction region with a velocity $E/B$ so that they travel parallel to the circuit,[19] until they are finally collected by the collector electrode. Since the cathode surface is rectangular, a rectangular sheet beam is formed.

The rf interaction is similar to that of the backward-wave oscillator. The electrons interact with a backward-wave space harmonic of the circuit, the energy on the circuit flowing opposite to the direction of the electron motion. This provides the feedback necessary for oscillation. The circuit is terminated at the collector end, and the rf signal is removed at the gun end.

---

[19]This neglects the rf perturbations of the electron motion which will be considered later.

The interaction with the backward-wave space harmonic may be discussed in connection with Figure 12.3-2. A line is drawn from the origin with a slope given by $E/B$, the drift velocity of the electrons. The intersection of this line with the backward-wave space harmonic determines the frequency at which the space-harmonic phase velocity equals the electron drift velocity, the synchronism condition. Thus oscillations are produced at a radian frequency $\omega_1$ provided that the beam current exceeds the starting value. Electronic tuning may be accomplished by changing the sole voltage $V_{So}$ or the cathode voltage $V_{Co}$, so as to vary the electron drift velocity. Amplitude modulation may be obtained by varying the accelerator voltage $V_{Ao}$, which varies the beam current.



Fig. 12.3-2 Interaction using the backward-wave space harmonic. $\omega_1$ is the radian frequency of oscillation.

Because of the phase focusing inherent in crossed-field devices, the exact frequency of oscillation is determined almost entirely by the circuit characteristics. That is, the frequency pushing (due to increasing the beam current) is very small.

Since the M-carcinotron is a crossed-field device, high-efficiency operation is possible. Efficiencies of 20 to 30 per cent are easily obtained.

The electron flow indicated in Figure 12.3-1 is perturbed by the rf electric field of the circuit. The perturbed electrons are shown in Figure 12.3-3 at a particular instant of time for a thin beam. Each electron moves in synchronism with the wave with a trajectory of the type previously discussed in connection with Figure 12-3.

For the purpose of discussion, half-cycle groups of electrons are identified by the letters $A$ through $G$. Electrons $A$ near the beginning of the circuit
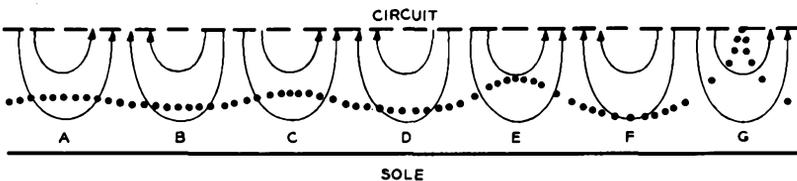


Fig. 12.3-3 Thin beam electrons and electric field lines at a particular instant of time in an M-carcinotron.

are moving toward the circuit, whereas electrons $B$ are moving toward the sole. Farther down the circuit, electrons $C$ are closer to the circuit, and electrons $D$ are closer to the sole. However, electrons $C$ have departed a greater distance from the unperturbed path than have electrons $D$. Thus, the electrons have lost a net amount of potential energy, this energy having been transferred to the rf field. The reason for the greater displacement of the electrons moving toward the circuit is that these electrons are in stronger rf fields, since they are closer to the circuit. Electrons $E$ and $F$ further illustrate this behavior. Electrons $G$ have moved so far from the un-perturbed position that some of them are being intercepted on the circuit.

In addition to the transverse motion of the electrons, there is also a bunching of the electrons in the longitudinal direction. Both of these effects cause induced currents in the circuit, and hence both effects must be considered in an analysis of the interaction process. Except for this added complexity, the analysis proceeds as for the backward-wave oscillator in Chapter 11. We shall not go into the details of this analysis. The interested reader is referred to the references.[20] The small-signal theory predicts the
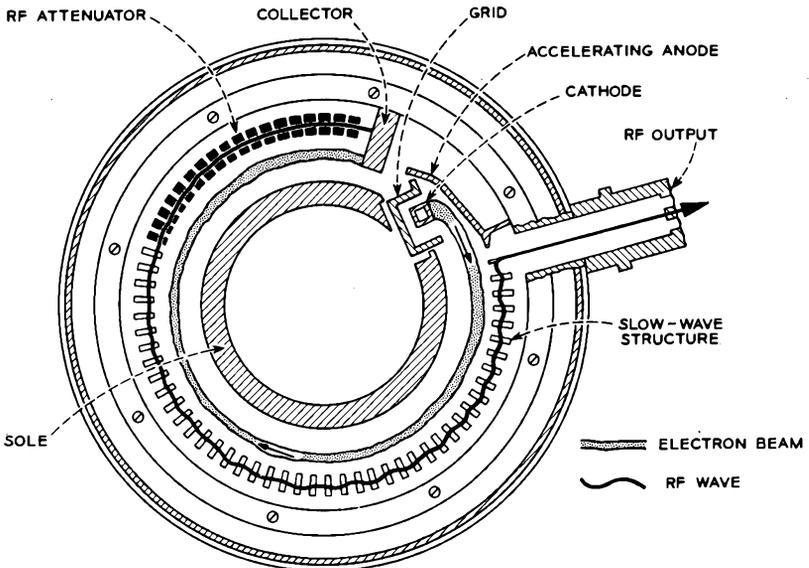


FIG. 12.3-4 Typical physical configuration of an M-carcinotron. The circular form of the device results in a lower magnet weight than would be required for a linear version. The rf attenuator shown serves to provide an rf termination for one end of the slow-wave structure. (Courtesy of Raytheon Company)

[20]Reference 12a, Vol. 1, pp. 395–495, 528–553, Reference 12.5.
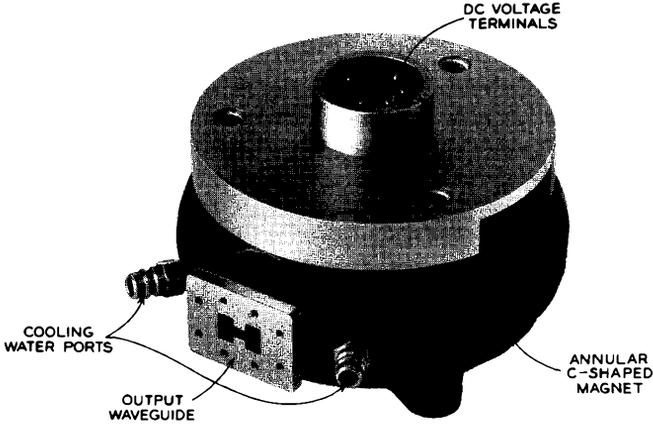
FIG. 12.3-5 Litton Industries L-3726 M-carcinotron. The tube weighs 16.5 pounds, with magnet. It measures approximately 18 cm in diameter and 13 cm high. (Courtesy of Litton Industries, Electron Tube Division)

starting current, and various large-signal, nonlinear theories allow one to estimate the efficiency. M-Carcinotrons are generally constructed in the circular re-entrant form shown in Figure 12.3-4. The slow-wave structure and sole are circular and nearly re-entrant to conserve magnet weight; the sole has the appearance of the cathode in a magnetron.

A typical M-carcinotron is the Litton L-3726, shown in Figure 12.3-5. The construction is similar to that of Figure 12.3-4. The circuit is an interdigital line,[21] a two-wire line with "fingers" extending alternately from one line toward the other. The tube delivers a minimum of 165 watts CW, voltage tunable over the frequency range 4800 to 6550 Mc.

TABLE 12.3-1. L-3726 TYPICAL OPERATING CHARACTERISTICS*

Frequency, Mc................................4800 to 6550
Power output, watts.......................... 167 to  305
Cathode voltage, $V_{Co}$, volts....................2045 to 4755
Accelerator voltage, $V_{Ao}$, volts....................1615
Sole voltage, $V_{So}$, volts...........................2450
Cathode current, ma............................. 300
Spurious output ratio, db (minimum)................ 15

*The symbols are defined in Figure 12.3-1.

Typical operating characteristics are summarized in Table 12.3-1. Tuning curves are shown in Figure 12.3-6. For this type of operation, tuning is

[21]See Problem 8.10.

accomplished by varying the cathode voltage $V_{Co}$. A faster rate of tuning is accomplished by varying the sole voltage, since the sole draws very little current and has only 65 picofarads of capacity to ground. However, with sole tuning the tuning range is limited to about 900 Mc with the other voltages held constant.
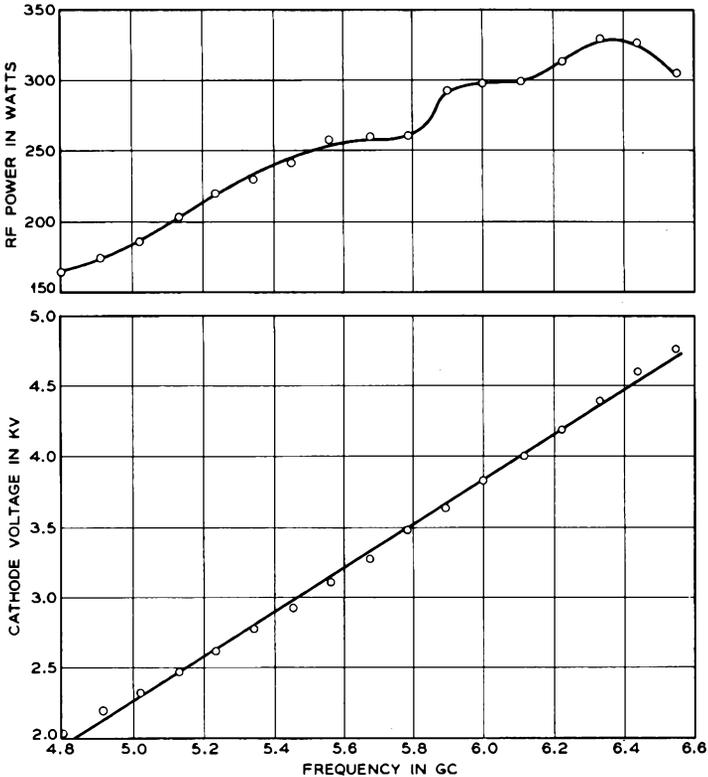


Fig. 12.3-6 Tuning curves for the L-3726. These data correspond to fixed values of sole voltage $V_{So}$ and cathode current of 2450 volts and 300 ma, respectively. (Courtesy of Litton Industries, Electron Tube Division)

Examination of Figure 12.3-6 reveals that the frequency tuning with voltage is nearly linear. This feature allows a simplification of the circuitry associated with electronic tuning. Linear tuning results chiefly from the fact that the electron drift velocity is linearly related to the dc electric field in the interaction space, which in turn is directly proportional to the cathode voltage $V_{Co}$. In contrast, the electron velocity in the O-type back-

ward-wave oscillator is proportional to the square root of the tuning voltage. As a result, the tuning sensitivity of the M-type oscillator is higher.

The electronic efficiency of the L-3726 varies from 27.5 per cent at the low-frequency end to 24 per cent at the high end for the operating conditions given above.

A disadvantage of the M-carcinotron is a relatively high degree of spurious output, due to the inherent instability of electrons in crossed fields. The spurious output of the L-3726 is as much as 15 db below the desired signal. Because of this high noise level, M-carcinotrons have found considerable applications as high efficiency jamming sources for electronic countermeasures.

We may summarize our discussion of the M-carcinotron by comparing its characteristics with those of the O-type backward-wave oscillator. Both tubes are electronically tunable over large bandwidths. The M-carcinotron has the advantages of high efficiency, linear tuning, and low-frequency pushing. On the other hand, the M-carcinotron has considerably more noise or spurious output.

## PROBLEMS

12.1 Derive Equation (12.1-19). Show that the velocity at the hub surface varies from $\omega/\beta$ at the Hartree voltage to $(eB/m)d$ at the Hull cutoff voltage.

12.2 Calculate the convection current density at the hub surface using the model of Figure 12.1-4 for a magnetron with the following parameters.

$$B = 7400 \text{ gauss}$$
$$V_{ao} = 17.5 \text{ kv}$$
$$d = 0.90 \text{ mm}$$

$$Ans.: 2560 \text{ amps/cm}^2$$

12.3 The electromagnetic fields in a magnetron operating in the $\pi$ mode may be considered to be either those corresponding to a standing wave or the result of the superposition of two equal traveling waves carrying equal amounts of power but



Problem 12.3

CROSSED-FIELD DEVICES
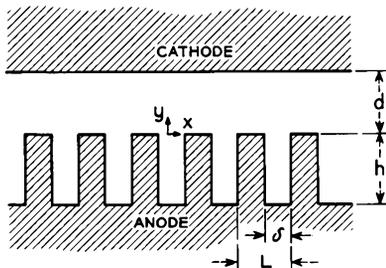
in opposite directions. The figure shows a linear version of a magnetron electrode configuration; this slow-wave structure is similar to that of Figure 8.7-4 and has similar properties.

(a) Sketch a Brillouin diagram for the lowest frequency mode of this structure over the range $-2\pi \leq \beta L \leq 2\pi$. Label the space harmonics. Indicate the point of interaction for the $\pi$-mode and indicate which two space harmonics are synchronous with the electrons at this point. Indicate which of the remaining space harmonics correspond to power flow in the same direction as each of the synchronous space harmonics.

(b) The $x$ component of the rf electric field in the region between the vane tips and the cathode for power flow in the positive $x$ direction is given by

$$E_{x+} = \frac{V_+}{L} \sum_{-\infty}^{\infty} \frac{\sinh \gamma_n(d-y)}{\sinh \gamma_n d} \frac{\sin \beta_n \delta/2}{\beta_n \delta/2} \epsilon^{-j\beta_n x}$$

where $\beta_n = \beta_o + (2\pi n/L)$ and $\gamma_n = \sqrt{\beta_n^2 - k^2}$, using an analysis similar to that given in Section 8.7(b). $V_+$ is the voltage across the vane-tip gap centered at $x = 0$. Write the corresponding expression for a wave having equal power flow in the negative $x$ direction, if $V_-$ instead of $V_+$ is the voltage across the gap centered at $x = 0$ in this case.

(c) For equal power flows in the two opposite directions $|V_+| = |V_-|$. Using this fact show that the four space harmonics discussed in part (a) have equal amplitudes at the frequency corresponding to $\pi$-mode operation.

(d) Show that the superposition of $E_{x+}$ and $E_{x-}$ corresponds to a standing wave, for $\pi$-mode operation.

12.4 Calculate the parameters of the equivalent circuit of Figure 12.1-19 as they apply to the Western Electric 7208B magnetron from the data given at the end of Section 12.1(c). Assume that $Y_e$ is purely conductive.

12.5 A crossed-field amplifier is being operated under conditions corresponding to maximum gain at a given anode voltage. Is it possible to increase the gain of the device by increasing the circuit length, assuming that the anode voltage and other circuit parameters are held constant? Is it possible to increase the rf power output for the same rf power input by this means? Is it possible to decrease the rf power input for the same rf power output by this means? Explain.
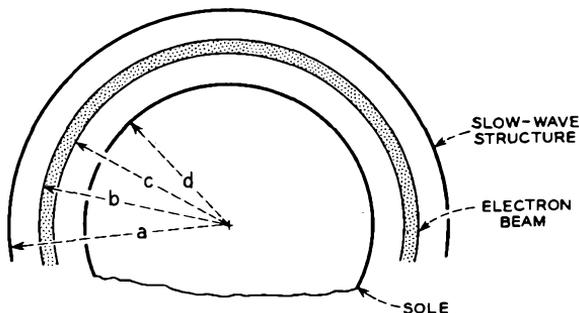


Problem 12.6

12.6  The figure shows ideal laminar flow in an M-carcinotron under dc conditions; that is, neglecting the effects of the rf fields.  All electrons are assumed to follow circular paths, and the angular velocity is the same for all electrons independent of the radius.  The latter condition allows exact synchronism of all electrons with the rf field despite the space-charge forces.  Assuming that the electron charge density is uniform, independent of the radius, show that this ideal flow requires

$$V_{S_o} = \frac{\rho c^2}{4\epsilon_o}\left(1 - 2\ln\frac{c}{d}\right)$$

and

$$V_{C_o} = -\frac{\rho b^2}{4\epsilon_o}\left(1 + 2\ln\frac{a}{b}\right)$$

where $\rho$ is the electron charge density and the voltages are defined in Figure 12.3-1.

12.7  The Hull cutoff parabola and Hartree line as plotted in Figure 12.1-7 have been derived from a linear model of the magnetron.  Show that curves of identical shape are obtained for the usual cylindrical model of the magnetron.

## REFERENCES

A general reference on crossed-field devices is:

12a.  E. Okress, Ed., *Crossed-Field Microwave Devices*, Academic Press, Inc., New York, 1961.

Two general references on magnetrons are:

12b.  G. B. Collins, Ed., *Microwave Magnetrons*, McGraw-Hill Book Co., Inc., New York, 1948.

12c.  J. C. Slater, *Microwave Electronics*, Chapter 13, D. Van Nostrand Co., Inc., Princeton, N. J., 1950.

Other references covering specific items are:

12.1  W. A. Smith, "A Wave Treatment of the Continuous Cathode Crossed-Field Amplifier," *Trans. IRE* **ED-9**, 379–387, September, 1962.

12.2  H. M. Olson and L. H. Von Ohlsen, "The Coaxial Magnetron, A Superior Microwave Power Source," *IRE Wescon Convention Record*, Vol. 2, Part 1, August, 1961.

12.3  W. C. Brown, "Description and Operating Characteristics of the Platinotron— A New Microwave Tube Device," *Proc. IRE* **45**, 1209–1222, September, 1957.

12.4  W. A. Smith and F. Zawada, "A 3-Megawatt, 15-Kilowatt S-Band Amplitron," *Microwave Journal* **2**, 42–45, October, 1959.

12.5  A. H. W. Beck, *Space-Charge Waves and Slow Electromagnetic Waves*, Pergamon Press, Inc., New York, pp. 256–270, 1958.

12.6  N. R. Mantena and T. Van Duzer. "Low Noise and Space-Charge Smoothing in a Crossed-Field Amplifier," *Proc. IEEE* **51**, 1662, 1663, November, 1963.

# Chapter 13

## NOISE

An important and fundamental limitation in the performance of all electron tubes and their associated circuits is noise. In this chapter we shall consider noise to be unwanted signals having a random or incoherent nature. The effects of noise may be of greater or lesser importance in communications systems, depending on the type of modulation used and the level of the intelligence-bearing signal relative to that of the noise. Noise may obscure weak signals in the case of amplitude or frequency modulation systems, and it may give false signals in pulse code modulation systems.

The word "noise" is also often used to include unwanted man-made signals, such as those due to electric motors, diathermy machines, automobile ignitions, transmitter harmonics, etc. However, such signals can be reduced to tolerable levels at their sources by proper engineering, and we shall not further consider this type of noise.

Random noise results from the finite charge on the electron or other carriers within the conductors and devices used in electronic circuits. Because an electric current within a conductor or an electron beam is made up of individual charged particles in motion, and because each particle carries a finite discrete charge, the current flow is never continuous but is subject to statistical fluctuations about an average value.

We shall find that the induced current flowing in the external circuit of a temperature-limited diode is characterized by a mean-square fluctuation current, or noise current, given by

$$\overline{i^2} = 2eI_{So}\Delta f \qquad (13\text{-}1)$$

where $-e$ is the electronic charge, $I_{So}$ is the dc beam current, and $\Delta f$ is the bandwidth over which the noise current is measured. This expression is valid only when the electron transit time is small compared with the period of the noise frequencies under observation. Noise resulting from a current flow which is limited by the random passage of electrons over a potential

barrier, such as the potential barrier at the surface of a temperature-limited cathode, is known as shot noise.

A second fundamental source of random noise results from the thermal motions of electrons or other charge carriers within a resistance. The motion of the charge carriers within the resistance causes a noise voltage to appear across the terminals of the resistance. We shall find that the mean square noise voltage across a resistance $R$ is given by

$$\overline{v^2} = 4kTR\Delta f \tag{13-2}$$

where $k$ is Boltzmann's constant, equal to $1.38 \times 10^{-23}$ joules/°K, $T$ is the absolute temperature of the resistance, and $\Delta f$ is the bandwidth over which the noise voltage is measured. If the resistance is connected to a second resistance, noise power is transmitted from the first resistance to the second resistance, and likewise noise power is transmitted from the second resistance to the first resistance. Maximum noise power is transmitted from one resistance to the other when the two resistances are equal. This maximum transmitted noise power, or the available noise power from the resistance, is given by

$$P = kT\Delta f \tag{13-3}$$

In grid-controlled tubes, the individual electrons of the beam pass the control grid at random times with the result that noise currents are induced in the grid circuit. This leads to a noise voltage at the control grid, and this noise voltage is amplified by the tube. Still another kind of noise, called partition noise, appears in the output of grid-controlled tubes and beam-type tubes when an electrode such as a screen grid intercepts a portion of the electron beam and hence adds random fluctuations to the remainder of the beam current. This type of noise is closely related to shot noise.

In beam-type microwave tubes, the random nature of the electron emission time and emission velocity gives rise to initial current and velocity modulation on the electron beam, and this in turn is amplified.

Noise sets a limit to the smallest signal that an amplifier can detect. The noise at the output of an amplifier consists of the amplified thermal noise due to the resistance of the source plus the noise added by the amplifier itself. For high receiver sensitivity, the noise added by the amplifier should be as small as possible.

The noise performance of an amplifier is specified by a quantity termed the noise figure (or noise factor). The noise figure $F$ is defined by[1]

$$F = \frac{\mathcal{G}_o k T_o B + N_a}{\mathcal{G}_o k T_o B} = 1 + \frac{N_a}{\mathcal{G}_o k T_o B} \tag{13-4}$$

---

[1] References 13.1, 13.2. Available power is the power delivered to a matched load. Available power gain is the ratio of available powers at the output and input of an amplifier.

where $N_a$ is the total available noise power due to noise sources within the amplifier, $\mathcal{G}_o$ is the available power gain of the amplifier at band center, $T_o$ is a standard reference temperature equal to 290°K, and $B$ is the amplifier bandwidth. The bandwidth $B$ is defined as

$$B = \frac{1}{\mathcal{G}_o} \int_0^\infty \mathcal{G}(f)df \qquad (13\text{-}5)$$

where $\mathcal{G}(f)$ is the available power gain at frequency $f$. The quantity $\mathcal{G}_o k T_o B$ in Equation (13-4) is then the total available noise power output for an ideal noiseless amplifier with the same variation of available power gain with frequency. Equation (13-4) states that the noise figure is equal to the ratio of the total available output noise to the available noise which would be present at the output if the amplifier itself were noiseless, assuming a source temperature of 290°K. The noise figure is usually expressed in decibels; that is, ten times the logarithm to the base 10 of the above quantity.

The expression for noise figure may be written in another form. The available signal power at the input $S_{in}$ is related to the available signal power at the output $S_{out}$ by

$$S_{out} = \mathcal{G}_o S_{in} \qquad (13\text{-}6)$$

Equation (13-4) may thus be written

$$F = \frac{\dfrac{S_{in}}{kT_o B}}{\dfrac{S_{out}}{\mathcal{G}_o k T_o B + N_a}} = \frac{\dfrac{S_{in}}{N_{in}}}{\dfrac{S_{out}}{N_{out}}} \qquad (13\text{-}7)$$

where $N_{in}$ and $N_{out}$ are the available noise powers at the input and output, respectively, assuming the standard source temperature.

The noise figure defined above is sometimes called the average noise figure, since the noise power is averaged over the entire bandwidth of the amplifier. Alternatively, for wide-band amplifiers, it is sometimes desirable to know the *spot noise figure*, a function of frequency. The spot noise figure is defined as above, except that noise power is measured in some small unit frequency interval centered at the signal frequency. $B$ in the above equations is then replaced by this unit frequency interval.

The above definitions of noise figure assume a standard source temperature $T_o$ of 290°K, room temperature. This is convenient for noise figure measurements. However, receivers are sometimes operated with receiving antennas pointed toward the sky. Under such conditions, the effective source temperature is often considerably less than room temperature.[2] At

---

[2] Reference 13.3.

frequencies proposed for satellite communications, 1 to 10 Gc, the effective sky temperature may be as low as 4°K. Let us examine the applicability of the noise figure concept under these conditions.

The available output noise added by the amplifier is obtained from Equation (13-4) as

$$N_a = (F - 1)kT_oB\mathcal{G}_o \qquad (13\text{-}8)$$

The total available output noise for a source at temperature $T_s$ is given by

$$N_{\text{out}}' = kT_sB\mathcal{G}_o + N_a$$

$$= kT_sB\mathcal{G}_o\left[1 + \frac{T_o}{T_s}(F - 1)\right] \qquad (13\text{-}9)$$

where the prime indicates that the actual source temperature is indicated. Using Equation (13-6), this may be written as

$$\frac{S_{\text{in}}}{kT_sB} = \frac{S_{\text{out}}}{N_{\text{out}}'}\left[1 + \frac{T_o}{T_s}(F - 1)\right] \qquad (13\text{-}10)$$

which relates the signal-to-noise ratios at the input and output of the amplifier.[3]

Equation (13-10) may be used to determine the minimum detectable input signal for a receiver. Suppose that we ask: How weak can the input signal be and still be distinguished from the noise? The signal-to-noise ratio at the output corresponding to this condition is somewhat arbitrary, so we shall take it as unity. That is, we shall set $S_{\text{out}}/N_{\text{out}}' = 1$. This corresponds to a doubling of the output power as a criterion to determine the presence of a signal.

The minimum detectable input signal $S_{\text{in}}$ from Equation (13-10) is plotted in Figure 13-1 as a function of noise figure for $S_{\text{out}}/N_{\text{out}}' = 1$ and for three source temperatures. The receiver bandwidth is taken arbitrarily as 1 Mc. Values for other bandwidths are easily obtained, since the minimum detectable input signal is directly proportional to the receiver bandwidth. For a source at room temperature, the relationship between minimum detectable input signal and noise figure is linear. However, for lower source temperatures, the minimum detectable input signal decreases more rapidly with decreasing noise figure. Thus, considerable improvement is realized by using low noise figure receivers ($F < 5$ db) in systems whose receiving antennas are directed toward the sky.

---

[3]Another useful noise quantity is the *effective input noise temperature* of the amplifier, equal to $T_o(F - 1)$. This is the temperature of an equivalent noise source at the input of an equivalent noiseless amplifier which replaces the actual amplifier. It is useful because it gives the relative contribution of the amplifier to the total output noise for any source temperature.
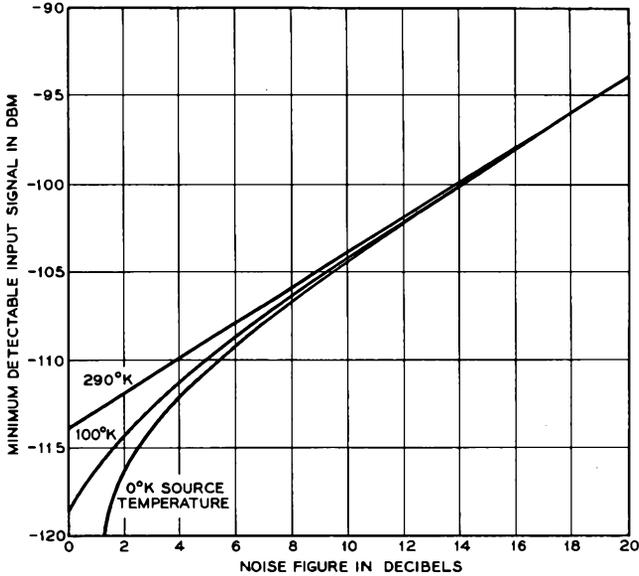
FIG. 13-1 Receiver sensitivity for a bandwidth of 1 Mc for signal sources at various temperatures as a function of the receiver noise figure. The signal-to-noise ratio at the receiver output is assumed to be unity.

We shall use the noise figure concept at several places in this chapter as a measure of the noise performance of an amplifier.

## 13.1   Fundamental Sources of Noise

There are certain fundamental sources of noise due to the fact that electron current is composed of individual electrons, each with its own velocity and position. Statistical methods have been used to analyze these effects.[4] Most of the analyses are rather complex for an introductory study of noise; hence, we shall limit our discussion to nonrigorous demonstrations of the fundamental theorems.

### (a)   Thermal Noise

Any passive resistance acts as a noise source with an available noise power given by Equation (13-3). The available noise power is the power

---

[4]Reference 13.4.
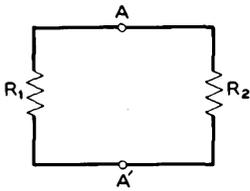
FIG. 13.1-1 Two equal resistors at
different temperatures connected to-
gether. $R_1$ is at temperature $T_1$, and
$R_2$ is at temperature $T_2$.

which is delivered to a matched load.
Thus in Figure 13.1-1, if $R_1 = R_2$, a
power $kT_1\Delta f$ flows from $R_1$ to $R_2$
within the frequency range $\Delta f$, and
similarly $kT_2\Delta f$ flows from $R_2$ to $R_1$
within the frequency range $\Delta f$, where
$T_1$ and $T_2$ are the respective resistor
absolute temperatures. If $T_1 > T_2$,
energy is transferred from $R_1$ to $R_2$.
If the resistors are at the same tem-
perature, no net transfer of energy
occurs; the powers flowing in the two directions are equal.

Equivalent circuits for a resistor as a noise source are given in Figure
13.1-2. $\overline{v^2}$ and $\overline{i^2}$ indicate time averages of the square of the voltage and
current, respectively. An equivalent rms quantity would be given by the



(a)                            (b)
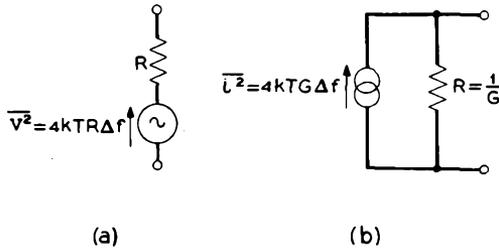
FIG. 13.1-2 Equivalent circuits for the thermal noise associated with a passive
resistance. $R$ in these equivalent circuits is taken to be noiseless. (a) Voltage
source equivalent circuit. (b) Current source equivalent circuit.

square root of such a quantity. It is simple to demonstrate that each of
these circuits has an available power equal to $kT\Delta f$. The resistor in each
equivalent circuit is considered noiseless; the voltage and current sources
represent the only sources of noise.

As an example of the use of these equivalent circuits, let us determine the
exchange of energy for the circuit of Figure 13.1-1, where $R_1 \neq R_2$ and
$T_1 \neq T_2$. The equivalent circuit is shown in Figure 13.1-3, using the voltage
source equivalent circuit of Figure 13.1-2(a). The two noise sources are
uncorrelated. That is, the random electron motion of one resistor is com-
pletely independent of that of the other resistor. Thus, if we were to com-
pare the relative phases of the two sources in a very narrow bandwidth, the
phase difference could be anywhere from zero to 360 degrees, with equal
probability, and is continuously varying. Each source acts completely in-
dependently of the other. We may use a very special type of superposition
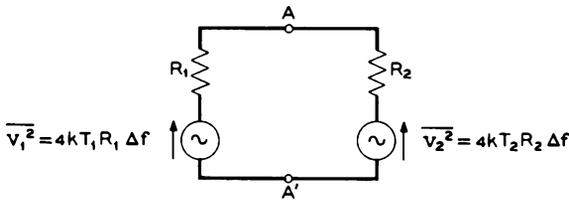
FIG. 13.1-3 Noise equivalent circuit for two unequal resistors $R_1$ and $R_2$, having temperatures $T_1$ and $T_2$, respectively, connected together.

in these problems of uncorrelated noise sources: The noise power delivered from each source to the other parts of the circuit may be evaluated by computing power with all other sources set to zero. Thus, in Figure 13.1-3 the power delivered to $R_2$ from $R_1$ is given by

$$P_{12} = \frac{\overline{v_1^2}R_2}{(R_1 + R_2)^2} = \frac{4kT_1R_1R_2\Delta f}{(R_1 + R_2)^2} \qquad (13.1\text{-}1)$$

Similarly, the power delivered to $R_1$ from $R_2$ is

$$P_{21} = \frac{\overline{v_2^2}R_1}{(R_1 + R_2)^2} = \frac{4kT_2R_1R_2\Delta f}{(R_1 + R_2)^2} \qquad (13\text{-}1.2)$$

The net flow of power from $R_1$ to $R_2$ is

$$P_{12} - P_{21} = \frac{4k(T_1 - T_2)R_1R_2\Delta f}{(R_1 + R_2)^2} \qquad (13.1\text{-}3)$$

If $R_1 = R_2 = R$, we have

$$P_{12} - P_{21} = k(T_1 - T_2)\Delta f \qquad (13.1\text{-}4)$$

which agrees with our previous result.

As another example, we shall determine the noise voltage appearing across the terminals $A$-$A'$ in Figure 13.1-1. This voltage would be of interest in the case where the resistors shown comprise the grid circuit in an
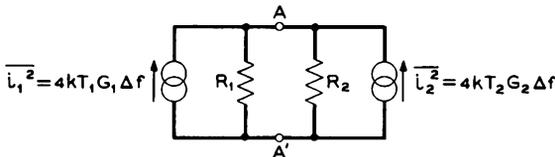


FIG. 13.1-4 Noise equivalent circuit of two resistors in parallel, both at the same temperature. It is demonstrated that the noise properties are given by the thermal noise of the equivalent parallel resistor.

amplifier; this noise voltage multiplied by the voltage gain of the amplifier appears as a portion of the output noise of the amplifier.

The noise voltage appearing across terminals $A$-$A'$ is most readily determined using the current source equivalent circuit of Figure 13.1-2(b). The noise equivalent circuit for the two resistors is shown in Figure 13.1-4. For convenience, we define the conductances $G_1 = 1/R_1$, $G_2 = 1/R_2$, and $G_T = G_1 + G_2 = 1/R_T$. The total noise voltage is obtained by superposition of the mean-square voltages produced by each noise source. The total noise voltage is given by

$$\overline{v_T^2} = \frac{4kT_1G_1\Delta f}{G_T^2} + \frac{4kT_2G_2\Delta f}{G_T^2}$$

$$= 4k\Delta f \frac{T_1G_1 + T_2G_2}{G_T^2} \qquad (13.1\text{-}5)$$

If $T_1 = T_2 = T$, Equation (13.1-5) becomes

$$\overline{v_T^2} = 4kTR_T\Delta f \qquad (13.1\text{-}6)$$

which is the open-circuit voltage of Figure 13.1-2(a), with $R$ replaced by $R_T$, the equivalent resistance of the parallel combination of resistors. This is a special case of a more general theorem; that is, the available output noise power at any two terminals of a passive isothermal network is equal to $kT\Delta f$.[5] Thus, the noise at two terminals of any passive isothermal network may be obtained from the equivalent circuits of Figure 13.1-2, where $R$ in Figure 13.1-2(a) is given by the real part of the two-terminal impedance and $1/R$ in Figure 13.1-2(b) is given by the real part of the two-terminal admittance.

Noise calculations in circuits containing reactive elements are made using the equivalent circuits of Figure 13.1-2 and the rules of ac circuit analysis. As an example let us compute the noise current flowing in an $R$-$L$-$C$ loop. The impedance of the loop is given by

$$|Z| = \sqrt{R^2 + (\omega L - 1/\omega C)^2}$$

Using the equivalent circuit of Figure 13.1-2(a), the noise current is given by

$$\overline{i^2} = \frac{\overline{v^2}}{|Z|^2} = \frac{4kTR\Delta f}{R^2 + (\omega L - 1/\omega C)^2}$$

A simple derivation of the expression for thermal noise, Equation (13-3), proceeds as follows.[6] In Figure 13.1-5 are shown two identical resistors of

---

[5]This general theorem requires only that each resistor possess bilateral network properties with respect to the output terminals. See Reference 13.4, pp. 185–189.

[6]Reference 13.5.

resistance $R$ joined by a lossless
TEM transmission line of character-
istic impedance $R$. All parts of the
circuit are at a temperature $T$.
Since the transmission line is matched
at both ends, each resistor is deliver-
ing its available power to the line.
This power then flows to the other
resistor where it is dissipated.

Now, suppose that the line is sud-
denly shorted at both ends. The
shorted line is a resonant element
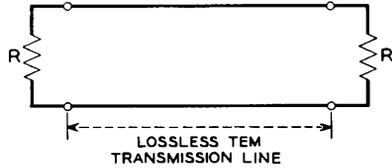
FIG. 13.1-5 A transmission line ter-
minated in its characteristic impedance
$R$ at both ends. The line and the ter-
minating resistors are all at the same
temperature $T$.

with resonant frequencies corresponding to the line length $L$ being an
integral number of half wavelengths, that is

$$f_n = \frac{nc}{2L} \qquad\qquad (13.1\text{-}7)$$

where $c$ is the velocity of propagation on the line, and $n$ is an integer. Thus
the modes of oscillation are spaced by a frequency interval $\Delta f$, where

$$\Delta f = \frac{c}{2L} \qquad\qquad (13.1\text{-}8)$$

From the equipartition theorem of statistical mechanics,[7] each mode of
oscillation will have associated with it an energy $kT$. Now the standing
wave corresponding to the resonant mode is comprised of two traveling
waves, traveling in opposite directions. Hence, each traveling wave has
an energy $\frac{1}{2} kT$ and an energy per unit length given by

$$W_l = \frac{kT}{2L} \qquad\qquad (13.1\text{-}9)$$

Since the group velocity for a TEM transmission line is $c$, the power flow
in one direction $P$ is given by

$$P = cW_l = \frac{kTc}{2L} \qquad\qquad (13.1\text{-}10)$$

We may substitute for $L$ from Equation (13.1-8) obtaining

$$P = kT\Delta f \qquad\qquad (13.1\text{-}11)$$

Since the line was shorted instantaneously, this must also be the power flow
in each direction before the lines were shorted. Thus, each resistor has an
available noise power given by Equation (13.1-11). A rigorous develop-
ment of an expression for thermal noise is outlined in Appendix XVI.

---

[7]Reference 13a, Chapter 11.

(b)   *Shot Noise*

In an electron tube the cathode emission current consists of individual electrons emitted at random. The random emission times for the individual electrons result in fluctuations of the current induced in the output circuit. The noise associated with these fluctuations is known as shot noise.

When the current drawn from the cathode is space charge limited, the space charge near the cathode acts to reduce the shot noise, an effect which will be studied later. On the other hand, when the emission is temperature limited, full shot noise is obtained. Diodes operated with temperature-limited emission are often used as standard noise sources.

The magnitude of full shot noise for a diode can be evaluated by statistical methods. Full shot noise is given by a current source of mean-square value[8]

$$\overline{i^2} = \frac{8eI_{So}}{(\omega\tau)^4}[(\omega\tau)^2 + 2(1 - \cos\omega\tau - \omega\tau\sin\omega\tau)]\Delta f \qquad (13.1\text{-}12)$$

for a planar diode, where $I_{So}$ is the temperature limited dc current, and $\tau$ is the electron transit time. The ac current $i$ is equal to the difference between the instantaneous value of the current and the average value, and $\overline{i^2}$ indicates the time average of the square of this quantity. Equation
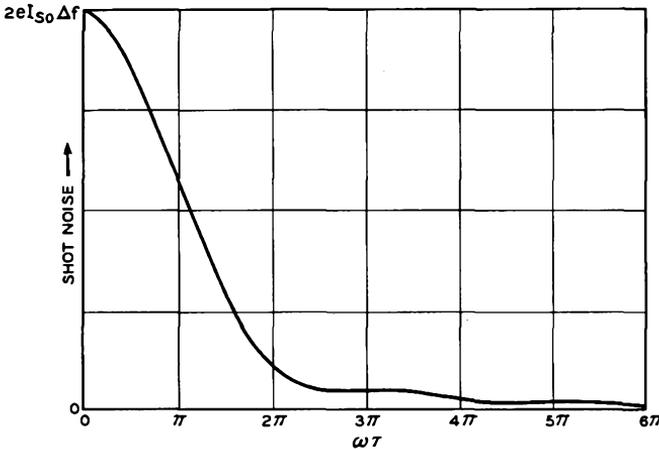


FIG. 13.1-6 Variation of full shot noise with transit angle in a parallel-plane diode (From *An Introduction to the Theory of Random Signals and Noise*, by W. B. Davenport and W. L. Root. Copyright 1958. McGraw-Hill Book Co., Inc. Used by permission)

[8]Reference 13.4, Chapter 7.

(13.1-12) is plotted in Figure 13.1-6. At low frequencies, full shot noise is given by

$$\overline{i^2} = 2eI_{So}\Delta f \tag{13.1-13}$$

The latter expression holds for diodes of any geometry. It is to be noted that full shot noise is not a function of temperature; one might expect this since it is caused by random emission time.

A simple derivation of Equation (13.1-13) proceeds as follows.[9] Consider the vacuum tube shown in Figure 13.1-7 consisting of two identical, planar, close-spaced cathodes parallel to each other. The entire structure is assumed to be enclosed in an isothermal oven at temperature $T$. Each cathode emits a small temperature limited current with noise fluctuations, intercepted by the opposite cathode, so that a noise current flows in the external wire connecting the two cathodes. This noise current is equal to

$$\overline{i^2} = 4kTG\Delta f \tag{13.1-14}$$

from Figure 13.1-2, where $G$ is the ac conductance of the diode formed by the two cathodes. This formula applies to the diode since it is a passive, isothermal structure in thermal equilibrium.

Next we must evaluate the ac conductance of the diode. This conductance is equal to $\partial I_o/\partial V_o$ in the limit as both the diode current $I_o$ and
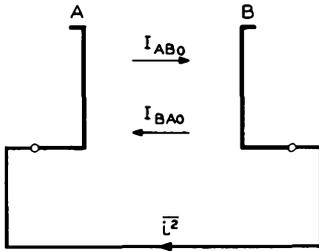


FIG. 13.1-7 Isothermal structure consisting of two identical, planar, close-spaced cathodes parallel to each other.
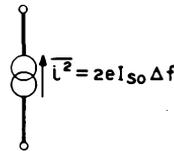
FIG. 13.1-8 Equivalent circuit for the noise from a temperature-limited diode at low frequencies.

the diode voltage $V_o$ approach zero. In order to evaluate the functional relationship between $I_o$ and $V_o$ let us assume that cathode $A$ is positive with respect to cathode $B$, the potential difference being equal to $V_o$. The diode current $I_o$ is given by

$$I_o = I_{ABo} - I_{BAo} \tag{13.1-15}$$

---

[9]Reference 13.6.

where $I_{ABo}$ corresponds to the current of electrons emitted from cathode $B$ arriving at cathode $A$, and $I_{BAo}$ corresponds to the current of electrons emitted from cathode $A$ arriving at cathode $B$.

Since electrode $A$ is positive, $I_{ABo}$ is equal to the maximum or saturated value of temperature limited current, which we call $I_{So}$. On the other hand, the electrons emitted from cathode $A$ see a repelling field so that only the electrons with normally directed emission energy greater than $eV_o$ electron volts are able to escape from $A$ and arrive at $B$. The probability distribution function for the emitted electrons is given by Equation (2.4-2) as a function of the kinetic energy associated with the normally directed velocity. The fraction of the emitted electrons with energies sufficient to overcome the potential barrier is obtained from this equation as:

$$\frac{I_{BAo}}{I_{So}} = \frac{\dfrac{1}{W_T}\displaystyle\int_{eV_o}^{\infty} \epsilon^{-W_n/W_T}dW_n}{\dfrac{1}{W_T}\displaystyle\int_{0}^{\infty} \epsilon^{-W_n/W_T}dW_n} = \epsilon^{-eV_o/kT} \qquad (13.1\text{-}16)$$

Thus, the total current is given by

$$I_o = I_{So} - I_{So}\epsilon^{-eV_o/kT} \qquad (13.1\text{-}17)$$

For $V_o$ approaching zero, the conductance is given by

$$G = \frac{\partial I_o}{\partial V_o}\bigg|_{V_o=0} = \frac{eI_{So}}{kT} \qquad (13.1\text{-}18)$$

from Equation (13.1-17).

When the value of conductance given by the last equation is substituted into Equation (13.1-14), one obtains

$$\overline{i^2} = 4kTG\Delta f = 4eI_{So}\Delta f$$

This noise current is the sum of two equal uncorrelated shot noise currents — that due to cathode $A$ and that due to cathode $B$. For a single cathode the shot noise is half of this, or

$$\overline{i^2} = 2eI_{So}\Delta f$$

identical with Equation (13.1-13).

The low-frequency equivalent circuit of the temperature limited diode for noise calculations is given in Figure 13.1-8.

### (c)   Velocity Fluctuations

In addition to a random emission time, each electron possesses a random emission velocity. In low-frequency tubes operating under space-charge limited conditions, we shall find in the next section that it is the randomness

of the electron velocity which sets a limit to the minimum noise. In micro-wave tubes, the velocity fluctuations act as initial velocity modulation on the electron beam, which may be amplified to produce output noise.

The probability distribution function for the normal component of electron emission velocity is given by Equation (1) of Appendix IV as

$$dP(u_n) = \frac{mu_n}{kT_c}\epsilon^{-mu_n{}^2/2kT_c}du_n \qquad (13.1\text{-}19)$$

where $T_c$ is the cathode temperature. As the emitted electrons move toward the potential minimum, those of lowest velocity are continuously sorted out and returned to the cathode.

Let us calculate the average velocity of the forward moving electrons at any point between the cathode and the potential minimum. The normal component of velocity of an electron at a point where the electrostatic potential relative to the cathode is $V$ may be written in terms of the emission component $u_n$ as

$$u^2 = u_n{}^2 + 2\frac{e}{m}V \qquad (13.1\text{-}20)$$

Since the electrostatic potential is independent of an individual electron's velocity, Equation (13.1-20) leads to

$$udu = u_n du_n \qquad (13.1\text{-}21)$$

Thus, Equation (13.1-19) can be written as

$$dP(u) = \frac{mu}{kT_c}\epsilon^{-(m/2kT_c)(u^2-2eV/m)}du \qquad (13.1\text{-}22)$$

The average velocity of the electrons moving away from the cathode at a point between the cathode and the potential minimum where the potential relative to the cathode is $V$ is given by

$$\bar{u} = \frac{\displaystyle\int_0^\infty \frac{mu^2}{kT_c}\epsilon^{-(m/2kT_c)(u^2-2eV/m)}du}{\displaystyle\int_0^\infty \frac{mu}{kT_c}\epsilon^{-(m/2kT_c)(u^2-2eV/m)}du}$$

$$= \sqrt{\frac{\pi kT_c}{2m}} \qquad (13.1\text{-}23)$$

a constant. This is an interesting result. Since $\bar{u}$ is independent of $V$, we must conclude that even though all electrons are being slowed down by the repelling field of the potential minimum, the average velocity of those passing the potential minimum is constant. Moreover, this average veloc-ity is also independent of the potential difference between the cathode and

the potential minimum. Thus, $\bar{u}$ remains constant even though the potential mimimum fluctuates due to some noise perturbation.

The average velocity given by Equation (13.1-23) is an average taken over a sufficiently long length of time. At any particular time the average electron velocity may be slightly higher or slightly lower than this value. The mean-square fluctuation of the average velocity at the potential minimum can be shown to be[10]

$$\overline{u_A{}^2} = \frac{ekT_c\Delta f}{mI_{ao}}(4 - \pi) \tag{13.1-24}$$

where $I_{ao}$ is the dc anode current. $u_A$ is the difference between the instantaneous value of the average velocity at the potential minimum and the value given by Equation (13.1-23), and $\overline{u_A{}^2}$ is the time average of the square of this quantity.

## (d)  Space-Charge Smoothing

When tubes are operated under space-charge limited conditions, the output noise is found to be considerably less than that corresponding to full shot noise. Thus, if the noise is represented by a current source in the plate circuit, it is of a strength,

$$\overline{i^2} = \Gamma^2 \cdot 2eI_{ao}\Delta f \tag{13.1-25}$$

where transit-time effects are assumed to be negligible, $I_{ao}$ is the dc anode current, and $\Gamma^2$ is a factor less than unity.

A physical explanation of the space-charge smoothing of the shot noise is easily given. Potential profiles for a space-charge limited diode are shown in Figure 13.1-9. The equilibrium potential profile is shown as a solid line. If, due to shot noise fluctuations, a larger than average number of electrons is emitted at a particular instant, the potential minimum would be depressed slightly as indicated by curve $A$. Because of the deepening of this minimum, a smaller percentage of the emitted electrons is able to pass the minimum and go on to the anode. Only those electrons with sufficiently high emission velocities can pass the minimum. Thus, the higher instantaneous emission current is somewhat neutralized by the rejection of a larger percentage of the emitted electrons by the deeper potential minimum. Curve $B$ corresponds to the opposite situation, where a fewer-than-average number of electrons are emitted at a particular instant. In this case, there is less space charge in front of the cathode, the potential minimum becomes shallower, and a larger percentage of the emitted electrons pass on to the anode.

---

[10]Reference 13.7.
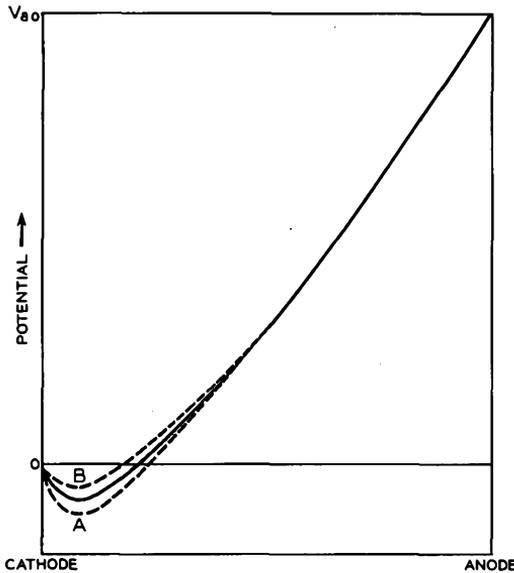
Fig. 13.1-9 Potential distribution in a space-charge limited diode under equilibrium conditions (heavy line). Dotted curves $A$ and $B$ indicate the potential distributions obtained due to an instantaneous excess or deficiency of emitted current, respectively.

This space-charge smoothing process is indeed very effective. At low frequencies, the current fluctuations are reduced greatly. $\Gamma^2$ may be as low as 0.02.

We can derive a simple expression for the space-charge smoothing factor $\Gamma^2$ using the Llewellyn and Peterson equations of Section 7.2.[11] The Llewellyn and Peterson equations are given by Equations (7.2-1), and the coefficients are tabulated in Appendix XI. The equations are applied to an open-circuited diode; that is, the diode is operated with an rf open circuit such as is obtained with a series rf choke of high inductance. We shall derive the noise-fluctuation voltage appearing across the diode.

In Figure 7.2-1(a) we take the potential minimum to correspond to plane $A$ and the anode to correspond to plane $B$. The first of Equations (7.2-1) (symbols as defined in Section 7.2) may be written as

$$V_B - V_A = B^*J_A + C^*U_A \qquad (13.1\text{-}26)$$

since $J_T = 0$ for an open-circuited diode. From Appendix XI we have

$$\zeta = 1 \qquad (13.1\text{-}27)$$

[11]Reference 13.7.

for a space-charge limited diode, and

$$u_{Ao} = 0 \qquad\qquad (13.1\text{-}28)$$

corresponding to zero dc velocity at the potential minimum, so that

$$B^* = \frac{T^2}{\epsilon_o \beta^3} u_{Ao}[2P - \beta Q] = 0 \qquad\qquad (13.1\text{-}29)$$

and

$$C^* = -\frac{2m}{e} u_{Bo} \frac{P}{\beta^2} \qquad\qquad (13.1\text{-}30)$$

At low frequencies, $\beta$ approaches zero and

$$\frac{P}{\beta^2} \rightarrow \frac{1}{2} \qquad\qquad (13.1\text{-}31)$$

so that $C^*$ is given by

$$C^* = -\frac{m}{e} u_{Bo} \qquad\qquad (13.1\text{-}32)$$

Thus, from Equations (13.1-26), (13.1-29), and (13.1-32), we have

$$\overline{v^2} = \left(\frac{m}{e} u_{Bo}\right)^2 \overline{u_A{}^2} = \frac{2m}{e} V_{ao} \overline{u_A{}^2} \qquad\qquad (13.1\text{-}33)$$

where $\overline{v^2}$ is the mean-square voltage fluctuation across the open-circuited diode, and $\overline{u_A{}^2}$ is the mean-square velocity fluctuation at the potential minimum; $V_{ao}$ is the dc anode voltage. But, as discussed in part (c) above, $\overline{u_A{}^2}$ is given by Equation (13.1-24). Thus, Equation (13.1-33) may be written as

$$\overline{v^2} = 2kT_c\Delta f \frac{V_{ao}}{I_{ao}}(4 - \pi) \qquad\qquad (13.1\text{-}34)$$

Equation (13.1-34) may be written in a more useful form. Let us define the dynamic resistance of the diode as

$$r_a = \frac{dV_{ao}}{dI_{ao}} \qquad\qquad (13.1\text{-}35)$$

as in Appendix X. Since the space-charge limited diode current is proportional to the three-halves power of the voltage, we have

$$r_a = \frac{2}{3} \frac{V_{ao}}{I_{ao}} \qquad\qquad (13.1\text{-}36)$$

Thus, Equation (13.1-34) may be written as

$$\overline{v^2} = 4k\theta T_c r_a \Delta f \qquad\qquad (13.1\text{-}37)$$

where $\theta = 3(1 - \pi/4) = 0.644$. Com-
paring this with the correspond-
ing expression for thermal noise, we
see that the diode appears as a ther-
mal noise source whose temperature
is 0.644 times the cathode tempera-
ture. The equivalent circuit is given
in Figure 13.1-10.
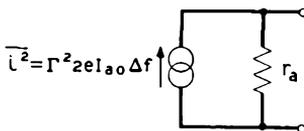


FIG. 13.1-10 Equivalent circuit for the
noise from a space-charge limited diode.
The resistance $r_a$ is noiseless.

The reader may wonder why the
formula for thermal noise does not apply directly, that is, why $\theta$ should
not be unity. The reason for this is that the space-charge limited diode
is not in thermal equilibrium, a necessary condition for the validity of
the thermal noise available power expression. By virtue of the dc current,
chemical energy from an external dc voltage source is being converted into
the heat associated with plate dissipation. On the other hand, the tube
of Figure 13.1-7 corresponded to thermal equilibrium since no dc voltages
were applied to it; thus thermal noise formulas could be applied directly.

Converting Equation (13.1-37) to an equivalent current source and com-
paring with Equation (13.1-25), we obtain an expression for $\Gamma^2$,

$$\Gamma^2 = \frac{3k\theta T_c}{eV_{ao}} \qquad (13.1\text{-}38)$$

For a typical oxide-coated cathode, with $T_c = 1000°K$, this becomes

$$\Gamma^2 = \frac{0.167}{V_{ao}} \qquad (13.1\text{-}39)$$

with $V_{ao}$ in volts.

Practical diodes usually produce more noise than is predicted by the
above expressions. In high-current diodes, the noise may be an order of
magnitude greater. This is due in part to elastic reflections of electrons
at the anode. The reflected electrons find their way back to the potential
minimum where they increase the fluctuations in the current passing the
potential minimum. Triodes, on the other hand, show good agreement
with the above expressions, since the negative grid deflects the reflected
electrons and prevents most of them from returning to the potential
minimum.[12]

The results of this section are applicable only to low frequencies, where
transit times are negligible. Effects at higher frequencies will be considered
later.

---

[12]Reference 13.8.

## 13.2   Noise in Grid-Controlled Tubes

Triodes, tetrodes, and pentodes have other sources of noise in addition to shot noise. In this section we shall study these other noise sources, and in addition we shall consider the modification of the shot noise expression so that it will apply to grid-controlled tubes.

The equivalent circuit for the noise sources of a tetrode or pentode is shown in Figure 13.2-1.[13] The reduced shot noise is given by $\overline{i_s^2}$. In tetrodes and pentodes there is partition noise $\overline{i_p^2}$ due to the random interception of current by the screen grid. The screen-grid noise current $\overline{i_{sg}^2}$ is
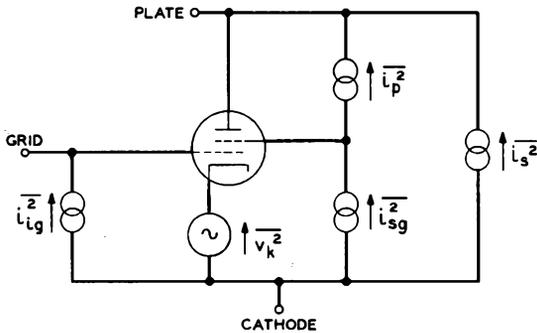


FIG. 13.2-1 Equivalent circuit for the noise of a tetrode or pentode. The noise
sources are as follows:

$i_{ig}^2$ Induced grid noise
$i_p^2$ Partition noise
$i_s^2$ Shot noise
$i_{sg}^2$ Screen grid interception noise
$v_k^2$ Cathode-coating noise

usually eliminated as a noise contributor by means of a bypass capacitor from screen grid to the cathode. The induced grid noise is represented by the generator $\overline{i_{ig}^2}$; this noise results from fluctuations of induced grid current due to the electrons in the cathode-grid region. Finally, cathode-coating noise or flicker noise is represented by the voltage source $\overline{v_k^2}$. All of these noise sources will be discussed in detail in this section.

The same equivalent circuit applies to a triode, except that the sources $\overline{i_p^2}$ and $\overline{i_{sg}^2}$ are eliminated.

Certain simplifications of the equivalent circuit are possible at high and low frequencies. Above 100 kc the flicker noise is usually negligible with

_____

[13]T. E. Talpey, Reference 13b, Chapter 4.

respect to shot noise so that $\overline{v_k^2}$ may be eliminated. Below 15 Mc the induced grid noise is usually negligible.

### (a)   Shot Noise in a Grid-Controlled Tube

To derive an expression for the shot noise in a grid-controlled tube, we consider the control grid and the cathode to comprise an equivalent diode.[14] The results of Section 13.1(d) are applied to this diode to determine the shot noise current. Assuming that the grid does not intercept or alter this current, this will then be the noise current reaching the anode and appearing in the output circuit, neglecting transit-time effects.

The reduced shot noise current for a space-charge limited diode is obtained from Equation (13.1-37) as

$$\overline{i_s^2} = 4k\theta T_c g_a \Delta f \tag{13.2-1}$$

where

$$g_a = \frac{1}{r_a} = \frac{3}{2}\frac{I_{ao}}{V_{geo}} \tag{13.2-2}$$

is the conductance of the equivalent diode consisting of cathode and control grid; $V_{geo}$ is the equivalent voltage which would have to be placed at the plane of the control grid so as to obtain the current $I_{ao}$ in a parallel-plane diode of the same electrode spacing. This voltage is obtained from Equation (5.2-13) as

$$V_{geo} = \sigma\left(V_{go} + \frac{V_{ao}}{\mu}\right) \tag{13.2-3}$$

where

$$\sigma = \left[1 + \frac{1}{\mu}\left(\frac{d_{ca}}{d_{cg}}\right)^{4/3}\right]^{-1} \tag{13.2-4}$$

We are assuming that the conditions leading up to Equations (5.2-13) and (13.1-37) are fulfilled; that is, we assume that the spacing from the potential minimum to the grid is much greater than the spacing from the potential minimum to the cathode. The transconductance $g_m$ is obtained by differentiating Equation (5.2-13) as

$$g_m = \frac{\partial I_{ao}}{\partial V_{go}} = \frac{3}{2}\sigma\frac{I_{ao}}{V_{geo}} \tag{13.2-5}$$

where use has been made of the definitions given by Equations (13.2-3)

---

[14]Reference 13.8.

and (13.2-4). Combining Equations (13.2-2) and (13.2-5), with $V_{ao}$ given by $V_{geo}$, we obtain

$$g_m = \sigma g_a \tag{13.2-6}$$

so that the noise current is given by

$$\overline{i_s^2} = \frac{4k\theta T_c g_m \Delta f}{\sigma} \tag{13.2-7}$$

using Equation (13.2-1).

Since $\sigma$ generally ranges between 0.5 and 1 and $\theta$ is approximately 2/3, Equation (13.2-7) states that the shot noise in a grid-controlled tube is equal to the thermal noise generated by a resistor whose temperature is 2/3 to 4/3 the cathode temperature, the conductance of the resistor being equal to the transconductance of the tube.

Noise calculations are often facilitated by representing the shot noise by a passive resistor at room temperature in series with the grid lead of value $R_{eq}$ and an equivalent tube with zero shot noise. The thermal noise current appearing in the plate lead due to this resistor is given by

$$\overline{i^2} = g_m^2 4kT R_{eq} \Delta f \tag{13.2-8}$$

Equating Equations (13.2-7) and (13.2-8) we obtain

$$R_{eq} = \frac{\theta T_c}{\sigma T g_m} \tag{13.2-9}$$

which, for a typical tube using an oxide-coated cathode, becomes approximately

$$R_{eq} = \frac{2.5}{g_m} \tag{13.2-10}$$

Measured values[15] of $R_{eq}$ are given in Table 13.2-1 for several common receiving tubes. All of these tubes are triode connected; that is, the plate is connected to the screen grid and suppressor grid in the case of tetrodes and pentodes.

Finally, one may express the reduced shot noise for a grid-controlled tube in terms of a space-charge reduction factor, that is,

$$\overline{i_s^2} = \Gamma^2 \cdot 2eI_{ao} \Delta f \tag{13.2-11}$$

where expressions for $\Gamma^2$ may be obtained from Equations (13.2-7) through (13.2-10)

---

[15] T. E. Talpey, Reference 13b, p. 175.

All of the above expressions may be used for the shot noise in a tetrode or pentode provided we define $\sigma$ properly. This is accomplished by re-placing $V_{ao}$ and $d_{co}$ in Equations (13.2-3) and (13.2-4) by the screen-grid voltage and the screen-grid-to-cathode spacing, respectively. $I_{ao}$ in Equation (13.2-11) is the anode current, which might be considerably less than the cathode current if the screen-grid interception is high.[16]

TABLE 13.2-1. MEASURED NOISE QUANTITIES AT 30 Mc*

| Tube (Triode Connected) | Transconductance $g_m$ Micromhos | Shot Noise, Equivalent Resistance $R_{eq}$, Ohms | Induced Grid Noise, Equivalent Conductance $G_{eq}$, Micromhos |
|---|---|---|---|
| 6AG5 | 6000 | 480 | 140 |
| 6AK5 | 5500 | 460 | 45 |
| 6AU6 | 6600 | 420 | 210 |
| 6BC5 | 5800 | 590 | 130 |
| 6BC6 | 7300 | 410 | 175 |
| 6J6 | 4400 | 720 | 60 |
| 396A /2C51 | 5400 | 550 | 40 |
| 404A | 16000 | 240 | 70 |

*Courtesy M.I.T. Press, Massachusetts Institute of Technology, Cambridge, Mass.

(b) Induced Grid Noise

At frequencies greater than 15 Mc the noise current induced in the anode circuit of a triode amplifier stage is found to be greater than that predicted by the shot noise theory of the previous section plus the additional thermal noise due to the grid resistance. This additional noise is caused by noise currents which are induced in the impedance of the grid circuit and which, in turn, cause noise voltages to appear between grid and cathode.

Figure 13.2-2 shows qualitatively the current induced in the grid circuit by the passage of a single electron from the cathode to the anode. The total current induced in the grid circuit is the sum of that arising from each of the electrons in transit. Since there are fluctuations in the rate of emission from the cathode, the total current induced in the grid circuit fluctuates so that a noise voltage appears across the impedance of the grid circuit in addition to that arising from thermal noise. This voltage causes the beam to be modulated and consequently increases the fluctuations in the current induced in the anode circuit. The induced grid noise increases as the square of the frequency; and at frequencies greater than 100 Mc, it

[16]*Ibid*, pp. 177–180.
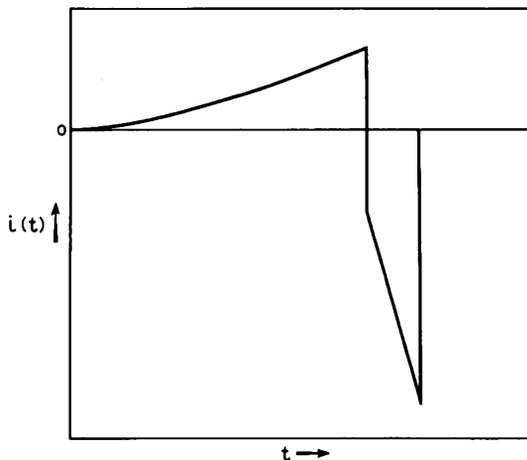
FIG. 13.2-2 Current induced in the grid circuit of a triode due to the passage of a single electron from the cathode to the anode.

is a principal limiting factor in the design of low-noise amplifiers using grid-controlled tubes.

From the physical description of the induced grid noise given above, it is evident that there is a relationship to the input conductance of the grid discussed in Section 7.3(b). The theory of that section may be applied to electrons having shot noise fluctuations with the result that the induced grid noise is given by a noise current generator of value[17]

$$\overline{i_{ig}^2} = 4k\beta T G_{in}\Delta f \tag{13.2-12}$$

where $T$ is equal to room temperature, $G_{in}$ is the input conductance discussed in Section 7.3(b), and

$$\beta = 1.43\frac{T_c}{T} \tag{13.2-13}$$

equal to approximately 5 for oxide-coated cathodes. The induced grid noise predicted by this equation arises from the shot noise fluctuations in the convection current passing through the grid. Thus, there is a correlation[18] between this noise and the shot noise in the anode current described in part (a) of this section. This correlation must be taken into account in any noise calculations.

---

[17]Reference 13.9.
[18]If the product of two noise currents has a nonzero time average, the currents are said to be correlated.

In addition to the correlated component of induced grid noise described above, there is an uncorrelated component[19] which may be as much as 70 per cent of the total induced grid noise. This uncorrelated component is thought to arise from three major sources;

1. Equation (13.2-12) is derived assuming a uniform transit time for all electrons. In reality, some electrons move closer to the negative grid wires than others, these electrons having thus a longer transit time. The randomness of this transit time introduces an additional induced grid noise.

2. A small percentage (2 to 3 per cent) of electrons elastically reflected from the anode find their way back to the grid, inducing additional noise currents. Since elastic reflections are purely random, and there is no space-charge smoothing of this current of electrons, this contributes significantly to the uncorrelated induced grid noise.

3. Electrons with insufficient emission velocity to pass the potential minimum induce current in the grid circuit if the total transit time is a significant portion of an rf cycle. This "total emission noise" is of course uncorrelated with the current passing the potential minimum.

Just as the shot noise may be replaced by a resistor $R_{eq}$ in the grid circuit, so also may we represent the total induced grid noise by an equivalent shunt conductance $G_{eq}$ in the grid circuit. This conductance is placed from the grid to the cathode; thus an rf short from grid to cathode eliminates induced grid noise. Since $R_{eq}$ is in series with the grid lead, a similar rf short does not eliminate the shot noise contribution. Measured values[20] of $G_{eq}$ are given in Table 13.2-1 for several common tubes.

(c) *Partition Noise*

Since the screen grid of a tetrode or pentode is at a positive potential, it collects a portion of the beam current, and since the electrons have random components of transverse velocity, the rate of collection of electrons by the screen grid fluctuates in time. As a result, the current passing on to the anode also fluctuates, these fluctuations being in addition to the reduced shot noise already present in the beam. Noise of this type, introduced by the random division of current between two or more positive electrodes, is known as partition noise.

As a result of partition noise, the noise generated by tetrodes and pentodes is often three to five times as great as that generated by the same tube connected as a triode.

Since the normal and transverse velocity components of electrons are uncorrelated, the shot noise and partition noise are also uncorrelated.

---

[19]T. E. Talpey, Reference 13b, pp. 166–176.
[20]*Ibid.*

The following expressions for the partition noise sources indicated in Figure 13.2-1 may be derived:[21]

$$\overline{i_p^2} = (1 - \Gamma^2) \cdot 2e\frac{I_{sgo}I_{ao}}{I_{ko}}\Delta f \qquad (13.2\text{-}14)$$

$$\overline{i_{sg}^2} = \Gamma^2 \cdot 2eI_{sgo}\Delta f \qquad (13.2\text{-}15)$$

where $I_{ao}$, $I_{ko}$, and $I_{sgo}$ are the dc anode, cathode, and screen-grid currents, respectively, and $\Gamma^2$ is the space-charge reduction factor defined by Equation (13.2-11).

If we assume that the noise source $\overline{i_{sg}^2}$ is effectively bypassed by the screen-grid capacitor, the total anode noise is given by the summation of Equations (13.2-11) (reduced shot noise) and (13.2-14) (partition noise), with the result

$$\overline{i_s^2} + \overline{i_p^2} = \left[\frac{\Gamma^2 I_{ao} + I_{sgo}}{I_{ko}}\right]2eI_{ao}\Delta f \qquad (13.2\text{-}16)$$

An upper limit to the anode noise is obtained by setting $\Gamma^2$ equal to unity in the above equation, corresponding to no space-charge smoothing of the shot noise. One obtains

$$\overline{i_s^2} + \overline{i_p^2} = 2eI_{ao}\Delta f \qquad (13.2\text{-}17)$$

the expression for full shot noise. Thus, despite partition noise effects, the total anode noise can never be greater than that corresponding to full shot noise in the anode current. Physically, this corresponds to the fact that without space-charge smoothing the electron arrival times are completely random, and random interception of some of these electrons cannot increase the randomness of a process that is already completely random.

One may define a noise equivalent grid resistor $R_{eq}$ by Equation (13.2-8) as for the triode. Its value may be obtained by equating Equations (13.2-8) and (13.2-16). This resistor replaces all the anode noise of the tube, that is, shot noise plus partition noise. Hence, values of $R_{eq}$ for tetrodes and pentodes are three to five times as large as the values given in Table 13.2-1.

### (d)   Flicker Noise and Other Miscellaneous Noise Sources

Flicker effect or cathode-coating noise becomes important at low frequencies and may determine the ultimate sensitivity of an audio amplifier. Its magnitude is inversely proportional to frequency, and it is usually negligible at frequencies above 100 kc.

Present theory[22] indicates that flicker effect is due to fluctuations of the

[21]T. E. Talpey, Reference 13b, pp. 177–180.
[22]A. van der Ziel, Reference 13b, Chapter 2.

voltage drop across a thin surface layer (depletion layer) on the oxide coating of the cathode. Because of the porous nature of the oxide coating, it is proposed that the fluctuating surface layer voltage modulates the current coming from the pores. For further discussion of the theory, the reader is referred to the reference cited.

Another possible source of noise in grid-controlled tubes is secondary-emission noise. Secondary electrons emitted from the anode which go to other electrodes result in fluctuations of the anode current. This noise is minimized by choosing proper dc electrode voltages such that the secondary electrons from the anode are returned to the anode.

Other sources of tube noise include noise due to ionized gas in the tube, poor electrode contacts, leakage currents along the surfaces of insulators, primary emission from the control grid, charges building up on the glass envelope, microphonics due to mechanical vibrations of electrodes, and hum caused by heater current induction. All of these sources of noise can be reduced to a negligible level with proper design and operating conditions.

### (e)  Amplifier Stage Noise Figure Calculation

In this part we shall derive an expression for the spot noise figure of an amplifier stage with an equivalent circuit similar to that given in Figure 13.2-3. This may be assumed to be the equivalent circuit of either a triode, tetrode, or pentode, where for simplicity we have assumed that the anode-to-grid capacitance has been neutralized and that input and output capacitances have been tuned out by shunt inductances. Thus, all impedances are purely real, for simplicity. This is done so that we can concentrate our attention on the noise properties of the amplifier without worrying about the complex circuit analysis required with reactances present.

The expression for the spot noise figure is obtained from Equation (13-4) as

$$F = 1 + \frac{N_a}{\mathcal{G}_o k T_o \Delta f} \tag{13.2-18}$$

where $N_a$ is the available noise power of the amplifier in the bandwidth $\Delta f$, and the other symbols are as defined in connection with Equation (13-4).

The available power gain $\mathcal{G}_o$ is defined as the ratio of the powers available at the output and input of the amplifier. The available power at the input is the power available from the source, given by

$$P_{a1} = \frac{|I_s|^2}{4G_s} \tag{13.2-19}$$

The available power at the output is given by

$$P_{a2} = \frac{g_m^2 |I_s|^2}{4(G_s + G_g + G_{in})^2 G_s} \tag{13.2-20}$$

where

$$G_e = \frac{1}{r_a} + \frac{1}{R_L} \qquad (13.2\text{-}21)$$

The available power gain is thus

$$\mathcal{G}_o = \frac{P_{a2}}{P_{a1}} = \frac{g_m^2 G_s}{(G_s + G_g + G_{\text{in}})^2 G_e} \qquad (13.2\text{-}22)$$

Next we calculate the available output noise due to noise sources within the amplifier. Assuming that the frequency is high, the noise sources are
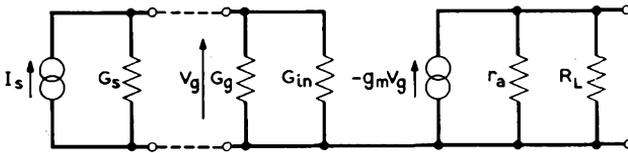


FIG. 13.2-3 Simplified equivalent circuit of a triode, tetrode, or pentode amplifier stage. The parameters are:

$I_s$ Current generator representing the source
$G_s$ Equivalent source conductance
$G_g$ Grid conductance
$G_{\text{in}}$ Grid input conductance due to beam loading (Section 7.3(b))
$r_a$ Dynamic anode resistance
$R_L$ Plate circuit load resistance

induced grid noise, anode noise due to shot noise and partition noise, and thermal noise arising from the grid conductance $G_g$ and the anode resistance $R_L = 1/G_L$. All of these noise sources are uncorrelated, except for the correlation between induced grid noise and shot noise discussed in Section (b). For simplicity we shall take the correlation between these noise sources to be zero. Expressions including this correlation and the effects of circuit reactance are derived elsewhere.[23]

The amplifier noise current produced in the grid circuit is given by

$$\overline{i_1^2} = 4kTG_g\Delta f + \overline{i_{ig}^2} = 4kT(G_g + G_{\text{eq}})\Delta f \qquad (13.2\text{-}23)$$

using the definition of $G_{\text{eq}}$ given in Section (b). The noise voltage appearing in the grid circuit is given by[24]

$$\overline{v_1^2} = \frac{\overline{i_1^2}}{(G_s + G_g + G_{\text{in}})^2} = \frac{4kT(G_g + G_{\text{eq}})\Delta f}{(G_s + G_g + G_{\text{in}})^2} \qquad (13.2\text{-}24)$$

---

[23]T. E. Talpey, Reference 13b, pp. 188–204.
[24]Note that the effect of induced currents in the grid circuit is taken into account by two separate factors. $G_{\text{in}}$ expresses the loading at the signal frequency and is noiseless. $G_{\text{eq}}$ gives the noise contribution but does not affect the loading at the signal frequency.

Note that, although $G_s$ and $G_{in}$ are not amplifier noise sources, they do influence the noise voltage developed. The noise current appearing in the anode circuit due to this noise voltage is given by

$$\overline{i_2{}^2} = g_m{}^2\overline{v_1{}^2} = \frac{4kTg_m{}^2(G_g + G_{eq})\Delta f}{(G_s + G_g + G_{in})^2} \qquad (13.2\text{-}25)$$

The noise current appearing in the anode circuit due to the shot noise, partition noise, and plate load resistor is given by

$$\overline{i_3{}^2} = \overline{i_s{}^2} + \overline{i_p{}^2} + 4kTG_L\Delta f = 4kT(g_m{}^2R_{eq} + G_L)\Delta f \qquad (13.2\text{-}26)$$

where $R_{eq}$ was defined in Section (c).

The available output noise is given by

$$N_a = \frac{\overline{i_2{}^2} + \overline{i_3{}^2}}{4G_s} = kT\Delta f\left[\frac{g_m{}^2(G_g + G_{eq})}{G_s(G_s + G_g + G_{in})^2} + \frac{g_m{}^2R_{eq} + G_L}{G_s}\right]$$
$$(13.2\text{-}27)$$

Substituting this equation and Equation (13.2-22) into (13.2-18), we obtain the desired expression for the spot noise figure:

$$F = 1 + \frac{G_g}{G_s} + \frac{G_{eq}}{G_s} + \left[R_{eq} + \frac{G_L}{g_m{}^2}\right]\frac{(G_s + G_g + G_{in})^2}{G_s} \qquad (13.2\text{-}28)$$

where it has been assumed that room temperature $T$ is equal to the standard noise temperature $T_o$ of $290°$K.

Let us calculate the noise figure of a typical high frequency (30 Mc) triode amplifier stage using this formula. Assume the following parameters:

$$G_s = 300 \text{ micromhos}$$

$$G_g = 35 \text{ micromhos}$$

$$G_{eq} = 50 \text{ micromhos}$$

$$G_{in} = 17 \text{ micromhos}$$

$$G_L = 10 \text{ micromhos}$$

$$R_{eq} = 450 \text{ ohms}$$

$$g_m = 5000 \text{ micromhos}$$

With these parameters substituted into Equation (13.2-28), one obtains

$$F = 1 + 0.117 + 0.164 + 0.186 + 0.0002 = 1.47$$

which is equal to 1.67 db. The factors in this expression are arranged in the same order as in Equation (13.2-28) so as to illustrate the relative contribution of each noise source. Immediately evident is the fact that thermal noise due to resistance in the anode circuit is negligible compared with the other quantities, and thus in general this term can be neglected.

The expression for the noise figure which applies when reactances are present is[25]

$$F = 1 + \frac{G_g}{G_s} + \frac{G_{\text{eq}}}{G_s}$$
$$+ \frac{1}{G_s}\left[R_{\text{eq}} + \frac{G_L}{g_m{}^2}\right][(G_s + G_g + G_{\text{in}})^2 + \omega^2(C_s + C_c + C_1)^2]  \quad (13.2\text{-}29)$$

where $C_s$ is the source capacitance, $C_c$ is the grid-circuit capacitance, and $C_1$ is the "cold" input tube capacitance (i.e., the input capacitance without beam loading). Here negative susceptances are represented by equivalent negative values of $C_s$ or $C_c$. Examining this equation, we see that minimum noise figure is obtained when the net inductance of the source and grid-circuit resonates with the "cold" input tube capacitance. This does not necessarily correspond to the frequency of maximum gain, due to the detuning effect of the susceptance due to beam loading.

## 13.3  Noise in Microwave Tubes

Either traveling-wave, backward-wave, or klystron amplifiers can serve as low noise amplifiers at microwave frequencies. The same basic theoretical considerations of low noise design apply to each of these devices. However, most low-noise research work has been done on the traveling-wave amplifier, with the result that this particular device has been produced with the lowest noise figures. Thus, we shall concentrate our study on the traveling-wave amplifier.

There are many sources of noise in the traveling-wave amplifier. Thermal noise arises from the attenuation of the helix. Random emission current (shot noise) and random emission velocities result in initial noise current and noise velocity on the electron beam as it enters the helix, both of which are amplified and appear as output noise. Partition noise may arise from interception on the electrodes of the electron gun or on the helix itself. Secondary electrons can also contribute noise. Most of these noise sources can be minimized in low-noise tubes by obvious means. Copper plating of the helix is used to minimize attenuation. Proper focusing of the beam minimizes beam interception, and the collector in low-noise tubes is usually operated at a positive potential with respect to the helix to recapture secondary electrons emitted from the collector. Minimizing the noise due to the emission fluctuations is much more subtle and will constitute the greater part of our discussion.

---

[25]T. E. Talpey, Reference 13b, p. 200.

## (a)  *Velocity and Current Fluctuations in the Electron Beam*

In order to study the effects of the velocity and current fluctuations in the electron beam, we shall make use of the space-charge wave theory of Section 9.3 and the Llewellyn-Peterson equations of Section 7.2. Both of these theories concern the propagation of small, sinusoidal perturbations on an electron beam. How then can we apply these theories to noise perturbations which are small but certainly not sinusoidal? Suppose we measure and plot as a function of time the instantaneous current and velocity fluctuations of an electron beam over some finite time interval. These functions may be analyzed by means of the Fourier integral to obtain an equivalent sinusoidal current and velocity for a frequency interval $\Delta f$. Thus, the propagation of noise perturbations may be determined from the propagation of the equivalent sinusoidal quantities. These equivalent sinusoidal quantities may be related to the corresponding noise fluctuations by equating the integral of the noise power density spectrum over the frequency interval $\Delta f$ to the power of the equivalent sinusoidal signal.

Let us study the propagation of such waves in a space-charge limited diode connected at the anode to a drift region. We shall first confine our study of the diode to the region from the potential minimum to the anode. The diode may be analyzed using the Llewellyn-Peterson equations of Section 7.2, where plane $A$ is taken at the potential minimum and plane $B$ at an arbitrary plane up to and including the anode. Assuming that the diode is open-circuited at the frequency of interest, the total rf beam current density $J_T$ is equal to zero in Equations (7.2-1). The entrance conditions at plane $A$ are assumed to be a convection current equivalent to full shot noise, Equation (13.1-13), and an ac velocity given by Equation (13.1-24). The initial current and velocity fluctuations are assumed to be uncorrelated, so that the equivalent sinusoidal quantities have no fixed phase relationship. Thus, we consider each initial quantity separately, independent of the other.

We may write our results directly in terms of the noise fluctuations. Since the mean-square values of the noise fluctuations are proportional to the square of the magnitudes of the equivalent sinusoidal quantities, we have from Equations (7.2-1) and (13.1-13) the following results for the shot noise:

$$\overline{i_1^2} = |E^*|^2 \overline{i_A^2} = |E^*|^2 2eI_o\Delta f$$

$$\overline{u_1^2} = \frac{1}{S^2}|H^*|^2 \overline{i_A^2} = \frac{1}{S^2}|H^*|^2 2eI_o\Delta f \qquad (13.3\text{-}1)$$

where the subscript 1 is used to denote the contribution of shot noise current fluctuations to the convection current and velocity at an arbitrary plane.

$S$ is the beam area. Similarly, we have for the uncorrelated initial velocity fluctuations, from Equations (7.2-1) and (13.1-24):

$$\overline{i_2^2} = S^2 |F^*|^2 \overline{u_A^2} = S^2 |F^*|^2 \frac{ekT_c \Delta f}{mI_o}(4 - \pi)$$

$$\overline{u_2^2} = |I^*|^2 \overline{u_A^2} = |I^*|^2 \frac{ekT_c \Delta f}{mI_o}(4 - \pi) \tag{13.3-2}$$
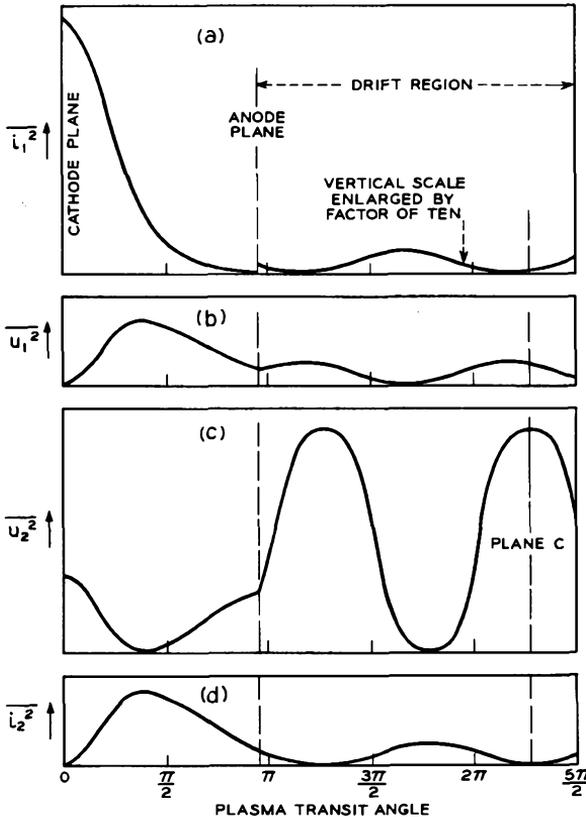


FIG. 13.3-1 Noise standing waves in a particular space-charge limited diode and the drift region following the anode. The noise due to shot noise and initial velocity fluctuations are assumed to be uncorrelated. Distance is measured in terms of plasma transit angle from the potential minimum. (a) Convection current fluctuations $\overline{i_1^2}$ due to shot noise. (b) Velocity fluctuations $\overline{u_1^2}$ due to shot noise. (c) Velocity fluctuations $\overline{u_2^2}$ due to initial velocity fluctuations at the potential minimum. (d) Convection current fluctuations $\overline{i_2^2}$ due to initial velocity fluctuations.

where the subscript 2 is used to designate the contribution of intitial veloc-
ity fluctuations to the convection current and velocity at an arbitrary
plane. The coefficients $E^*$, $H^*$, $F^*$, and $I^*$ are functions of the location of
this plane.

The results of this analysis are shown in Figure 13.3-1, calculated for a
particular diode. In Figure 13.3-1(a) is shown the mean-square value of the
convection current fluctuations $\overline{i_1^2}$ resulting from the shot noise current at
the potential minimum. In Figure 13.3-1(b) is shown the mean-square
value of the associated rf velocity fluctuation $\overline{u_1^2}$. Note that the abscissa
is the plasma transit angle measured from the potential minimum. The
location of the anode has been arbitrarily chosen. In parts (c) and (d) of
the figure are shown the corresponding quantities $\overline{u_2^2}$ and $\overline{i_2^2}$ due to the
initial velocity fluctuations at the potential minimum.

At the anode plane, the beam enters a drift region at anode potential,
and the rf quantities vary in the axial direction with an envelope deter-
mined by the plasma wavelength as in Figure 9.3-4. The maxima of $\overline{i_1^2}$
and $\overline{u_1^2}$ in the drift region are separated by a half-plasma wavelength. It
should be emphasized that these are standing wave patterns, stationary in
space. The phase and amplitude of these standing waves are obtained
using Equations (9.3-31) and (9.3-33) by matching boundary conditions
at the anode plane. These boundary conditions are that the rf convection
current and velocity must both be continuous at the anode plane.

We note from the figure that it is the initial velocity fluctuations rather
than the initial shot noise which contributes the major part of the noise of
the drifting beam. If we assume a space-charge smoothing factor $\Gamma^2$ at
the potential minimum, the contribution of the shot noise becomes even
smaller in the drifting beam. Thus, the noise produced by using this beam
unmodified in a traveling-wave tube results almost entirely from the initial
velocity fluctuations.

Since the initial velocity fluctuations contribute so greatly to the noise
in the drifting beam, let us study a scheme which has been proposed for
reducing this source of noise.[26] Assume that the anode voltage of the diode
$V_{ao}$ is small compared with the desired helix voltage. Suppose that the
beam is suddenly accelerated to the helix voltage $V_o$ at plane $C$, shown in
Figure 13.3-1. This plane is chosen at the position where the velocity
$\overline{u_2^2}$ is a maximum and $\overline{i_2^2}$ is zero. If $u_{o-}$ and $\Delta u_{2-}$ are the dc velocity and
instantaneous velocity fluctuation just before the *velocity jump*, and $u_{o+}$
and $\Delta u_{2+}$ are the corresponding quantities just after the velocity jump, we
have from conservation of energy:

$$(u_{o+} + \Delta u_{2+})^2 - (u_{o-} + \Delta u_{2-})^2 = \frac{2e}{m}(V_o - V_{ao}) \qquad (13.3\text{-}3)$$

[26]Reference 13.10.

where $V_{ao}$ is the diode anode voltage.  Expanding the left-hand side and neglecting the terms $(\Delta u_{2+})^2$ and $(\Delta u_{2-})^2$, we obtain

$$2u_{o+}\Delta u_{2+} - 2u_{o-}\Delta u_{2-} = 0$$

or

$$\frac{\Delta u_{2+}}{\Delta u_{2-}} = \frac{u_{o-}}{u_{o+}} = \sqrt{\frac{V_{ao}}{V_o}} \qquad (13.3\text{-}4)$$

Thus, the bigger the velocity jump, the larger is the reduction of the rf velocity fluctuation.
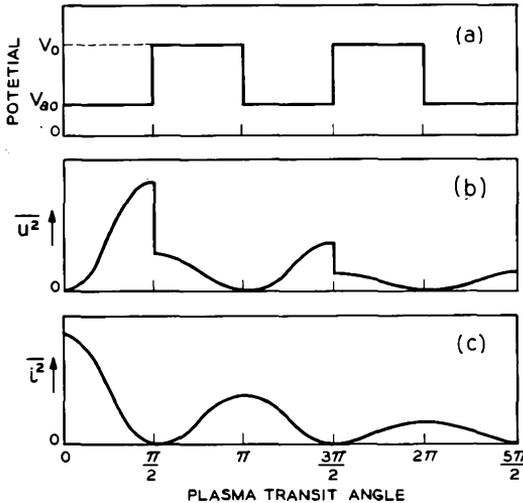


FIG. 13.3-2 Noise reduction resulting from successive velocity jumps.  (a) Potential profile along the beam.   (b) Velocity fluctuations.   (c) Convection current fluctuations.

We may carry the above noise reduction scheme even further.  Suppose that at another plane located a quarter-plasma wavelength beyond plane $C$ we suddenly decelerate the beam again.  Since the rf velocity is zero at this plane, the principle behind Equation (13.3-4) does not apply and the noise properties of the beam are essentially unchanged.  However, now we may introduce another velocity jump a quarter-plasma wavelength farther on and achieve a further noise reduction.  Thus, if this process of acceleration-deceleration is repeated $N$ times, the final noise velocity fluctuation $\overline{u_{2(N)}}^2$ is in principle related to the initial fluctuation $\overline{u_2}^2$ by

$$\frac{\overline{u_{2(N)}}^2}{\overline{u_2}^2} = \left(\frac{V_{ao}}{V_o}\right)^N \qquad (13.3\text{-}5)$$

Since the rf convection current is proportional to the rf velocity from the theory of space-charge waves, the corresponding convection current fluctuations are reduced by the same factor. Thus, one may reduce the noise due to the velocity fluctuations to any desired level merely by choosing enough velocity jump cycles. This process is illustrated in Figure 13.3-2.

Does all of the above discussion mean that we can actually produce a noiseless beam? To answer this question, we must examine the waves due to the shot noise, $\overline{i_1^2}$ and $\overline{u_1^2}$. As drawn in Figure 13.3-1 these waves do not reach their maxima and minima at exactly the same planes as those waves due to the initial velocity fluctuations, $\overline{i_2^2}$ and $\overline{u_2^2}$.[27] Thus, although the noise due to the initial velocity fluctuations can in principle be reduced
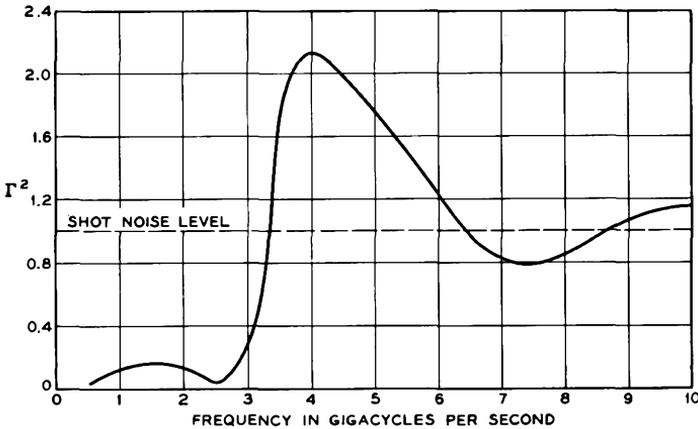


FIG. 13.3-3 Space-charge smoothing factor $\Gamma^2$ for the shot noise at the potential minimum as calculated for a typical space-charge limited diode at microwave frequencies. A pronounced dip is obtained at 2500 Mc; this particular diode has a plasma frequency of 3784 Mc at the potential minimum (Reference 13.11). (Courtesy of *J. Appl. Phys.*)

without limit by the velocity jump process, the noise waves due to shot noise cannot be correspondingly reduced. In fact, the lowest noise figure that one can obtain under the above conditions is approximately 6 db.[28]

The foregoing discussion assumed full-shot noise at the potential minimum. Numerical calculations[29] have demonstrated that a space-charge smoothing factor $\Gamma^2$ considerably less than unity may be obtained under

[27]This disparity in the phases of the standing waves can be shown to be a general principal. See Reference 13b, Chapters 3 and 5.

[28]H. A. Haus, Reference 13b, Chapter 3.

[29]Reference 13.11.

conditions suitable to a traveling-wave tube. The results of such a cal-
culation are shown in Figure 13.3-3 for a particular diode for which the
smoothing is optimized at a frequency of 2.5 Gc. The frequency for mini-
mum $\Gamma^2$ is a function of the plasma frequency at the potential minimum;
in this case the plasma frequency is equal to 3784 Mc. Different values of
the plasma frequency would cause the minimum value of $\Gamma^2$ to occur at
higher or lower frequencies. This calculation was made numerically
assuming charge discs emitted at random times and with random velocities
from a cathode, including the forces due to space charge. Physically the
smoothing process is analogous to that described in Section 13.1(d), except
that transit-time effects become important.

For reduced shot noise, uncorrelated with the velocity fluctuations, the
minimum noise figure is given by[30]

$$F = 1 + \Gamma \sqrt{4 - \pi} \frac{T_c}{T_o} \qquad (13.3\text{-}6)$$

which approaches 1 (0 db) as $\Gamma$ approaches zero. $T_c$ is the cathode tem-
perature.

One additional mechanism of noise reduction remains to be considered.
So far we have considered the initial current fluctuations and velocity
fluctuations at the potential minimum to be uncorrelated. It can be shown
in general[31] that the minimum noise figure decreases as the amount of
positive correlation increases. We can see intuitively how such correlation
permits further reduction of the noise. Correlation implies a definite
phase relation between part of the noise from each of the two noise sources
(shot noise and initial velocity fluctuations). If this phase is of the correct
value, the two noise sources can be made to cancel each other.

Correlation of the proper type may be produced in the low-velocity re-
gion of the diode just beyond the potential minimum. In this region,
thermal velocities are of the same order of magnitude as the dc velocity,
so that the physical interaction is highly nonlinear. Some ultra-low-
noise guns are designed especially to extend this low-velocity region using
proper electrode voltages so that the nonlinear noise reduction mechanism
is enhanced. Theoretical calculations[32] have indicated that such a low-
velocity region can result in a noise figure approaching 0 db. Experimental
measurements have also been made demonstrating the existence of such
correlation in beams produced by ultra-low-noise electron guns.[33]

Let us return to a further consideration of the velocity-jump noise re-

[30]R. W. Peter, Reference 13b, p. 232.
[31]H. A. Haus, Reference 13b, Chapter 3.
[32]Reference 13.12, 13.13.
[33]References 13.14, 13.15.

duction scheme illustrated in Figure 13.3-2. The discontinuities in potential produce strong electric lens effects (Section 3.1) which, despite the strong axial magnetic field used in the gun region, produce a conversion of transverse velocity fluctuations into longitudinal velocity fluctuations. Thus, in practice, it is found that this noise reduction scheme does not result in the lowest noise figures. It is found[34] that a gradual increase in voltage, of the
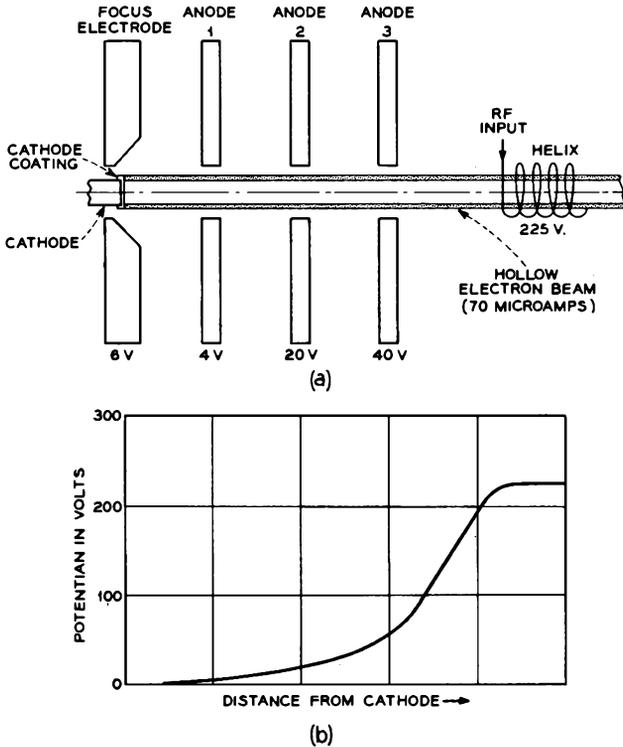


Fig. 13.3-4 Typical ultra-low-noise electron gun. (a) Gun geometry showing voltages applied to the electrodes. (b) Potential variation at the position of the electron beam.

proper variation with distance between the low-velocity diode region and the helix input, can achieve an equivalent noise reduction. In addition, the gradual taper of voltage is effective over a wider frequency band. This voltage variation is analogous to matching two transmission lines of dif-

[34]R. W. Peter, Reference 13b, Chapter 5.

ferent characteristic impedance; the exponential taper is superior electri-
cally to a sudden change of the characteristic impedance.

We may summarize the construction of a typical ultra-low-noise gun in
Figure 13.3-4. A strong axial magnetic field confines the electrons to mo-
tion in the longitudinal direction; this field is usually stronger than that in
the helix-interaction region. Voltages applied to the electrodes produce an
extended low-velocity region for the beam near the cathode. This potential
distribution results in electron emission which is primarily from the edge of
the cathode, producing a hollow beam. Because of this fact some cathodes
are made annular, and some are constructed with an additional probe
electrode at the cathode center. The oxide coating applied to the end of the
cathode also extends a small distance down the cylindrical outer surface of
the cathode as shown. It is thought that the intense charge concentration
in the beam near the corner of the cathode produces conditions similar to
those producing the dip in the space-charge smoothing factor, as predicted
in Figure 13.3-3. The extended low-velocity region permits correlation to
take place. The gradual increase of voltage to helix potential is carefully
shaped so as to result in minimum tube noise figure.

At the time of writing this chapter, the lowest spot noise figure measured
on a traveling-wave tube was 1.0 db at a frequency of 2600 Mc.[35] Probably
still lower noise figures could be obtained with additional research. One
may contrast this value with noise figures which have been measured at
microwave frequencies on microwave triodes; values of 16 db or so are
common.[36]

Low-noise traveling-wave amplifiers are commonly used as rf amplifiers
in sensitive receiver systems. In this application the traveling-wave ampli-
fier provides built-in protection from burnout of the crystal detector follow-
ing it. As the rf power input to the tube is increased, the rf power output
reaches a limiting or saturated value, this value being less than that re-
quired to damage the crystal diode.

(b)  *Description of the Western Electric GA53851 Low-Noise Traveling-
Wave Amplifier*

The Western Electric GA53851 is a typical ultra-low-noise traveling-
wave amplifier with a noise figure of 4.9 db at frequencies between 5 and
6 Gc. The tube is shown in Figure 13.3-5. Part (a) of the figure shows an
outline of the tube together with the corresponding variation of the axial
magnetic field used to focus the electron beam. It is to be noted that the
magnetic field is peaked in the vicinity of the cathode in order to restrict

---

[35]Reference 13.16.
[36]Reference 13.17.
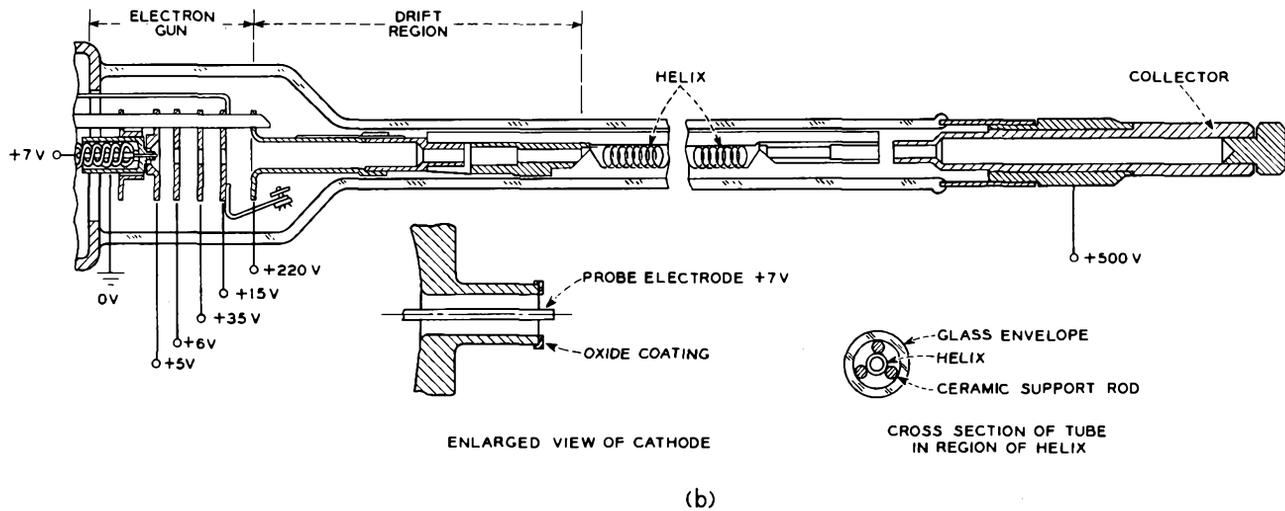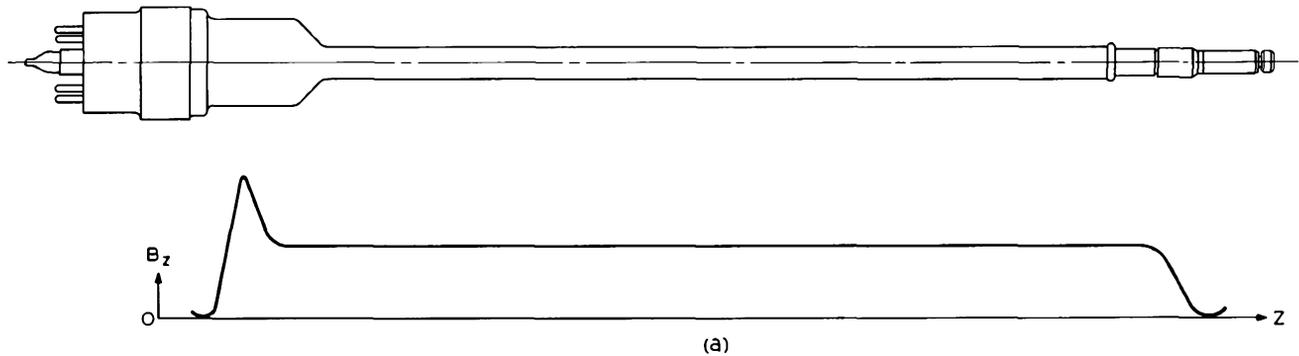
FIG. 13.3-5 Western Electric GA53851. The complete package with magnetic circuit weighs 36 pounds and is 34 cm long. (a) Tube outline showing the corresponding variation of the axial magnetic field. (b) Cross-sectional drawing of the vacuum tube alone showing the values of the applied voltages. The waveguide couplers and magnetic circuit are similar to those of Figure 10.3-7.

transverse motion of the electrons as much as possible in the low velocity region of the electron gun.

The general construction of the tube is shown in Figure 13.3-5(b), exclusive of the waveguide couplers and magnetic circuitry. These additional parts are somewhat similar to those previously shown in Figure 10.3-7. The cathode coating is applied to an annular region at the end of a cylindrical sleeve of nickel. Coaxial with the cathode sleeve is a probe electrode with a positive bias of about 7 volts. The voltages applied to the other electrodes of the electron gun are also shown in the figure. These voltages lead to a potential profile along the electron beam much like that of Figure 13.3-4(b).

The operating characteristics of the tube are presented in Table 13.3-1. The beam current is typically low to minimize the total beam noise at the entrance to the helix. Despite the low beam current, high gain is obtained merely by making the helix long enough. The collector voltage (500 volts) is made much higher than the helix voltage (220 volts) so as to capture within the collector all secondary electrons emitted from the collector. With a lower collector voltage, some secondary electrons from the collector would reach the region of the beam which is inside the helix, and this has been found to degrade the noise figure.

TABLE 13.3-1. WE GA53851 TYPICAL OPERATING CHARACTERISTICS

Frequency, Mc................................5000 to 6000
Noise figure, db..................................4.9
Small-signal gain, db.............................21
Beam voltage, volts..............................220
Beam current, ma.................................60
Saturation power output, mw......................0.10

The helix is copper plated molybdenum wire of 0.064 mm diameter. The helix is 13.3 cm long, with a pitch of 0.175 mm and a mean diameter of 1.5 mm. It is glazed to three ceramic rods which support it in the glass envelope of this tube.

The electron gun forms a hollow beam of 0.51 mm inner diameter and 0.74 mm outer diameter. Since the electron trajectories very nearly follow the magnetic field lines, and since the magnetic field decreases from 1250 gauss at the electron gun to 600 gauss in the region of the helix, the beam expands to approximately $1\frac{1}{2}$ times its original diameter by the time it is within the helix.

The complete package of tube, magnet, and waveguide couplers weighs 36 pounds and is 34 cm long.

## PROBLEMS

13.1 A temperature-limited diode is connected in series with a resistance $R$ and a battery. The resistance $R$ is at a temperature of 290°K. The dc diode current $I_o$ can be adjusted by varying the cathode temperature. Suppose it is adjusted so that the mean-square noise voltage $\overline{v^2}$ across $R$, in a frequency band $\Delta f$, is just twice the value that existed before the diode was turned on. What is the dc voltage in volts across $R$ with the diode operating under these conditions?
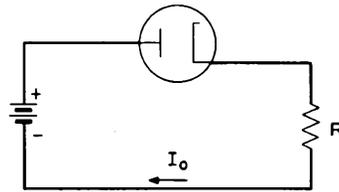


Problem 13.1

13.2 Derive an expression for the mean-square noise voltage appearing at the terminals of a series connection of two resistances $R_1$ and $R_2$ at temperatures $T_1$ and $T_2$, respectively.
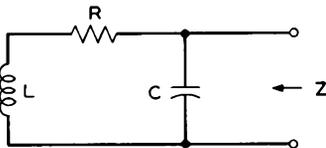


Problem 13.3

13.3 A resistance $R$, inductance $L$, and capacitance $C$ are connected together as shown in the figure. The thermal noise generated by the resistance gives rise to voltage fluctuations across the terminals of the capacitance. The input impedance as seen at the terminals of the capacitance is denoted by the symbol $Z$.

(a) Using one of the noise equivalent circuits of Figure 13.1-2, derive an expression for the mean-square voltage fluctuaiton across the capacitance.

(b) Show that the noise voltage across the capacitance is equal to that of a resistance of value equal to the real part of $Z$.

13.4 Two identical resistances, each of resistance $R$, are connected together; both are at the same temperature. Each resistance generates thermal noise which is dissipated in the other resistance. A capacitance $C$ shunts the two resistances; this capacitance represents the stray capacitance inevitably present in any physical circuit.
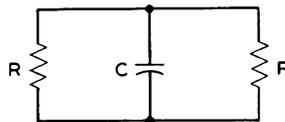


Problem 13.4

(a) Show that the total noise power flowing from one resistance to the other, as integrated over the entire frequency range from zero to infinity, is given by $kT/2RC$.

(b) Show that the average stored energy in the capacitance due to the noise fluctuations is given by $\frac{1}{2}kT$.

13.5 A resistive pad inserted into a transmission line is designed such that its image impedances are equal to the characteristic impedance of the transmission line. The definition of noise figure may be generalized to apply to passive networks,

where the available power gain is less than or equal to unity. Find an expression for the noise figure of a resistive pad inserted into a lossless transmission line which is matched at both ends. What is the noise figure of a pad designed to produce an attenuation of 3 db?

**13.6** An amplifier consists of two stages of amplification. The first stage has an available power gain $G_1$ and a noise figure $F_1$. The second stage has an available power gain $G_2$ and a noise figure $F_2$.

(a) Show that the noise figure of the whole amplifier $F_{12}$ is given by

$$F_{12} = F_1 + \frac{F_2 - 1}{G_1}$$

(b) If the first stage has a noise figure of 4 db and an available power gain of 20 db, calculate the maximum noise figure of the second stage for an overall amplifier noise figure of 4.5 db.

**13.7** Assuming that one has a power meter capable of measuring noise power over the appropriate frequency bandwidth, explain how one could measure the induced grid noise and the reduced shot noise for a triode with the equivalent circuit given in Figure 13.2-3.

**13.8** Calculate $R_{eq}$ for a pentode, including the partition noise. The control grid interception is negligible.

$$g_m = 5000 \text{ micromhos}$$

$$I_{ao} = 7 \text{ ma}$$

$$I_{sgo} = 2 \text{ ma}$$

**13.9** The noise figure of the amplifier in Figure 13.2-3 is a function of the source conductance $G_s$. Show that the lowest noise figure is obtained for $G_s$ given by

$$G_s = \sqrt{\frac{G_g + G_{eq}}{R_{eq} + G_L/g_m{}^2} + (G_g + G_{in})^2}$$



MATCHED
TERMINATION
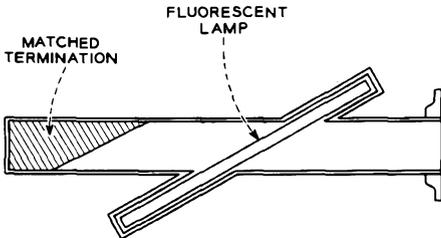
FLUORESCENT
LAMP

Problem 13.10

**13.10** In the fluorescent lamp noise source shown in the figure, the matched termination causes a noise power $kTB$ to flow to a matched load whether the lamp is on or not. When the lamp is on, an *additional* noise power equal to 38 times $kTB$ flows to the load. Suppose that the noise source is connected to the matched input of a traveling-wave amplifier (TWT), and a detector which measures noise power over a band $B$ is connected to the output of the tube. The gain of the traveling-wave amplifier is assumed to be uniform over

the band $B$. Let

$$X = \frac{\text{noise power output from TWT with lamp on}}{\text{noise power output from TWT with lamp off}}$$

Show that the noise figure of the amplifier is given by

$$F = \frac{38}{X - 1}$$

## REFERENCES

Two general references to the material of this chapter are:

13a. A. van der Ziel, *Noise*, Prentice-Hall, Inc., New York, 1954.

13b. L. D. Smullin and H. A. Haus (Eds.), *Noise in Electron Devices*, Technology Press of M.I.T. and John Wiley and Sons, Inc., New York, 1959.

Other references covering specific items are:

13.1 J. M. Pettit and M. M. McWhorter, *Electronic Amplifier Circuits, Theory and Design*, McGraw-Hill Book Co., Inc., New York, pp. 283–285, 1961.

13.2 *IRE Standards on Electron Tubes:* Methods of Testing 1962, Part 9, 1962.

13.3 R. L. Forward and F. Richey, "Effects of External Noise on Radar Performance," *Microwave Journal* **III**, 73–80, December, 1960.

13.4 W. B. Davenport, Jr., and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Co., Inc., New York, 1958.

13.5 H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.* **32**, 110–113, July, 1928.

13.6 J. R. Pierce, "Noise in Resistances and Electron Streams," *Bell System Tech. J.* **27**, 158–174, January, 1948.

13.7 A. J. Rack, "Effect of Space Charge and Transit Time on the Shot Noise in Diodes," *Bell System Tech. J.* **17**, 592–619, October, 1938.

13.8 D. O. North, "Fluctuations in Space-Charge Limited Currents at Moderately High Frequencies," Part II, "Diodes and Negative Grid Triodes," *RCA Rev.* **4**, 441–472, April, 1940; **5**, 106–124, July, 1940.

13.9 D. O. North and W. R. Ferris, "Fluctuations Induced in Vacuum Tube Grids at High Frequencies," *Proc. IRE* **29**, 49–50, February, 1941.

13.10 D. A. Watkins, "Noise Reduction in Beam Type Amplifiers," *Proc. IRE* **40**, 65–70, January, 1952.

13.11 P. K. Tien and J. Moshman, "Monte Carlo Calculation of Noise Near the Potential Minimum of a High-Frequency Diode," *J. Appl. Phys.* **27**, 1067–1078, September, 1956.

13.12 A. E. Siegman, D. A. Watkins, and H. C. Hsieh, "Density-Function Calculations of Noise Propagation on an Accelerated Multi-velocity Electron Beam," *J. Appl. Phys.* **28**, 1138–1148, October, 1957.

13.13 A. W. Shaw, A. E. Siegman, and D. A. Watkins, "Reduction of Electron Beam Noisiness by Means of a Low-Potential Drift Region," *Proc. IRE* **47**, 334–335, February, 1959.

13.14 A. Zacharias and L. D. Smullin, "Noise Reduction in Electron Beams," *Trans. IRE* **ED-7**, 172–173, July, 1960.

13.15 J. M. Hammer, "Measured Values of Noise Spectra, S and Π, of Ultra-Low-Noise Beams," *Proc. IRE* **51,** 390–391, February, 1963.

13.16 J. M. Hammer and E. E. Thomas, "Traveling-Wave-Tube Noise Figures of 1.0 db at S-Band," *Proc. IEEE* **52,** 207, February, 1964.

13.17 M. T. Vlaardingebroek, "Small-Signal Performance and Noise Properties of Microwave Triodes," *Philips Res. Reports* **15,** 124–221, April, 1960.

# Chapter 14

## CONDUCTION THROUGH A GAS
## AT LOW CURRENTS

Gas tubes are a class of electron tubes in which conduction takes place through a partially ionized gas between the electrodes. Some types of gas tubes include: rectifiers, voltage-reference tubes, voltage-regulator tubes, counting tubes, tubes used as switching elements, storage elements, indicator lamps, and sources of illumination. Gas lasers also make use of a gas discharge. For each type of service the design of the tube is different.

As an introduction to the subject of gas tubes we shall discuss in the present chapter some of the mechanisms involved in gas discharges at low currents between unheated electrodes. The range of currents we shall consider extends from perhaps a few electrons per second to a few microamperes. In the two chapters that follow we shall consider discharges at higher currents, and we shall describe a number of specific devices.

Both ions and electrons contribute to the conduction of current in a gas tube, their motions being relatively complex because of frequent collisions with other particles in the interelectrode space. Since many of the processes taking place in a discharge depend on particle collisions, and since collision processes are inherently statistical in nature, it is usual to describe the discharge phenomena in terms of average particle behavior rather than individual particle histories. In regions of uniform electric field intensity a statistical description of the particle behavior in terms of the average particle density, the particle drift velocity, and the distribution of particle energies can sometimes be given. Some data are also available on the excitation and ionization caused by electrons moving through a gas under the influence of a uniform electric field. However, discharges often involve regions of rapidly changing fields, and when this is the case, much less is generally known about the particle behavior.

In addition to ions and electrons, photons and excited atoms called metastables often play an important part in the discharge.

Many gas tubes have cathodes that are not heated and hence are called cold-cathode tubes. When a discharge is established in such a tube, an electron leaving the cathode gives rise to a number of discharge products, ions, metastables, and photons, as well as additional electrons. Many of the ions are drawn to the cathode by the electric field, and some of the metastables and photons reach the cathode. Each has a certain probability, less than one, of releasing another electron. At currents at which cold-cathode tubes are normally operated, the current drawn by the tube must be limited by a series resistance, and the voltage across the tube adjusts itself so that the average number of discharge products generated by an electron leaving the cathode is sufficient to cause the release of one more electron from the cathode. Electron emission is thus supported by the discharge itself, and the discharge is said to be self-sustaining.

In order to establish a self-sustained discharge in a cold-cathode tube, the voltage applied to the tube must be raised above a value known as the breakdown voltage. Over a range of currents covering several orders of magnitude up to about 10 microamps, the anode-to-cathode voltage of the self-sustained discharge is equal to the breakdown voltage, the condition for breakdown being the same as that for a self-sustained discharge at low currents. At higher currents, above about 10 microamps, the sustaining voltage falls below the breakdown voltage, and a visible glow can be seen in the interelectrode space. Above 200 microamps the glow is fairly intense, and the discharge is said to be a glow discharge. Most cold-cathode tubes are normally operated in the glow-discharge region.

It should be noted at this point that provided the cathode area is sufficiently large, the discharge current for a given applied voltage is nearly independent of the cathode area at low discharge currents. This occurs because a portion of the cathode surface will always have slightly more favorable surface conditions for electron emission than the rest of the cathode, and the discharge tends to start opposite this part of the cathode. Once the discharge is initiated, the ion bombardment of the cathode in the region of the discharge tends to keep this part of the cathode somewhat more free of absorbed contaminent gases, with the result that it remains a better electron emitter. Thus for discharge currents below 200 microamps, the remainder of the cathode surface plays very little part in the discharge, provided the cathode area is sufficiently large. It is therefore customary to describe the discharge at low discharge currents in terms of the total discharge current rather than the discharge current density.

Voltage-reference tubes, voltage-regulator tubes, counting tubes, and

neon indicator lamps are all cold-cathode devices. Some switching tubes and storage elements are also cold-cathode devices.

The gas fillings of cold-cathode tubes are usually noble gases, since these do not interact with the electrode material, and more stable electrical characteristics can be obtained. Often neon, or argon, or a mixture of neon and argon are used. One material frequently used for the cathodes of voltage-reference and voltage-regulator tubes is molybdenum, since it can be processed to give a particularly clean surface. Recently niobium has also been shown to give very stable cathode surfaces when used with the noble gases. Cold-cathode tubes used in switching applications often have oxide-coated nickel cathodes and are filled with a mixture of neon and argon.

In the first sections of this chapter we shall describe the motions of ions, electrons, and metastables through the noble gases. Later we shall discuss the low-current discharge in neon gas between planar molybdenum electrodes. In Chapter 15 we shall discuss the glow discharge and some specific cold-cathode tubes. In Chapter 16 we shall describe a quite different form of discharge known as the hot-cathode discharge.

In this chapter and in the remaining chapters we shall consider only dc discharges in which both the electrode voltages and currents are constant with time. Consequently we shall omit the subscript "o" which was used in previous chapters to designate dc quantities.

## 14.1  The Kinetic Nature of a Gas

The molecules of the noble gases are single atoms which for many purposes behave like hard elastic spheres. They travel at average velocities of the same order of magnitude as the speed of sound in the gas and experience frequent elastic collisions with each other. At pressures normally used in cold-cathode tubes the average distance traveled in the interval between collisions is very large compared with the molecular diameter and small compared with typical electrode dimensions and spacings.

The density of molecules in a gas at 0°C is equal to $3.54 \times 10^{16}\, p$ molecules per cubic centimeter, where $p$ is the pressure in millimeters of Hg. Cold-cathode tubes are often filled to pressures of the order of 10 to 100 mm of Hg. (1 mm of Hg is equivalent to 1/760 of atmospheric pressure or 133 newtons/meter²). At $p = 50$ mm of Hg there is an average volume of space per molecule equal to the volume of a cube of side 83 angstroms (one angstrom $= 10^{-8}$ cm). In this chapter and in subsequent chapters we shall use $p$ to designate pressure in millimeters of Hg and $P$ to designate pressure in mks units.

Approximate molecular diameters[1] for the noble gases neon and argon

---

[1] *Viscosity Diameters, Landolt-Bornstein Tables*, Julius Springer, Berlin, 1950, Vol. 1, Part 1, p. 325.

are 2.67 angstroms and 3.76 angstroms, respectively. Equation (14.1-8) below gives the mean distance traveled by a molecule between collisions (mean free path) as $1/(\sqrt{2}n\pi d^2)$, where $n$ is the number of molecules per unit volume, and $d$ is the molecular diameter. Using the approximate molecular diameters given above we find that at 50 mm of Hg and 0°C, the mean free path of a neon molecule in neon is 18,000 angstroms, and the mean free path of an argon molecule in argon is 9000 angstroms.

At 0°C the average velocity of neon molecules is 538 meters/sec, and the average velocity of argon molecules is 381 meters/sec. We shall find shortly that the average kinetic energy of the molecules is directly proportional to the absolute temperature of the gas.

From statistical mechanics we learn how the molecular velocities are distributed. The law that expresses this distribution is known as the Maxwell-Boltzmann distribution law.[2] It states that the probability $M$ that a molecule has velocity $u$ with components in the range $u_x$ to $u_x + du_x$, $u_y$ to $u_y + du_y$, $u_z$ to $u_z + du_z$ is given by

$$M = (2\pi kT/m)^{-3/2}\epsilon^{-mu^2/2kT}du_x du_y du_z \qquad (14.1\text{-}1)$$

where $u^2 = u_x^2 + u_y^2 + u_z^2$, $k$ = Boltzmann's constant, $T$ is the absolute temperature of the gas, and $m$ is the mass of the molecules.

It is convenient to imagine a velocity space in which three rectangular coordinate axes measure the velocity components $u_x$, $u_y$, and $u_z$. Each molecule in a given enclosed volume can at any time be represented by a point in such a space, and from Equation (14.1-1) we see that the fractional density of points in the space, or the number of points per unit volume di-
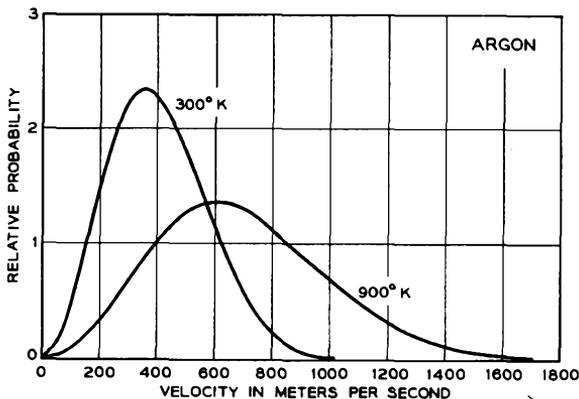


FIG. 14.1-1 Velocity distributions of argon molecules for gas temperatures of 300°K and 900°K.

[2]Reference 14.1, p. 90; Reference 2, p. 52.

vided by the total number of points, is equal to $(2\pi kT/m)^{-3/2}\epsilon^{-mu^2/2kT}$. It follows that the fraction of the total points lying in a spherical shell of radius $u$ and thickness $du$, and hence the fraction of the total number of molecules having velocities in the range $u$ to $u + du$, is given by

$$f(u)du = (2\pi kT/m)^{-3/2}\epsilon^{-mu^2/2kT}4\pi u^2 du \qquad (14.1\text{-}2)$$

The velocity distribution given by this expression for $T = 300°K$ and $T = 900°K$ is plotted in Figure 14.1-1 for argon molecules.

Using Equation (14.1-2), the average kinetic energy of the molecules is easily shown to be

$$(\text{k.e.})_{\text{avg}} = 2\pi m(2\pi kT/m)^{-3/2}\int_0^\infty u^4\epsilon^{-mu^2/2kT}du$$

$$= (3/2)kT \qquad (14.1\text{-}3)$$

The average kinetic energy of the molecules is thus directly proportional to the absolute temperature of the gas and is independent of the mass of the molecules and hence of the kind of gas.

In a similar manner, the average molecular velocity can be shown to be given by

$$u_{\text{avg}} = 4(kT/2\pi m)^{1/2} = (8kT/\pi m)^{1/2} \qquad (14.1\text{-}4)$$

Next let us determine the pressure on the walls of the containing vessel. From Equation (14.1-1) we can determine the fraction of the total number of molecules with velocity in the range $u_x$ to $u_x + du_x$. This fraction is given by

$$F(u_x)du_x = (2\pi kT/m)^{-3/2}\left[\int_{-\infty}^\infty \int_{-\infty}^\infty \epsilon^{-mu^2/2kT}du_y du_z\right]du_x$$

$$= (2\pi kT/m)^{-1/2}\epsilon^{-mu_x^2/2kT}du_x \qquad (14.1\text{-}5)$$

The number of molecules striking unit area of surface normal to the $x$ direction per unit time with velocities in the range $u_x$ to $u_x + du_x$ is $u_x nF(u_x)du_x$, where $n$ is the number of molecules per unit volume. If the molecules are reflected from the wall so that their $x$ component of velocity is reversed, each molecule suffers a change in momentum $2mu_x$ upon collision with the wall. The pressure is therefore given by

$$P = 2mn(2\pi kT/m)^{-1/2}\int_0^\infty u_x^2\epsilon^{-mu_x^2/2kT}du_x = nkT \qquad (14.1\text{-}6)$$

(Actually collisions between the gas molecules and the walls are not always elastic and, in fact, sometimes result in the molecule "sticking" to the surface. However, in a steady-state condition in which the temperature of the walls equals the temperature of the gas, the current of molecules

arriving at the surface equals the current leaving, and the average energy of the incident molecules equals that of the departing molecules. Hence, for the purpose of determining the pressure on the walls, it is possible to treat the interactions at the surface as though only elastic reflections are taking place.)

Rearranging Equation (14.1-6), we find that $n$, the number of molecules per unit volume, is equal to $P/kT$. Since $P/kT$ is independent of the mass of the molecules, we can conclude that for the same conditions of temperature and pressure all gases have the same number of molecules per unit volume. This result is known as Avogadro's Hypothesis.

If we let $N$ equal the number of molecules in one gram molecular weight of gas and let $V$ be the volume occupied by this amount of gas, then

$$P = (N/V)kT$$

or

$$PV = RT \tag{14.1-7}$$

where $R = Nk$ is a universal constant for one mole of gas equal to 8.31 joules/°K. Equation (14.1-7) is known as the equation of state for a perfect gas.

As a final point let us consider the average distance traveled by molecules between collisions. We shall first assume that only one molecule is moving and that the rest are stationary and randomly distributed throughout space with an average density of $n$ molecules per cubic meter. If the molecules have an effective diameter $d$ for the collision process, a collision takes place whenever the center of the moving molecule passes within $d$ meters of the center of a stationary molecule. The effective volume swept out by the moving molecule per unit time is therefore $u\pi d^2$ cubic meters per second, where $u$ is the velocity of the moving molecule. The average number of collisions made by the moving molecule per unit time is $nu\pi d^2$, and the mean distance traveled between collisions (mean free path) is $u/nu\pi d^2 = 1/n\pi d^2$.

This derivation of the mean free path is only an approximate one, however, for we have neglected the effect of the motion of the other molecules. When this additional motion is taken into account, it can be shown that the mean free path $L$ is given by[3]

$$L = \frac{1}{\sqrt{2}\,n\pi d^2} \tag{14.1-8}$$

[3]Reference 14.1, p. 110.

## 14.2   Diffusion and Mobility

### (a)   Diffusion

When two dissimilar gases are mixed in such a manner that concentration gradients exist within the mixture, the concentration gradients tend to become diminished by a process known as diffusion.  As an example of this process we shall consider what happens when the gases CO and $N_2$ are mixed.  We shall assume that the gases are initially separated by a partition as in the apparatus shown in Figure 14.2-1 and that the pressure on both
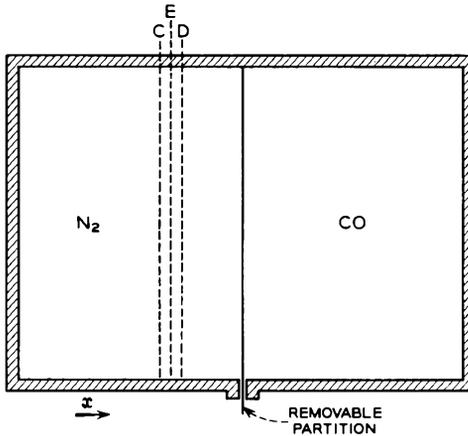


FIG. 14.2-1  A chamber with removable partition.

sides of the partition is the same.  The gases CO and $N_2$ have been selected because their molecules have the same mass and travel with the same average velocity.  This means that each diffuses into the other at approximately the same rate, which considerably simplifies the problem.

When the partition is removed, there is a net flow of $N_2$ molecules to the right and a flow of CO molecules to the left.  (We shall assume that there is no turbulence or convection within the gas.)  The flow results from the thermal motions of the molecules and continues until the partial pressure of the two gases is uniform throughout the entire chamber.

Diffusion processes are always characterized by a net flow of the diffusing particles away from regions of high concentration.  To explain this, let us first calculate the current density of $N_2$ molecules crossing a plane within the gas.  Using Equation (14.1-5), we find that the current of $N_2$ molecules crossing unit area within the gas is given by

$$n \int_0^\infty u_x F(u_x) du_x = n(kT/2\pi m)^{1/2} = n u_{avg}/4 \qquad (14.2\text{-}1)$$

where $n$ is the density of $N_2$ molecules and $u_{avg}$ is the average molecular velocity given by Equation (14.1-4). Hence the current of $N_2$ molecules crossing a plane within the gas is proportional to the density of $N_2$ molecules at the plane and to the average molecular velocity.

Next let us consider the flow of $N_2$ molecules across planes $C$ and $D$ in Figure 14.2-1. If the partial pressure of $N_2$ molecules at plane $C$ is greater than that at plane $D$, the current of $N_2$ molecules crossing plane $C$ in the direction from left to right will be greater than that crossing plane $D$ from right to left. As a result, there will be a net current of $N_2$ molecules flowing in the direction from left to right across plane $E$ midway between planes $C$ and $D$.

As a very rough approximation let us assume that the total current of $N_2$ molecules crossing plane $E$ from left to right is characteristic of the density of $N_2$ molecules $L$ units to the left of plane $E$, where $L$ is the mean free path. Similarly, let us assume that the total current of $N_2$ molecules crossing plane $E$ from right to left is characteristic of the density of $N_2$ molecules $L$ units to the right of plane $E$. Then the net current density $J$ of $N_2$ molecules crossing plane $E$ is given by

$$
\begin{aligned}
J &= \frac{u_{avg}}{4}[n_{(x_o-L)} - n_{(x_o+L)}] \\
&= \frac{u_{avg}}{4}\left(n_{x_o} - L\frac{dn}{dx} - n_{x_o} - L\frac{dn}{dx}\right) \\
&= \frac{-Lu_{avg}}{2}\frac{dn}{dx}
\end{aligned}
\tag{14.2-2}
$$

where the $x$ direction is taken normal to the planes $C$, $E$, and $D$ and the coordinate of plane $E$ is given by $x = x_o$. Although this equation is only approximate, it is apparent that we can express the current density $J$ in the form

$$
J = -D\frac{dn}{dx}
\tag{14.2-3}
$$

where $D$ is a constant of dimensions length$^2$/time. The constant $D$ is known as the diffusion coefficient. More generally, diffusion takes place in three dimensions, and we can write that

$$
\mathbf{J} = -D\nabla n
\tag{14.2-4}
$$

Since the $N_2$ molecules are conserved, we can invoke an equation of continuity on them, namely that

$$
\nabla\cdot\mathbf{J} = -\frac{\partial n}{\partial t}
\tag{14.2-5}
$$

Combining Equations (14.2-4) and (14.2-5), we obtain

$$D\nabla^2 n = \frac{\partial n}{\partial t} \tag{14.2-6}$$

This equation is known as the diffusion equation. It is of similar form to that which applies to the flow of heat in conducting bodies, and a mental picture of a particular diffusion process can often be obtained in terms of heat flow in conducting bodies of similar geometry. (In the case of heat flow, $n$ would be replaced by temperature or heat energy, and $J$ would represent flow of heat energy.)

The diffusion coefficient $D$ is inversely proportional to the total gas pressure, and experiment shows that it is not in general critically dependent on the relative composition of the two gases. An approximate expression for $D$ is given by Equations (14.2-2) and (14.2-3), namely $D = Lu_{avg}/2$. A more accurate expression that takes into account the distribution of molecular velocities is given by[4]

$$D = \frac{3\pi\sqrt{2}}{32}Lu_{avg}\sqrt{\frac{m + M}{M}} \tag{14.2-7}$$

where it is assumed that molecules of mass $m$ and average velocity $u_{avg}$ are diffusing into a gas of molecular mass $M$, and where the molecules are assumed to behave like hard elastic spheres upon collision with one another. The diffusion coefficient for electrons diffusing through a gas is given by[5] $D = Lu_{avg}/3$, where $u_{avg}$ is the average electron velocity, and $L$ is the electron mean free path for collisions with the gas molecules.

Diffusion takes place in a number of physical processes which are characterized by random particle motion and mean free paths. When ions, metastable atoms, or electrons are present in a gas, there is a tendency for nonuniformities in the particle concentration to be evened out as a result of diffusion.

If a boundary of the chamber is also a sink for the diffusing particles, as is the case when metastable atoms or ions diffuse through a gas to a metal surface, the boundary condition $n = 0$ applies at the surface.

It is sometimes useful to regard diffusing molecules as being driven by their partial pressure gradient. To understand this view of the process let us first consider the diffusion of a gas through a porous carbon block as illustrated in Figure 14.2-2. Here again the gas penetrates the block because of the random motion of the gas molecules. If the block has unit cross

---

[4]Reference 14.1, Sections 109, 267.

[5]See Reference 14.1, Section 269, for a derivation of the electron mobility. Using Equations (14.1-4) and (14.2-12) of the present text, the above expression for the electron diffusion coefficient can be obtained from the electron mobility.
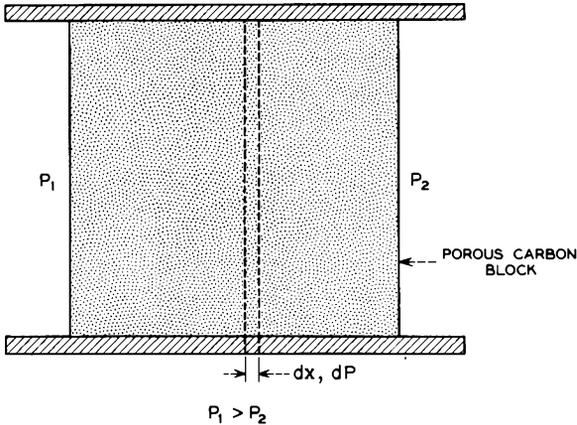
FIG. 14.2-2 A porous carbon block. The left-hand side of the block is in contact with a gas at pressure $P_1$, and the right-hand side is in contact with the same gas at pressure $P_2$.

section, the gradient of partial pressure of the gas $dP/dx$ in the direction of diffusion is equal to the resistive force per unit length that the block offers to the flow of gas. The partial pressure difference $dP$ over the length $dx$, effectively exerts an average force $dP/ndx$ on each gas molecule, where $n$ is the density of gas molecules in the slab of thickness $dx$.

Similarly in the diffusion of $N_2$ into CO the gradient of the partial pressure of the $N_2$ molecules equals the resistive force per unit area and per unit length in the direction of diffusion offered by the CO molecules. In this case the resistance offered by the CO molecules arises from the partial pressure gradient of the CO molecules and is directed oppositely to the partial pressure gradient of the $N_2$ molecules. We shall find this view of the diffusion process useful in the discussion of the relationship between diffusion and mobility given below.

### (b)   Mobility

Next let us consider the motion of positively charged atomic ions of the noble gases through their parent gases. We shall assume that the density of the ions is sufficiently small that the field resulting from the charge of an ion has negligible effect on the motion of the other ions. If no external fields are applied, the ions behave much like their parent gas molecules. They have essentially the same mass as the parent molecules and exhibit much the same velocity distribution. (However, their cross section for collision is probably somewhat larger than that of the parent molecules.) Any net motion of the ions in this case results from diffusion.

If, on the other hand, there exists a uniform electric field intensity $E$ in the region, the ions acquire a drift velocity in the direction of the field. If the ions are singly charged, the electric field acts on them with a force $eE$ newtons at all times and gives them an acceleration $eE/m$ in the direction of the field, where $m$ is the mass of an ion. In the steady-state condition the energy gained by the ions in drifting under the influence of the field is transmitted to the gas molecules in elastic collisions, so that the average drift velocity of the ions is a function only of the magnitude of the electric field intensity and not of how long the ions have drifted under the influence of the field. Consequently at pressures normally used in cold-cathode tubes the power expended by the external source applying the field appears largely as heat energy of the neutral gas molecules rather than as kinetic energy of the ions.

For weak electric fields the average drift velocity $\bar{u}$ of the ions is linearly proportional to the applied electric field intensity $E$ and can be expressed as

$$\bar{u} = \mu E \qquad (14.2\text{-}8)$$

where $\mu$ is a constant known as the mobility of the ions.

The resistance which ions encounter in drifting through a gas under the influence of a weak electric field is essentially the same as the resistance encountered by ions diffusing through the gas. It is reasonable therefore to expect the existence of a direct correlation between the diffusion coefficient $D$ and the mobility $\mu$.

In the diffusion of ions there is effectively a force $dP/ndx$ acting on each ion causing it to drift in the direction of the partial pressure gradient, the $x$ direction, where $P$ measures the partial pressure of the ions, and $n$ measures the ion density. It is convenient to associate a velocity $u'$ with the diffusion process such that $J = nu' = -Ddn/dx$, where $J$ is the number of ions crossing unit area normal to the partial pressure gradient per second, and $n$ is the ion density. The velocity $u'$ can be expressed in the form

$$u' = -\frac{D}{n}\frac{dn}{dx} = -\frac{D}{P}\frac{dP}{dx} \qquad (14.2\text{-}9)$$

It follows that the force $dP/ndx$ acting on each ion as a result of the partial pressure gradient is given by

$$\left|\frac{dP}{ndx}\right| = \frac{Pu'}{nD} \qquad (14.2\text{-}10)$$

If, on the other hand, there is no gradient in the concentration of ions, but an electric field $E$ exists in the region causing the ions to drift with a velocity $\bar{u}$ equal to $u'$, then the force $eE$ must equal $dP/ndx$, or

$$eE = \frac{e\bar{u}}{\mu} = \frac{eu'}{\mu} = \frac{Pu'}{nD} \qquad (14.2\text{-}11)$$

Hence[6]

$$\frac{D}{\mu} = \frac{P}{en} = \frac{kT}{e} \qquad (14.2\text{-}12)$$

where the relationship $P = nkT$ given by Equation (14.1-6) is assumed to apply to the ions diffusing through their parent gas. This equation gives the desired relation between the mobility and the diffusion constant.

If it is assumed that collisions between the ions and the gas molecules are elastic, Equations (14.2-7) and (14.2-12) can be combined to give an expression for $\mu$, namely that

$$\mu = \frac{e}{kT}\frac{3\pi\sqrt{2}}{32}L u_{\text{avg}}\sqrt{\frac{m+M}{M}} \qquad (14.2\text{-}13)$$

where $m$ is the mass of the ions and $M$ is the mass of the gas molecules.

### 14.3   Motion of Ions and Electrons through a Gas under the Influence of an Electric Field — Experimental Results

In this section we shall first present some results of measurements of the drift velocities of noble-gas ions through their parent gases when a uniform electric field is applied. Later we shall discuss the motion of free electrons through a noble gas under the influence of a uniform electric field.

Figure 14.3-1 shows some measured values[7] of the drift velocities of noble-gas ions through their parent gases as a function of $E/p$, the ratio of the electric field intensity to gas pressure. The quantity $E/p$ is an independent parameter in a number of gas discharge phenomena since at constant $E/p$ the average energy which an ion or electron gains in the interval between collisions is a constant. Furthermore, since the ion or electron loses kinetic energy with each collision, it must reach an equilibrium kinetic energy which depends only on the energy gained between collisions. It is customary in gas-discharge work to measure $E$ in volts/per centimeter and $p$ in millimeters of mercury rather than in mks units. This convention will be adhered to in the discussion that follows.

The log-log plot in Figure 14.3-1 shows a slope of one at low values of $E/p$, indicating that $\bar{u}$ is proportional to $E/p$ and hence to $E$ in agreement with Equation (14.2-8). On the other hand, at high values of $E/p$ the curves have a slope of $1/2$, indicating that $\bar{u}$ is proportional to $\sqrt{E/p}$. An analysis of the problem[8] which takes into account the nature of the ion-atom forces is able to predict this variation of $\bar{u}$ with $E/p$.

---

[6]When the ion velocity distribution is not Maxwellian, this relationship is not strictly true.

[7]References 14.3 and 14.4.

[8]Reference 14.5.
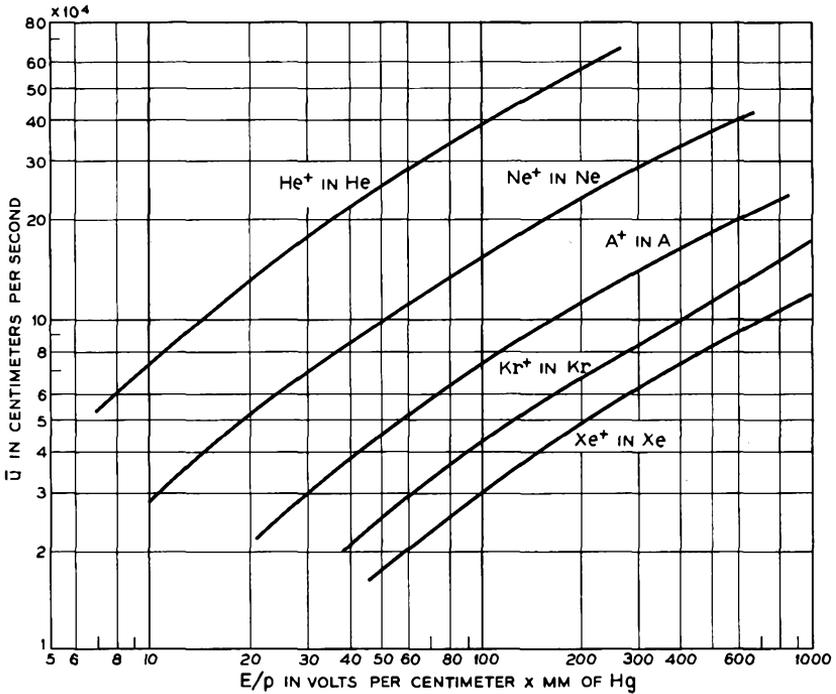
Fig. 14.3-1 Drift velocities of noble-gas ions through their parent gases as a function of $E/p$.

In Chapter 15 we shall find that the maximum value of $E/p$ which is obtained in the normal glow discharge with molybdenum electrodes and neon gas is of the order of 200 volts/cm × mm of Hg. At this value of $E/p$ the drift velocity of neon ions in neon is about 2250 meters/sec, or approximately 4.2 times the thermal velocity of neon molecules. Since the thermal energy of a neon molecule is approximately 0.025 electron volt and since the lowest excited state of neon lies at 16.6 electron volts, there is essentially no excitation or ionization of neutral atoms by ions in the discharge. The same result applies to discharges in the other noble gases.

When free electrons drift through a noble gas under the influence of an electric field, the picture is considerably different. The concept that the drift velocity for weak electric fields is determined by a mobility is less meaningful for electrons because the electron mean free path is found to be a function of the electron energy. (It will be recalled that the expression for the ion mobility given by Equation (14.2-13) is directly proportional to the mean free path and that the mobility was assumed to be independent of ion

energy.)  Figure 14.3-2 shows some experimentally measured[9] values of a quantity $P_c$ proportional to the reciprocal of the mean free path of free electrons in the noble gases as a function of the square root of the electron energy.  The abscissa is therefore proportional to the electron velocity.
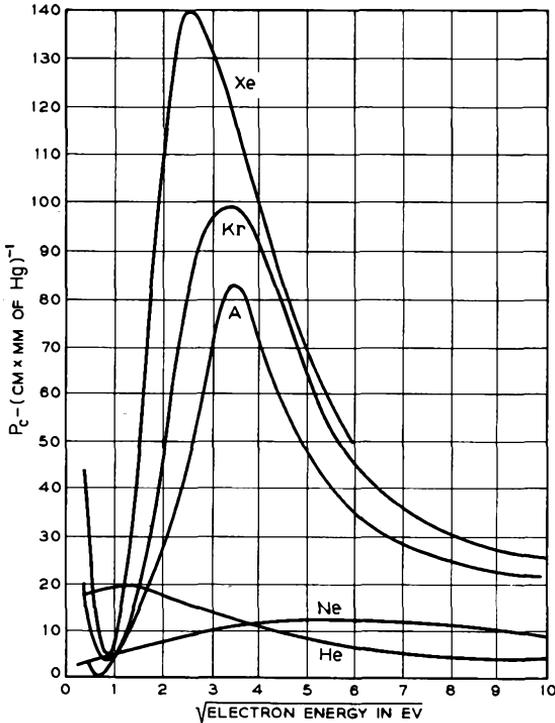


FIG. 14.3-2  The collision probability $P_c$ plotted as a function of the square root of the electron energy for electrons in the noble gases.  (From R. B. Brode, *Revs. Modern Phys.* **5,** 257, 1933)

The quantity $P_c$, known as the collision probability, equals $1/pL$, where $p$ is the gas pressure in millimeters of Hg and $L$ is the electron mean free path in centimeters.  In the case of the heavier gases, $P_c$ drops to relatively small values at electron energies of about 1 ev.  This effect, known as the Ramsauer-Townsend effect, can be explained quantum mechanically, and is discussed in several texts.[10]

---

[9]Reference 14.6.
[10]See, for instance, L. I. Schiff, *Quantum Mechanics*, McGraw-Hill Book Co., Inc., New York, 1949, p. 109.

For a given value of $E/p$, electrons gain much greater energies in drifting under the influence of the field than do ions. This effect can be attributed to the much smaller mass of the electrons. When an ion drifts through a gas under the influence of an electric field, the average energy transferred to the gas molecules per elastic collision is an appreciable part of the total kinetic energy of the ion. In a "head-on" elastic collision between an ion and a molecule in which the molecule is at rest and is of the same mass as the ion, the ion transfers all its kinetic energy to the struck particle. However, in a head-on elastic collision between an electron and a molecule, conservation of energy and momentum requires that the electron be reflected with almost all its kinetic energy. (The collision is in some ways similar to a collision between a ping pong ball and a bowling ball.) Consequently, even at relatively low values of $E/p$, of the order of 1 or 2 volts/cm $\times$ mm of Hg, free electrons in neon gas may scatter their way up to sufficient energies to cause excitation of the gas molecules, the energy being gained as the electrons advance in the direction of greater potential. At higher values of $E/p$ many of the electrons gain sufficient energy to cause ionization.

Table 14.3-1 lists some of the energies needed to excite and ionize the molecules of the noble gases. The metastable states referred to in the table are excited states for which the quantum-mechanical probability per unit time of direct transitions to the ground state by radiation is very small. An atom excited to such a state is frequently called a metastable.

TABLE 14.3-1. EXCITATION AND IONIZATION ENERGIES OF
NOBLE GAS MOLECULES IN ELECTRON VOLTS

|  | $W_m$ | $W_r$ | $W_i$ |
|---|---|---|---|
| He | 19.80 | 21.21 | 24.58 |
| Ne | 16.62 | 16.85 | 21.58 |
| A | 11.55 | 11.61 | 15.77 |
| Kr | 9.91 | 10.02 | 14.01 |
| Xe | 8.32 | 8.45 | 12.14 |

$W_m$ = excitation energy of lowest metastable state.
$W_r$ = excitation energy of lowest level from which direct radiation to the ground state can take place.
$W_i$ = first ionization potential, or energy needed to remove a single electron from the outer shell of a neutral molecule.

When electrons make inelastic collisions with molecules of the noble gases in which excitation or ionization takes place, the electrons lose a large part of their kinetic energy. If a collision results in excitation of a neutral molecule, the kinetic energy lost by the electron equals the excitation energy of the molecule plus the kinetic energy gained by the molecule. Even

more energy is lost if ionization of a neutral molecule takes place. Consequently, the "steady-state" condition in which electrons drift through a gas under the influence of a moderate or large electric field is a dynamic condition in which the electrons scatter their way up to energies sufficient to ·cause excitation or ionization, whereupon they abruptly lose most of their energy and start over again.

Each time an electron ionizes a neutral molecule, an additional electron is released which is also accelerated by the field and which may cause further ionization. An "avalanche" therefore results in which a single electron falling through a sufficiently large potential drop gives rise to many electrons.
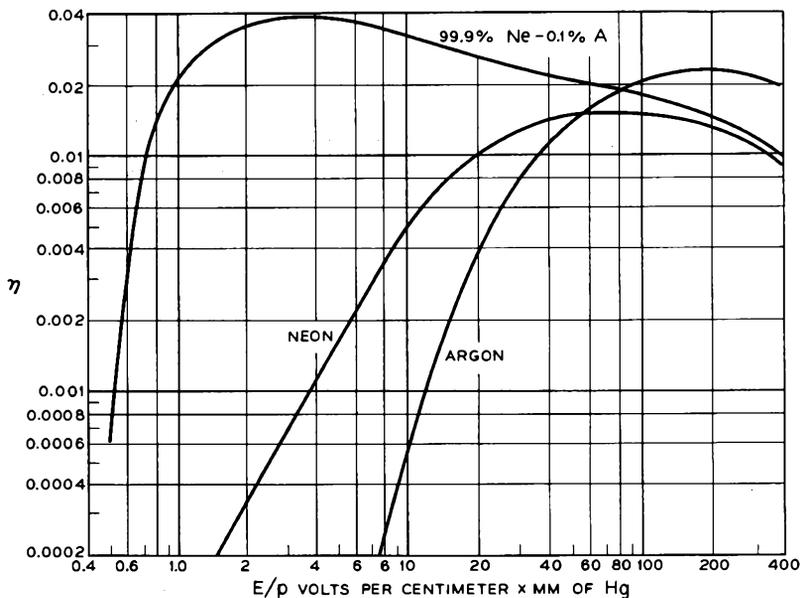


FIG. 14.3-3 The ionization coefficient $\eta$ as a function of $E/p$. (From A. A. Kruithof and F. M. Penning, *Physica* **4**, 430, 1937)

Figure 14.3-3 shows some measured[11] values of a quantity $\eta$ known as the ionization coefficient per volt as a function of $E/p$. $\eta$ is defined by the relation $dN = \eta N dV$, where $dN$ is the incremental number of ion-electron pairs produced by $N$ electrons in falling through $dV$ volts. This definition implies that an "average" avalanche of $\epsilon^{\eta V}$ electrons is generated by a single electron in falling through $V$ volts. However, the values of $\eta$ shown in Figure

———————————

[11]Reference 14.7.

14.3-3 apply only to electrons that have a velocity distribution characteristic of the "steady-state" condition in a region of uniform $E/p$. If the field is established between planar electrodes and if the electrons begin at the cathode ($V = 0$) with relatively small energy, the size of the avalanche is more accurately expressed by $\epsilon^{\eta(V-V_o)}$, where $V_o$ is a voltage which accounts for the potential drop through which the electrons must fall in order to acquire the energy distribution characteristic of the established field. Figure
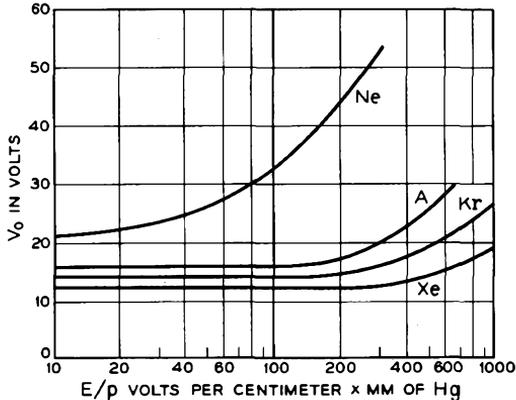


FIG. 14.3-4 The voltage $V_o$ as a function of $E/p$. (From M. J. Druyvesteyn and F. M. Penning, *Revs. Modern Phys.* **12**, 87, 1940)

14.3-4 shows some measured values[12] of $V_o$ as a function of $E/p$ for several noble gases.

Of particular interest in Figure 14.3-3 is the curve for the gas mixture 99.9 per cent Ne − 0.1 per cent A since it is well above the curves for pure neon and pure argon over a considerable range of $E/p$ values. The greater ionization in the Ne-A mixture results from a two-stage, modified $\eta$ process in which neon metastables ionize neutral argon atoms. Reference to Table 14.3-1 shows that the lowest metastable state of neon is higher than the ionization potential of argon. Furthermore, it is known that when a neon atom is excited, the probability that it goes into a metastable state is high[13], probably greater than 50 per cent for $E/p$ less than 12 volts/cm × mm of Hg. A current of fast electrons passing through the Ne-A mixture therefore generates a considerable number of neon metastables. The metastables diffuse through the gas with a relatively high probability of striking neutral argon atoms before being de-excited (see next section). Upon collision with

---

[12]Reference 14.8, p. 100.
[13]Reference 14.9, p. 463; Reference 14.8, p. 100.

an argon atom the excitation energy of the metastable is transferred to an outer-shell electron in the argon atom giving the electron sufficient energy to leave the atom. The neon metastable thereby becomes an unexcited atom, and an argon ion is generated. The phenomenon is known as the Penning effect.

For pure neon and argon the curves of $\eta$ fall off at low values of $E/p$ because of an increasing ratio of excitation to ionization and because of energy losses in elastic collisions. The curve of $\eta$ for the Ne-A mixture falls off at low $E/p$ because an increasing portion of the energy received by the electrons is lost in elastic collisions.

A measure of the voltage drop needed for a single electron to produce one more electron is obtained by setting $\epsilon^{\eta V} = 2$, or $V = (1/\eta) \ln 2$. For pure neon $\eta$ has a maximum value of 0.015 at which $(1/\eta) \ln 2 = 46$ volts. For pure argon $\eta$ has a maximum value of 0.025 at which $(1/\eta) \ln 2 = 28$ volts; and for the Ne-A mixture $\eta$ has a maximum value of 0.037 at which $(1/\eta) \ln 2 = 19$ volts. The latter value of $\eta$ is the maximum value so far measured for any gas or gas mixture.

It is sometimes useful to define an additional ionization coefficient $\alpha$ such that $\alpha = \eta E$. $\alpha$ is known as the ionization coefficient per centimeter. In analogy to the definition of $\eta$ given earlier, $\alpha$ satisfies the relation $dN = \alpha N dx$, where $dN$ is the number of ion-electron pairs produced by $N$ electrons in traveling $dx$ centimeters through the gas. In a discharge between planar electrodes in which there is a uniform electric field intensity between the electrodes, the electron current can be expressed as $I = I_o \epsilon^{\alpha(x-x_o)} = $
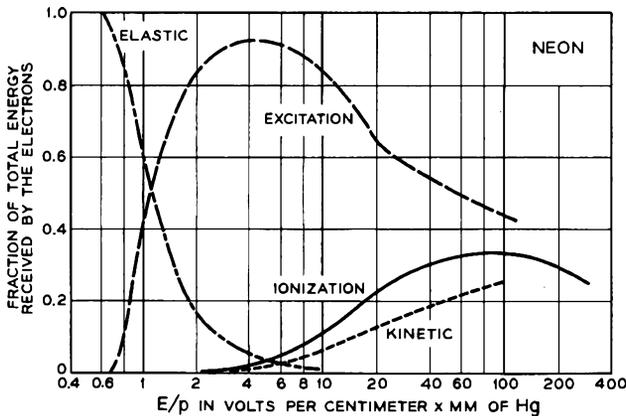


FIG. 14.3-5 The figure shows how the energy received by free electrons drifting through neon gas under the influence of a uniform electric field is spent as a function of $E/p$. From F. M. Penning, *Physica* 5, 286, 1938)

$I_o \epsilon^{\eta(V-V_o)}$, where $x_o$ is the distance from the cathode at which the potential is $V_o$, and $I_o$ is the electron current emitted from the cathode. Since $\alpha = \eta E$, and $\eta$ is a function of $E/p$, the ratio $\alpha/p$ is also a function of $E/p$.

Figure 14.3-5 shows how the total energy received by free electrons drifting through neon under the influence of an electric field is spent[14] as a function of $E/p$. Part of the energy received by the electrons is expended in elastic collisions (Elast.), part in inelastic collisions leading to excitation and ionization (Excit., Ion.), and part in accelerating newly released electrons up to the average energy (Kin.).

## 14.4  Metastable Generation, Diffusion, and Destruction

Metastables play an important part in a number of gas-discharge phenomena. Metastables are generated by high-energy electrons striking
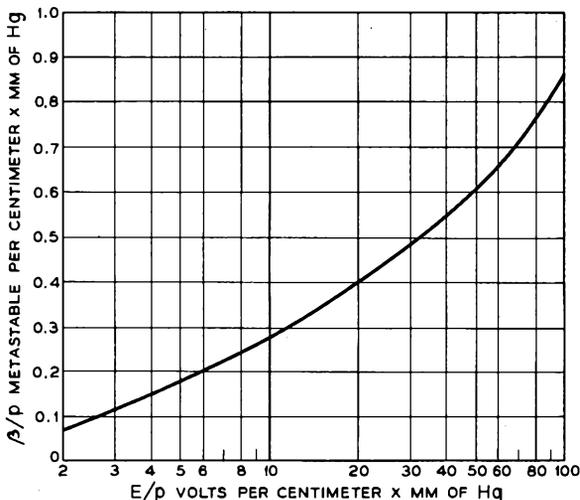
FIG. 14.4-1 $\beta$ measures the average number of metastables produced by an electron in advancing one centimeter in the direction of the field. The data are for free electrons in neon.

neutral molecules and by radiative transitions from other excited states. Since the metastables are uncharged, they diffuse through the gas and behave much like unexcited molecules (except that their effective diameter for collision is somewhat larger than that of the unexcited molecules).

---

[14]Reference 14.10, p. 288; Reference 14.8, p. 103.

Eventually each metastable experiences a collision which causes it to lose its identity as a metastable. In this section we shall present some data relating to metastable generation and diffusion, and the results of calculations based on the data. We shall also discuss briefly the mechanisms by which metastables are destroyed.

In analogy to the ionization coefficient $\alpha$ we can define a coefficient $\beta$ which expresses the average number of metastables produced by a single electron in advancing one centimeter in the direction of the field. Just as $\alpha/p$ is a function of $E/p$, so also is $\beta/p$. Rogowski[15] has estimated $\beta/\alpha$ for neon as a function of $E/p$ by reducing Kruithof and Penning's[16] data for the ionization coefficient in neon-argon mixtures. His results are given in Table 14.4-1 together with Kruithof and Penning's values of $\eta$. The right-hand column in the table gives $\beta/p = (\beta/\alpha)\eta E/p$. Figure 14.4-1 shows a plot of $\beta/p$ vs. $E/p$.

TABLE 14.4-1. IONIZATION AND METASTABLE GENERATION
COEFFICIENTS FOR NEON

| $E/p$ Volts/ cm $\times$ mm of Hg | $\beta/\alpha$ Metastables/ Ion-Electron Pairs (Ref. 14.11) | $\eta$ Ion-Electron Pairs/ Volt (Ref. 14.7) | $\beta/p$ Metastables/ cm $\times$ mm of Hg |
|---|---|---|---|
| 2 | 102 | 0.00034 | 0.0694 |
| 5 | 22 | 0.0016 | 0.176 |
| 10 | 5.3 | 0.0050 | 0.265 |
| 20 | 2.1 | 0.0096 | 0.404 |
| 30 | 1.3 | 0.012 | 0.468 |
| 50 | 0.85 | 0.0142 | 0.604 |
| 100 | 0.6 | 0.0145 | 0.87 |

Metastables generated in a discharge of *low current density* are lost principally by (a) diffusion to the electrodes or walls of the containing vessel where de-excitation takes place, (b) collisions with neutral atoms leading to excitation or de-excitation to a nearby radiating state, from which a transition to the ground state can take place, and (c) collisions with two neutral atoms leading to the formation of an excited molecule which may be metastable.[17] In discharges of *higher current density* additional interactions be-

---

[15]Reference 14.11.
[16]Reference 14.7.
[17]Reference 14.12. In collisions leading to the formation of a metastable molecule, one of the neutral atoms becomes bound to the metastable, thus forming the molecule, and the other neutral atom carries away the binding energy of the molecule. The fact that the metastable molecule is stable, and the unexcited diatomic molecule is not, can be attributed to different electron configurations in the two cases.

tween metastables and metastables, between metastables and electrons, and between metastables and ions contribute to the destruction of the meta-stables. In the metastable-metastable interactions, one metastable becomes ionized and the other becomes de-excited.

In a discharge of low current density in which only the first three types of interaction are important, the equation expressing the density of meta-stables within the discharge is

$$\frac{\partial M}{\partial t} = D \frac{\partial^2 M}{\partial x^2} - ApM - Bp^2M + J(x)\beta(x)/e \qquad (14.4\text{-}1)$$

where $M$ is the number of metastables per unit volume, $D$ is the meta-stable diffusion coefficient, $p$ is the gas pressure in millimeter of Hg, $A$ is the frequency of metastable destruction by two-body collisions at a pressure of 1 mm of Hg, $B$ is the frequency of metastable destruction by three-body collisions at a pressure of 1 mm of Hg (both $Ap$ and $Bp^2$ have the dimensions $(\text{Time})^{-1}$), $J(x)$ is the electron current density, and $e$ is the electronic charge. It is assumed that the discharge takes place between planar parallel electrodes. The left-hand term in the equation and the first term on the right comprise the diffusion equation, Equation (14.2-6). The terms $ApM$ and $Bp^2M$ express the volume rate of destruction of metastables, and the term $J(x)\ \beta(x)/e$ expresses the volume rate of gener-ation of metastables. In the steady-state condition both sides of the equa-tion are zero.

Phelps and Molnar[18] have measured the coefficients $D$, $A$, and $B$ for neon metastables in neon. They found that $D = 146/p$ cm$^2$/sec, where $p$ is the pressure in millimeters of Hg, $A = 50$ sec$^{-1}$ (mm of Hg)$^{-1}$, and $B = 0.5$ sec$^{-1}$ (mm of Hg)$^{-2}$.

McClure[19] has used Equation (14.4-1) and the values of $D$, $A$, and $B$ given above to determine the steady-state current of metastable atoms reaching the cathodes of several devices having planar electrodes, a neon gas filling, and particular applied anode voltages. The current of meta-stables reaching the cathode was determined by calculating $D\partial M/\partial x$ at $x = 0$ using the boundary conditions that $M = 0$ at $x = 0$ and $x = d$, where $d$ is the electrode spacing.

The results of McClure's calculations are shown in Table 14.4-2. The applied anode voltages $V_a$ given in the table are the experimental break-down voltages for devices with planar molybdenum electrodes and neon-gas fillings. The quantity $G$ is the average number of metastables striking the cathode for each electron leaving the cathode. Later when we come to

---

[18]References 14.12 and 14.13.
[19]Reference 14.14.

TABLE 14.4-2. DATA RELATING TO AVALANCHES IN NEON GAS BETWEEN PLANAR ELECTRODES*

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $pd$ mm of Hg $\times$ cm | 1.25 | 1.50 | 1.75 | 2.0 | 2.5 | 3 | 4 | 5 | 7 | 9 | 11 |
| $V_a$ volts | 148.5 | 146.0 | 145.0 | 145.1 | 147.6 | 151.9 | 162 | 173 | 196 | 220 | 239 |
| $E/p$ volts/cm $\times$ mm of Hg | 119 | 97 | 83 | 73 | 59 | 51 | 40 | 35 | 28 | 24 | 22 |
| $\epsilon^{\eta(V_a-V_o)} = \epsilon^{\alpha(d-z_o)}$ | 5.3 | 5.4 | 5.5 | 5.6 | 5.8 | 6.1 | 6.5 | 7.0 | 7.7 | 8.6 | 8.9 |
| $G$ metastables/electron | 0.62 | 0.67 | 0.71 | 0.75 | 0.79 | 0.81 | 0.80 | 0.74 | 0.62 | 0.52 | 0.43 |
| $H$ metastables/electron | 2.5 | 2.7 | 2.9 | 3.1 | 3.7 | 4.3 | 5.6 | 6.9 | 9.4 | 12 | 15 |
| $K$ ions/electron | 4.3 | 4.4 | 4.5 | 4.6 | 4.8 | 5.1 | 5.5 | 6.0 | 6.7 | 7.6 | 7.9 |

*Pressure $p = 40$ mm of Hg. The voltages $V_a$ are the breakdown voltages for devices with planar molybdenum electrodes and neon-gas fillings.

discuss the breakdown phenomenon in Section 14.6 we shall make use of these results.

A second calculation of interest is the average number of metastables produced by the release of a single electron from the cathode. Assuming that few metastables are generated closer to the cathode than a distance $x_o = (V_o/V_a)d$, where $V_o$ is the voltage plotted in Figure 14.3-4, this average number of metastables is easily shown to be

$$H = \beta \int_{x_o}^{d} \epsilon^{\alpha(x-x_o)}dx = \frac{\beta}{\alpha}(\epsilon^{\alpha(d-x_o)} - 1) \qquad (14.4\text{-}2)$$

The ratio $\beta/\alpha$ can be obtained by plotting a smooth curve of the data given in Table 14.4-1. Table 14.4-2 lists the values of $H$ calculated in this manner for the particular applied voltages and $pd$ values selected in the table. The quantity $K$ given in the table is the average number of ions generated by each electron leaving the cathode. $K$ is equal to $\epsilon^{\alpha(d-x_o)} - 1$.

## 14.5    Emission of Electrons from Metal Surfaces by Ion, Metastable, and Photon Impact

The cold-cathode glow discharge is self-sustaining because products of the electron avalanches — ions, metastables, and photons — strike the cathode and release additional electrons. The self-sustaining condition is attained when the average number of discharge products generated by the release of a single electron from the cathode is sufficient to release one more electron. In normal glow discharges with noble gases and clean molybdenum or tungsten electrodes, about two thirds of the cathode emission results from ions striking the cathode, about one third results from incident metastables, and photoelectric emission may account for a few per cent.[20]

The emission of electrons from clean metal surfaces under impact by ions of the noble gases has been investigated in detail by Hagstrum.[21] The results of his investigations of the electrons emitted from molybdenum and tungsten by singly charged ions of the noble gases are shown in Figure 14.5-1. The quantity $\gamma$ plotted in the figure is the average number of electrons emitted per incident ion. The measurements were taken using an exceptionally well evacuated system. Each observation was made immediately after the metal surface had been heated to a high temperature to drive off adsorbed layers of gas and before appreciable readsorption had taken place.

---

[20]Reference 14.15.
[21]References 14.16 to 14.19.
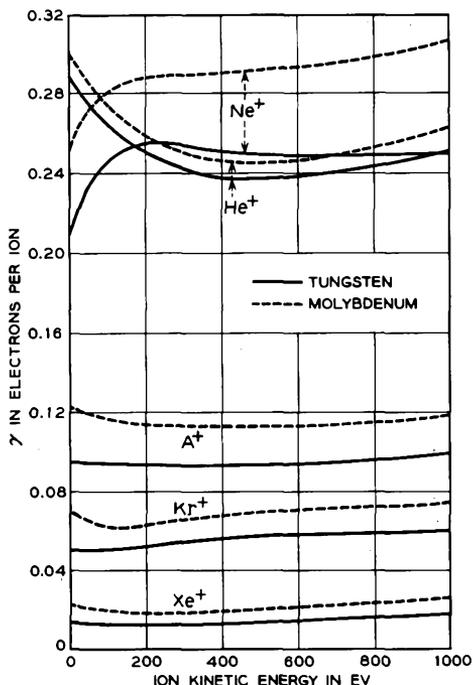
FIG. 14.5-1 The quantity $\gamma$ for ions of the noble gases incident upon clean tungsten and molybdenum surfaces. (From H. D. Hagstrum, *Phys. Rev.* **104**, 672, 1956)

From his measurements Hagstrum concluded that for helium ions of kinetic energy less than 400 electron volts and for Ne, A, Kr, and Xe ions of kinetic energy less than 1000 electron volts the ion kinetic energy plays only a secondary role in the electron emission and is not directly responsible for the kinetic energy imparted to the emitted electrons. Instead, the emitted electrons derive their kinetic energy from the potential energy of the ions. The mechanism is as illustrated in Figure 14.5-2. As the ion approaches the metal surface, an interaction between the field of the ion and the fields of two electrons $e_1$ and $e_2$ in the conduction band of the metal causes one electron $e_1$ to be captured by the ion, and the second electron $e_2$ to be excited to an energy level well above the conduction band. The electron $e_1$ goes into the lowest unfilled state of the ion making it an unexcited atom. This releases an amount of energy $\epsilon'$ which is imparted to the second electron. Hagstrum concluded that the probability of this interaction taking place when noble-gas ions of 10 electron volts of kinetic energy or less are incident upon the metal surface is very nearly unity.
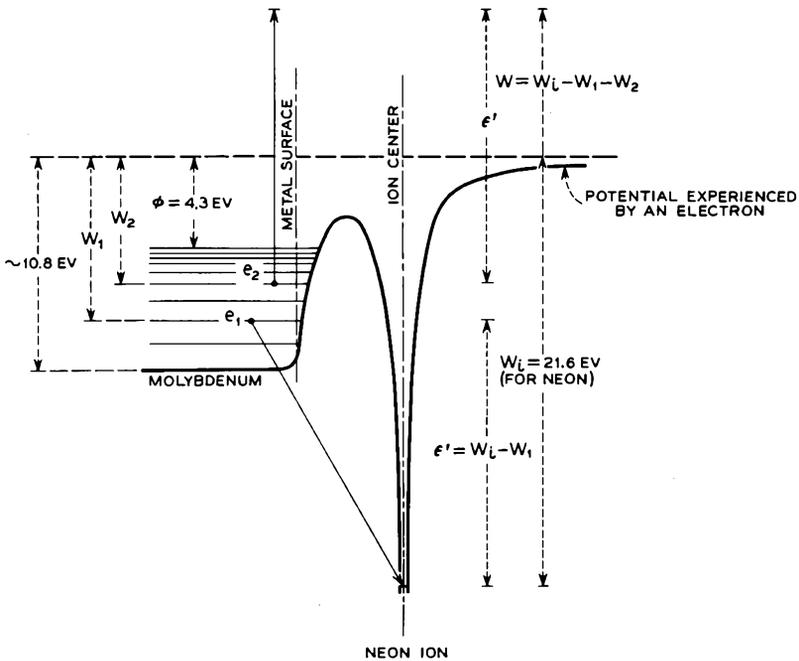
FIG. 14.5-2 The interaction that takes place at a molybdenum surface between an incident neon ion and two electrons in the conduction band of the metal.

The quantity $\gamma$ therefore measures the probability that the electron $e_2$ will escape from the metal when excited to the higher state.

Since the electrons $e_1$ and $e_2$ may initially have any energy within the conduction band, the excitation energy of the $e_2$ electron just after the interaction may have a range of values. From Figure 14.5-2 it is easily shown that the maximum possible energy of the emitted electrons is $W_i - 2\phi$, where $W_i$ is the ionization potential of the ion in electron volts, and $\phi$ is the work function of the metal surface in electron volts. The emitted electron $e_2$ has maximum energy only if both electrons $e_1$ and $e_2$ were originally in energy states at the top of the conduction band and if the electron $e_2$ suffers no loss of energy before it leaves the surface. It is reasonable to expect that for different ions incident upon the same surface, high $\gamma$ would be associated with a high value of $W_i - 2\phi$, and vice versa. From Figure 14.5-1 it is evident that, for low incident ion energies, $\gamma$ decreases with increasing ion mass in going from He to Xe. This can be attributed to a corresponding decrease in $W_i$ with increasing mass of the ions.

In investigating the emission of electrons from molybdenum by $He^+$

ions Hagstrum found that the electron emission at low ion energies ($\sim$10 electron volts) was reduced by about a third when the surface was covered with a monolayer of helium gas. Varney[22] has measured $\gamma$ for noble-gas ions striking a molybdenum surface through observations taken with a pulsed discharge. It is likely that the molybdenum cathode used in his measurements was covered with approximately a monolayer of noble gas, and hence more closely approximated conditions which are normally experienced in a gas-discharge device. Varney's measured values of $\gamma$ are shown in Figure 14.5-3 as a function of $E/p$. The curves show nearly
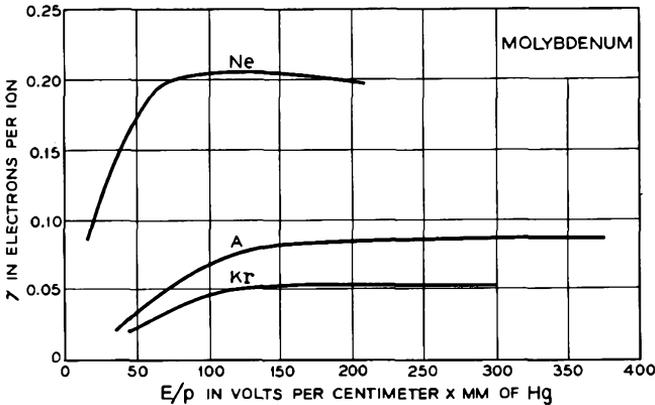


Fig. 14.5-3 Measurements of $\gamma$ taken with a pulsed discharge. (From R. N. Varney, *Phys. Rev.* **93**, 1156, 1954)

constant values of $\gamma$ for $E/p$ greater than 150 volts/cm $\times$ mm of Hg. These maximum values of $\gamma$ are less than those observed by Hagstrum apparently because of the effects of adsorbed gas on the metal surface. For low values of $E/p$ the curves become almost straight lines approximately directed at the origin. The reduction in $\gamma$ at lower $E/p$ is attributable to diffusion of the emitted electrons back to the surface as a result of collisions with gas molecules. As $E/p$ increases, the probability of diffusion back to the cathode decreases; and for $E/p$ greater than 150 volts/cm $\times$ mm of Hg, virtually none of the emitted electrons return to the cathode.

Gases with polyatomic molecules, such as hydrogen, nitrogen, oxygen, and carbon dioxide, tend to become adsorbed on metal surfaces much more than the noble gases. When a tube is filled with a gas such as hydrogen, the cathode becomes covered with several monolayers, whereas a noble

---

[22]Reference 14.20.

gas may only cover the cathode with a single monolayer. As a result, $\gamma$ for a discharge in hydrogen is very small, perhaps a few times $10^{-4}$ at low $E/p$, where the ions have little energy with which to penetrate the adsorbed layers. At higher $E/p$, $\gamma$ takes on higher values, perhaps $10^{-2}$ to $10^{-1}$, as the ions gain sufficient energy to approach the surface more closely.

Hagstrum concluded that, when a metastable atom of a noble gas is incident upon a clean molybdenum or tungsten surface, it experiences a two-stage interaction in which it first becomes ionized by losing its excited electron to the conduction band of the metal. The excited electron tunnels through[23] the potential barrier between the atom and the metal and occupies a vacant state in the metal directly opposite the metastable level. The incident particle is then indistinguishable from an incident ion so that an interaction similar to that illustrated in Figure 14.5-2 follows. It is therefore thought that the values of $\gamma$ for metastable atoms of the noble gases incident upon clean molybdenum or tungsten surfaces are the same as for the corresponding ions.

The contribution made by photoelectric emission to the total cathode emission depends markedly on the work function of the cathode surface. Photoelectrons are emitted from the cathode because photons generated in the discharge strike the cathode and interact with electrons in the conduction band. The interaction is such that the energy of an individual photon is imparted to a single electron. The electron is energetically able to escape from the metal if it is excited to a state at least $\phi$ volts above the conduction band, where $\phi$ is the work function of the cathode surface. For a given energy $h\nu$ of the incident photon, the maximum energy of the emitted photoelectrons is $h\nu - \phi$, where $h$ is Planck's constant, and $\nu$ is the frequency of the radiation. Low work function cathodes are in general more efficient emitters of photoelectrons, so that the photoelectron emission from an oxide-coated cathode (work function 1 to 2 electron volts) is likely to be much greater than that from a metal surface such as molybdenum (work function 4.3 electron volts). Because $\gamma$ is quite small for discharges in gases with polyatomic molecules, the proportion of the total cathode emission arising from the photoelectric process is thought to be greater in these discharges than in discharges with the noble gases.

## 14.6   The Townsend Discharge and Breakdown

A gas discharge can be considered to be a state of conduction through a partially ionized gas between two electrodes. The current density of the

---

[23]A quantum-mechanical effect in which an electron can pass through a potential barrier, even though classically the electron should not do so because it does not have sufficient kinetic energy to overcome the barrier.

discharge may vary over many orders of magnitude, from perhaps a few electrons per second per square centimeter ($\sim 10^{-18}$ amp/cm²) to thousands or millions of amperes per square centimeter in the arc discharge. It happens that as the discharge current is increased above about 10 micro-amps, two independent phenomena start to show their influence on the discharge. Events which depend upon the square of the current density, such as metastable-metastable interactions and electron-metastable inter-actions, become of importance, and the space charge density in the inter-electrode space starts to modify the potential distribution between the electrodes. A Townsend discharge[24] is characterized by a current density which is sufficiently small that neither of these effects are of importance.

In the present section we shall use the experimental data presented earlier to discuss the Townsend discharge and the phenomenon known as breakdown. Two discharges occurring at higher current densities, the glow discharge and the arc discharge, will be described in the first section of Chapter 15.
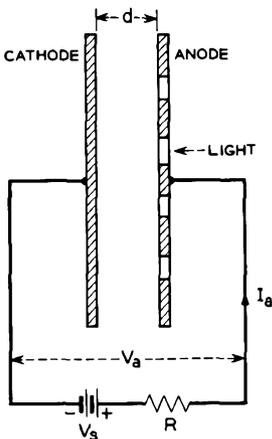


FIG. 14.6-1 A discharge device consisting of two large planar electrodes.

Let us suppose we have a device consisting of two large planar elec-trodes, a cathode and an anode, spaced by a distance of $d$ centi-meters, as illustrated in Figure 14.6-1. The interelectrode space is filled with a noble gas at a pressure of $p$ millimeters of Hg. The applied voltage $V_a$ is less than the break-down voltage $V_b$ of the device. The anode has a number of holes in it, and light is shone through the holes at the cathode. We shall assume that the incident light causes a cur-rent of $I_o$ photoelectrons to be drawn from the cathode. Each photo-electron gives rise to an avalanche which in an average case consists of $(\epsilon^{\eta(V_a - V_o)} - 1)$ ion-electron pairs plus the original electron. The $(\epsilon^{\eta(V_a - V_o)} - 1)$ ions generated in the ava-lanche are drawn to the cathode where they release $\gamma(\epsilon^{\eta(V_a - V_o)} - 1)$ addi-tional electrons. Metastables generated in the avalanches also strike the cathode and release electrons. If $I_c$ is the total cathode emission, and $G$ is

---

[24]After J. S. Townsend who contributed much to the early understanding of gas-discharge phenomena.

the average number of metastables striking the cathode as a result of the emission of a single electron from the cathode, then

$$I_c = I_o + \gamma(\epsilon^{\eta(V_a - V_o)} - 1 + G)I_c \qquad (14.6\text{-}1)$$

or

$$I_c = \frac{I_o}{1 - \gamma(\epsilon^{\eta(V_a - V_o)} - 1 + G)} \qquad (14.6\text{-}2)$$

where it is assumed that the contribution to $I_c$ made by photons generated in the discharge is negligible. The current of electrons reaching the anode is $I_c \epsilon^{\eta(V_a - V_o)}$, or

$$I_a = \frac{I_o \epsilon^{\eta(V_a - V_o)}}{1 - \gamma(\epsilon^{\eta(V_a - V_o)} - 1 + G)} \qquad (14.6\text{-}3)$$

This equation expresses the anode current of a Townsend discharge, when the discharge takes place between planar electrodes with a voltage $V_a$ applied between the electrodes, and when a photoelectron current $I_o$ is drawn from the cathode. The equation is known as the Townsend Equation.

As the applied voltage $V_a$ increases, the denominator of the Townsend Equation decreases and the numerator increases.[25] At a particular applied voltage $V_b$, the denominator becomes zero and the current $I_a$ may increase many orders of magnitude. The condition is known as breakdown, and the voltage $V_b$ is called the breakdown voltage of the device.

The condition that the denominator of the Townsend Equation be zero is essentially the condition for a self-sustaining discharge, namely that the average number of discharge products generated by the release of a single electron from the cathode be sufficient to release one more electron from the cathode. In such a discharge the current $I_a$ is no longer proportional to the photocurrent $I_o$, and in fact would continue if the light source were turned off. To limit the current of a self-sustained discharge, a resistance $R$ is placed in series with the supply voltage. The voltage $V_a$ is then $V_s - I_a R$, where $V_s$ is the supply voltage.

Figure 14.6-2(c) shows qualitatively the variation of $I_a$ with $V_a$ for a particular device. The broken curve in the figure shows the $I_a$-$V_a$ characteristic for an increased photocurrent $I_o'$. The horizontal part of the solid curve in the Townsend region and the portions of the solid curve to the right of the Townsend region correspond to self-sustained discharges. Evidently, the current of the self-sustained discharge in the Townsend

---

[25]It should be noted that at low applied voltages, and hence at low $E/p$, many of the photoelectrons diffuse back to the cathode, so that increasing the applied voltage increases the current $I_o$ of photoelectrons drawn from the cathode.

FIG. 14.6-2 Current voltage characteristics of a cold-cathode diode. (a) Current is plotted on a linear scale, and (c) current is plotted on a logarithmic scale.

region can be varied over several orders of magnitude with essentially no change in $V_a$.

To understand the stability of the self-sustained discharge we should note that the anode voltage of a self-sustained discharge is that voltage which makes the average avalanche contain enough discharge products to release one more electron from the cathode. Suppose that a discharge current $I_a$ and an anode voltage $V_a$ are obtained when the supply voltage is $V_s$ and the series resistance is $R$. If the current $I_a$ were to increase a small amount to $I_a + \Delta I_a$, the anode voltage $V_a = V_s - I_a R$ would fall $\Delta I_a R$ volts below the voltage needed for the discharge to be self-sustaining. The avalanches would then contain less than enough discharge products to release one more electron from the cathode, and the current $I_a + \Delta I_a$ would decrease. The current $I_a$ therefore adjusts itself to the value which makes $V_a$ the voltage needed to sustain the discharge.

The condition for breakdown can be expressed as

$$\gamma(\epsilon^{\eta(V_b - V_o)} - 1 + G) = 1 \qquad (14.6\text{-}4)$$

or

$$V_b = V_o + \frac{1}{\eta}\ln\left(\frac{1}{\gamma} + 1 - G\right) \qquad (14.6\text{-}5)$$

where $V_b$ is the breakdown voltage. Figure 14.6-3 shows a plot of $V_b$ as a function of the product of the gas pressure $p$ and the electrode spacing
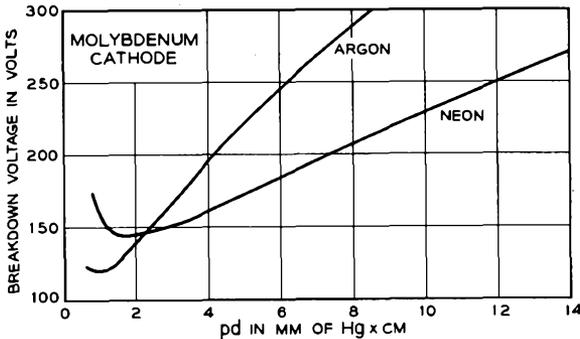


FIG. 14.6-3 The breakdown voltage for devices with planar molybdenum electrodes and neon and argon fillings.

$d$ for devices with planar molybdenum electrodes and neon and argon fillings. To a first approximation $V_b$ is a function of the product $pd$, since at constant $pd$ and constant applied voltage, both the average number of collisions made by an electron in traveling from the cathode to the anode and the average voltage gained in the interval between collisions are constant. The relationship that

$$V_b = f(pd) \qquad (14.6\text{-}6)$$

is known as Paschen's Law.

It will be of interest to evaluate the quantity $\gamma(\epsilon^{\eta(V_a - V_o)} - 1 + G)$ for one of the $pd$ values shown in Table 14.4-2. The voltages $V_a$ given in the table are the measured breakdown voltages for devices with planar molybdenum electrodes and neon fillings at a pressure of 40 mm of Hg. Consequently if we use the value of $V_a$ indicated in the table and the value of $\gamma$ for neon ions incident on molybdenum, the quantity $\gamma(\epsilon^{\eta(V_a - V_o)} - 1 + G)$ should equal 1.0. The device we shall consider has a $pd$ of 5 and hence an electrode spacing of 0.125 cm. The measured breakdown voltage is 173 volts. As the applied voltage is increased toward breakdown, $E/p$ approaches

34.6 volts/cm × mm of Hg. From Figure 14.5-3 we find that $\gamma = 0.14$ for this value of $E/p$. From Table 14.4-2 the quantity $\epsilon^{\eta(V_a - V_o)}$ is found to be 7.0 and $G = 0.74$. Hence $\gamma(\epsilon^{\eta(V_a - V_o)} - 1 + G) = 0.94$. The difference between 0.94 and 1.0 can easily be explained in terms of the accuracy of the experimental data we have used.

Since each value of $pd$ defines a unique $V_b$, it also defines a unique $E/p$, namely $V_b/pd$. The breakdown voltage $V_b$ can therefore also be plotted as a function of the value of $E/p$ which applies at breakdown. Such a plot could be made from the data given in Table 14.4-2. Figure 14.6-3 shows that minimum breakdown voltage for a tube with a neon gas filling and planar molybdenum electrodes occurs at a pd of about 1.8 mm of Hg × cm. At this $pd$, $V_b = 145$ volts, and $E/p = 145/1.8 = 80$ volts/cm × mm of Hg. Reference to Figures 14.3-3 and 14.5-3 shows that both $\eta$ and $\gamma$ are at their maximum values at approximately this value of $E/p$. Table 14.4-2 shows that $G$ is a relatively slowly varying function of the value of $E/p$ at breakdown. Consequently, $V_b = V_o + (1/\eta) \ln [(1/\gamma) + 1 - G]$ is a minimum at $E/p = 80$ volts/cm × mm of Hg and hence at $pd = 1.8$ mm of Hg × cm.

When it is important that breakdown voltage of a cold-cathode tube be reproducible to within narrow limits, the pressure and electrode spacing are often chosen so that $V_b$ is at its minimum value. In this case small variations in $p$ and $d$ from tube to tube make little change in the breakdown voltage.

Equation (14.6-6) is valid for electrode geometries other than the planar geometry if all electrode dimensions are scaled in proportion to the electrode spacing $d$.

## REFERENCES

Several texts on gas tubes and gas-discharge phenomena are available. Among these are:

14a. S. Flügge, *Encyclopedia of Physics*, Vols. 21 and 22, Julius Springer, Berlin, 1956.

14b. A. Von Engel, *Ionized Gases*, Oxford University Press, Oxford, England, 1955.

14c. L. B. Loeb, *Basic Processes of Gaseous Electronics*, University of California Press, Berkeley, 1955.

14d. J. M. Meek and J. D. Craggs, *Electrical Breakdown of Gases*, Oxford University Press, Oxford, England, 1953.

14e. H. W. S. Massey and E. H. S. Burhop, *Electronic and Ionic Impact Phenomena*, Oxford University Press, Oxford, England, 1952.

14f. W. G. Dow, *Fundamentals of Engineering Electronics* 2nd Ed. Chapters 14–18, John Wiley and Sons, Inc., New York, 1952.

14g. J. Millman and S. Seely, *Electronics*, Chapters 8–11, McGraw-Hill Book Co., Inc., New York, 1951.

14h. J. D. Cobine, *Gaseous Conductors*, McGraw-Hill Book Co., Inc., New York, 1951.

14i. F. A. Maxfield and R. R. Benedict, *Theory of Gaseous Conduction and Electronics*, McGraw-Hill Book Co., Inc., New York, 1941.

14j. L. B. Loeb, *Fundamental Processes of Electrical Discharges in Gases*, John Wiley and Sons, Inc., New York, 1939.

Other references covering specific subjects discussed in this chapter are:

14.1. E. H. Kennard, *Kinetic Theory of Gases*, McGraw-Hill Book Co., Inc., New York, 1938.

14.2. J. C. Slater, *Introduction to Chemical Physics*, McGraw-Hill Book Co., Inc., New York, 1939.

14.3. J. A. Hornbeck, *Phys. Rev.* **84,** 615, 1951.

14.4. R. N. Varney, *Phys. Rev.* **88,** 362, 1952.

14.5. G. H. Wannier, *Bell System Tech. J.* **32,** 170, 1953.

14.6. R. B. Brode, *Revs. Modern Phys.* 5, 257, 1933.

14.7. A. A. Kruithof and F. M. Penning, *Physica* 4, 430, 1937.

14.8. M. J. Druyvesteyn and F. M. Penning, *Revs. Modern Phys.* **12,** 87, 1940.

14.9. A. A. Kruithof and M. J. Druyvesteyn, *Physica* 4, 450, 1937.

14.10 F. M. Penning, *Physica* 5, 286, 1938.

14.11. W. Rogowski, *Zeits F. Phys.* **115,** 257, 1940.

14.12. A. V. Phelps and J. P. Molnar, *Phys. Rev.* 89, 1202, 1953.

14.13. A. V. Phelps, *Phys. Rev.* **99,** 1657, 1955.

14.14. B. T. McClure, Private communication.

14.15. J. P. Molnar, *Elec. Eng.* **69,** 1071, 1950.

14.16. H. D. Hagstrum, *Phys. Rev.* **89,** 244, 1953.

14.17. H. D. Hagstrum, *Phys. Rev.* **91,** 543, 1953.

14.18. H. D. Hagstrum, *Phys. Rev.* **96,** 325 and 336, 1954.

14.19. H. D. Hagstrum, *Phys. Rev.* **104,** 309, 317, 672, and 1516, 1956.

14.20. R. N. Varney, *Phys. Rev.* **93,** 1156, 1954.

# Chapter 15

## THE GLOW DISCHARGE AND
## COLD-CATHODE TUBES

If the current drawn by a neon-molybdenum device with planar, parallel electrodes is increased beyond about 10 microamps, events which occur with a frequency proportional to the square of the current density such as metastable-metastable interactions and electron-metastable interactions begin to show their influence on the discharge. These interactions lead to the generation of a greater number of ions for each electron leaving the cathode.[1] Since almost every ion is drawn to the cathode, but less than half the metastables reach the cathode, the voltage needed to sustain the discharge is reduced below that of the Townsend discharge, and hence below the breakdown voltage.

At about the same current, or perhaps at somewhat lower current, the space charge caused by relatively slow-moving ions drifting toward the cathode becomes sufficient to raise the potential close to the cathode. This causes an increase in $E/p$ at the cathode surface. If $E/p$ at breakdown is less than that which gives maximum emission coefficient $\gamma$, the emission coefficient increases with the increase in $E/p$, and the voltage needed to sustain the discharge falls still further.

At a current of about 20 microamps, a region of diffuse glow can be seen in the interelectrode space, and the glow is observed to fill only a portion of that space. As the current is increased from 20 microamps to 200 microamps, the boundaries of the glow region become more clearly defined, and, for relatively large $pd$, the glow is observed to be closer to the cathode than the anode. At currents above 200 microamps, the glow area tends to increase linearly with current up to the point where the cathode is covered with glow.

---

[1] When the two metastables interact, one becomes ionized, and the other returns to the ground state. Collisions between fast electrons and metastables frequently result in ionization of the metastables.

The discharge in this range of currents is said to be a normal glow discharge and is characterized by a nearly constant sustaining voltage. If the current is increased beyond the point where the cathode is covered with glow, the sustaining voltage rises, and the discharge is said to be an abnormal glow discharge. At still higher currents it becomes an arc, and the anode-to-cathode voltage falls to a value comparable with the ionization potential of the gas.

The sustaining voltage of the normal glow discharge depends on such quantities as the emission coefficient $\gamma$, the product $pd$, and the functions which give the probabilities of excitation and ionization for electron collisions with the molecules as a function of the incident electron energy. Because it is possible to process tubes so that the cathode remains relatively clean over long periods, and hence $\gamma$ remains nearly constant, and because the other factors determining the sustaining voltage are essentially constant with time, it is possible to obtain sustaining voltages which change only a fraction of a per cent in a year of continuous service. Voltage-reference tubes used in regulated power supplies are so processed.

In voltage-regulator tubes, use is made of the fact that the sustaining voltage is nearly constant with current over a range of currents from about 200 microamps up to the current at which the cathode is covered with glow. The tubes are operated in parallel with a load of variable impedance, and the parallel combination is connected through a series resistance to a power supply. The voltage drop across the load therefore equals the sustaining voltage of the tube, the circuit parameters' being so chosen that the tube operates in the normal glow discharge. As the current drawn by the load increases, the current drawn by the tube decreases, and the voltage drop across the combination remains nearly constant.

A cold-cathode diode is essentially a two-state device. If a voltage less than the breakdown voltage is applied to the tube, practically no conduction takes place, and the device is characterized by almost infinite impedance. However, if the supply voltage, in series with a suitable resistance, is raised above the breakdown voltage, the glow discharge is established. The discharge in this second state is characterized by appreciable current conduction, a much lower impedance, and a sustaining voltage which is less than the breakdown voltage. By reducing the supply voltage sufficiently, the discharge can be caused to return to the low-conduction, or essentially nonconducting, state. This two-state nature of the tubes makes possible their use both as storage elements and as switching elements. In some tubes designed for switching applications, a third electrode located closer to the cathode is added to initiate the glow discharge.

## 15.1   The Glow Discharge, Ionization Time, and the Arc Discharge

### (a)   The Glow Discharge

The visible glow of the normal glow discharge is caused by excited atoms undergoing radiative transitions to lower states. It is observed only where there is an appreciable current of fast moving electrons which have sufficient energy to excite neutral atoms. Since the glow region of the normal glow discharge covers only a portion of the electrode area, this must be the only region where there is an appreciable current of fast electrons, and we must conclude that nearly all the cathode emission comes from the
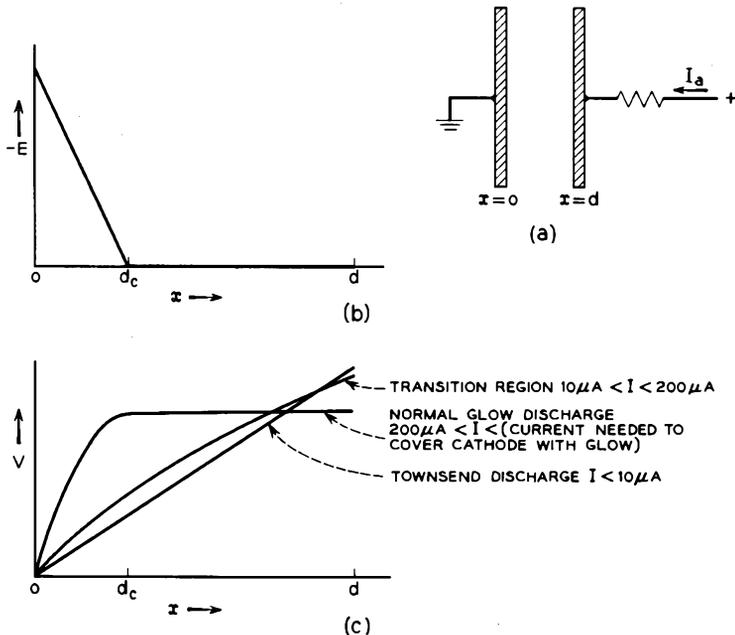


Fɪɢ. 15.1-1 Approximate plots of the electric field intensity and potential in the interelectrode space of a gas tube with planar, parallel electrodes.

glow-covered part of the cathode. Furthermore, since the glow area of the normal glow discharge is linearly proportional to the current drawn by the tube, it follows that the normal glow discharge is characterized by a constant current density over the glow area. For molybdenum electrodes and neon gas, this current density is found to be $J = 5p^2$ microamps/cm$^2$, where $p$ is the pressure in millimeters of Hg.

In a search for a better understanding of the mechanisms involved in the glow discharge, a number of investigators have made probe measurements of the potential in the interelectrode space of glow-discharge devices.[2] Unfortunately, such measurements always tend to disturb the field under observation, so that the information obtained is only approximate. However, most observers have concluded that the electric field intensity $E$ falls off in a nearly linear manner with distance from the cathode, becoming zero, or almost zero, at some distance $d_c$ from the cathode. A plot of the electric field intensity in the interelectrode space of a glow-discharge device may be something like that shown in Figure 15.1-1(b). Qualitative plots of the potential in the interelectrode space for three different ranges of currents are shown in Figure 15.1-1(c). The plots for the transition region and the normal glow discharge are made along a line running through the glow area perpendicular to the cathode.

The part of the glow discharge extending from the cathode out to $d_c$ is known as the cathode-fall region, since nearly all the voltage drop in the tube takes place in this region. A linear decrease of $E$ from the cathode out to $d_c$ and no electric field intensity beyond $d_c$ would imply a uniform net density of positive charge from the cathode out to $d_c$ and no net charge beyond that point. Beyond $d_c$ there is an ion-electron "plasma" consisting of approximately equal numbers of ions and electrons diffusing through the gas and perhaps drifting under the influence of weak electric fields. It is believed that the probability of ion-electron recombination in this region is small.

Because most of the potential drop in the glow discharge takes place close to the cathode, the metastables and photons produced in the discharge are, for the most part, generated closer to the cathode than in the Townsend discharge. This increases the probability of their reaching the cathode and provides a further reason for the lower sustaining voltage of the glow discharge.

The foregoing picture of the glow discharge leads to a number of questions about the mechanisms involved. Unfortunately, present understanding of the glow discharge is not at all complete, and a quantitative explanation of the shape of the potential distribution between the electrodes is not available. However, we can write down several conditions which mathematically describe the discharge. For a device with planar electrodes these are:

1. The discharge is self-sustaining. This implies that at the cathode

$$J_-|_{z=0} = \gamma(J_+|_{z=0} + GJ_-|_{z=0}) + \text{(contribution caused by photons striking the cathode)} \quad (15.1\text{-}1)$$

where $J_+|_{x=0}$ is the ion current density at the cathode, $J_-|_{x=0}$ is the electron current density at the cathode, and $G$ is the average number of metastables striking the cathode per electron released from the cathode.

2. The electric field intensity is given by

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon_o} = (n_+ - n_-)\frac{e}{\varepsilon_o} \qquad (15.1\text{-}2)$$

where $n_+$ and $n_-$ are the ion and electron densities, respectively.

3.
$$V(x) = \int_0^x E\,dx$$

$$V_a = \int_0^d E\,dx \qquad (15.1\text{-}3)$$

where $V_a$ is the sustaining voltage of the discharge, and $d$ is the electrode spacing.

4.
$$J_+ = n_+u_+e$$
$$J_- = n_-u_-e \qquad (15.1\text{-}4)$$
$$J_+ + J_- = J$$

where $u_+$ and $u_-$ are, respectively, the ion and electron mean velocities, and $J$ is the total current density.

The quantities $J_+$, $J_-$, $n_+$, $n_-$, $u_+$ and $u_-$ are all functions of distance $x$ from the cathode, whereas $J = J_+ + J_-$ is independent of $x$. The values of $u_+$ in the cathode-fall region are probably close to those plotted in Figure 14.3-1, but may differ slightly because of the rapidly changing field. However, very little is known about the distribution of electron velocities in the cathode-fall region and beyond. This means that $\eta$ cannot be obtained from the data given in Figure 14.3-3 which are valid only for a uniform electric field intensity. For similar reasons few data are available on the generation of metastables and photons in the cathode-fall region and beyond. At the present time opinions differ between experimenters on what fraction of the ions is generated in the cathode-fall region and what fraction is generated beyond the cathode-fall region.

If the coefficient $\eta'(x)$ giving the average number of ions generated by a single electron per volt potential rise were known together with the rates of generation of ions from metastable interactions, we could add one more equation to those given above, namely

5. $\quad \dfrac{dJ_+}{dx} = \dfrac{dJ_+}{dx} = \eta'(x)\,|J_-|\,|E| +$ (contributions from processes leading to metastable ionization)  $\quad(15.1\text{-}5)$

Because so many of the processes taking place in the glow discharge are interrelated, a complete mathematical description of the discharge would be very complex indeed. (It may be well to note once more that the voltage drop $V_c$ of the cathode fall is determined by the condition that the electrons passing through the cathode-fall region must gain sufficient energy to generate the current of discharge products which strike the cathode and which serve to maintain the electron emission. If $\gamma$ is 0.2, about 5 ions and metastables must strike the cathode for each electron leaving the cathode. Of course, additional energy is lost by the electrons in exciting atoms to radiating states, in increasing the thermal energy of the gas, and in generating metastables 'which do not reach the cathode. Consequently the voltage drop of the cathode fall must be at least several times the ionization potential of the gas. In a neon-molybdenum device, $\gamma$ is about 0.2 and $V_c$ is close to 110 volts, or 5.1 times the ionization potential of neon.)

Some further understanding of the problem can be obtained by assuming that the electric field intensity really does vary as shown in Figure 15.1-1(b). In this case we can use the preceding equations to estimate $E$, $\rho$, and $d_c$ in the cathode-fall region in terms of the voltage drop $V_c$ across the cathode-fall region and the observed cathode current density $J$.

Let the electric field intensity in the cathode-fall region be given by

$$E = -\frac{2V_c}{d_c}\left(1 - \frac{x}{d_c}\right) \tag{15.1-6}$$

where $V_c$ is the voltage drop of the cathode fall, and the minus sign in front of the right-hand side implies that the direction of $E$ is that of decreasing $x$. $V_c$ is approximately equal to $V_a$ over a range of $pd$ values up to $pd = 20$ mm of Hg $\times$ cm for neon-molybdenum devices, but at higher $pd$ there is a voltage rise in front of the anode. Since several ions on the average are needed to release one electron from the cathode and since the electron velocities in the cathode-fall region are many times those of the ions, the space charge in the cathode-fall region is primarily caused by ions. We can therefore write that

$$\frac{dE}{dx} = \frac{\rho_+}{\varepsilon_o} = \frac{J_+}{\varepsilon_o u_+} = \frac{2V_c}{d_c{}^2} \tag{15.1-7}$$

Note that since $E$ decreases linearly with $x$, $\rho_+$ is independent of $x$, and $J_+$ is proportional to $u_+$. From Equation (15.1-7), it follows that

$$(pd_c)^2 = \frac{2\varepsilon_o V_c u_+}{J_+/p^2} \tag{15.1-8}$$

It will be convenient to rewrite Equation (15.1-1) in the form

$$J_-|_{x=0} = \gamma'J_+|_{x=0} \tag{15.1-9}$$

where $\gamma'$ is greater than $\gamma$ by an amount that accounts for the contributions to the electron current density at the cathode caused by metastables and photons striking the cathode. Combining Equations (15.1-9) and (15.1-4), we obtain

$$J_+|_{x=0} = \frac{J}{1 + \gamma'} \qquad (15.1\text{-}10)$$

Since $J_+$ is proportional to $u_+$, in the cathode-fall region, we can use the values of both $J_+$ and $u_+$ at $x = 0$ in Equation (15.1-8). Combining that equation with Equation (15.1-10), we obtain

$$(pd_c)^2 = \frac{(1 + \gamma')2\varepsilon_o V_c u_+|_{x=0}}{J/p^2} \qquad (15.1\text{-}11)$$

Now from Equation (15.1-6),

$$\frac{E}{p}\bigg|_{x=0} = -\frac{2V_c}{pd_c} \qquad (15.1\text{-}12)$$

For a planar molybdenum cathode and neon gas, $V_c = 110$ volts and $J/p^2 = 5$ microamps/cm$^2$ $\times$ (mm of Hg)$^2$. $E/p$ at the cathode is of the order of 200, so that $\gamma = 0.20$ (Figure 14.5-3). $\gamma'$ may be of the order of 0.30 assuming that the metastables striking the cathode give rise to about a third of the cathode emission and that the contribution caused by photons is small. (A change of 20 per cent in the assumed value of $\gamma'$ would only change $pd_c$ by about 2 per cent.) Equations (15.1-11) and (15.1-12) are consistent for one particular value of $pd_c$ only. A choice of $pd_c = 1.08$ mm of Hg $\times$ cm is easily shown to satisfy both equations. From the second equation we find that $(E/p)|_{x=0} = 204$ volts/cm $\times$ mm of Hg. According to Figure 14.3-1 the corresponding value of $u_+|_{x=0}$ is $2.3 \times 10^3$ meters/sec. Substituting for $J/p^2$, $u_+|_{x=0}$, and $1 + \gamma'$ into the right side of the Equation (15.1-11), we obtain $pd_c = 1.08$ mm of Hg $\times$ cm, confirming our choice. (One must be a little careful of units in this last substitution. If $J$ is expressed in amperes per square centimeter, $u_+$ in meters per second, and $V_c$ in volts, then $d_c$ is in centimeters.) If the gas pressure in the device is known, $d_c$ is determined from $pd_c = 1.08$ mm of Hg $\times$ cm. The electric field intensity can be obtained using Equation (15.1-6), and $\rho_+$ can be obtained using Equation (15.1-7).

The fact that $pd_c$ is a constant means that the cathode-fall region is a definite number of mean free paths in thickness, the number being independent of pressure. A rough estimate of the number of mean free paths involved can be obtained using the data for the collision probability $P_c$ plotted in Figure 14.3-2. The electron mean free path is given by the reciprocal of $P_c$ divided by the gas pressure in millimeters of Hg. For electrons of 15 to 30 electron volts in neon $P_c \simeq 12$ (cm $\times$ mm of Hg)$^{-1}$, and

the mean free path $\simeq 1/(12p)$ cm, where $p$ is in millimeters of Hg. Combining this with $d_c = 1.08/p$ cm, we find that the cathode-fall region is approximately 13 mean free paths in thickness. (However this does not mean that only 13 collisions are made by the electrons on their way through the cathode-fall region, since in some of the collisions the electrons are deflected through large angles.)

Still another calculation of interest is the ratio of ions to gas molecules in the cathode-fall region. Now from Equation (15.1-7)

$$\rho_+ = \frac{2\varepsilon_o V_c p^2}{(pd_c)^2} \qquad (15.1\text{-}13)$$

The density of ions is given by $\rho_+/e$, and the ratio of ions to gas molecules is $\rho_+/ne$, where $n$ is the density of gas molecules. For a gas at $0°C$, $n = 3.54 \times 10^{22}\, p$ molecules/meter$^3$, where $p$ is the pressure in millimeters of Hg. In the cathode-fall region the gas temperature is usually somewhat higher than $0°C$, but for the purpose of an order-of-magnitude calculation it will suffice to use $n = 3.54 \times 10^{22}\, p$. For neon gas and planar molybdenum electrodes the ratio of ions to gas molecules in the cathode-fall region is therefore

$$\frac{\rho_+}{ne} = \frac{2\varepsilon_o V_c p}{3.54 \times 10^{22} e (1.08 \times 10^{-2})^2} = 3.0 \times 10^{-9} p \qquad (15.1\text{-}14)$$

For $p = 50$ mm of Hg, there is one ion for approximately 7 million gas molecules in the cathode-fall region.

Next let us consider the scaling of gas-discharge devices. It was pointed out earlier that the breakdown voltage of a gas diode with planar, parallel electrodes is a function of $pd$, where $p$ is the gas pressure and $d$ is the electrode spacing. It was also noted that if all the linear dimensions of a gas diode having arbitrarily shaped electrodes are scaled by a factor $k$, and if the
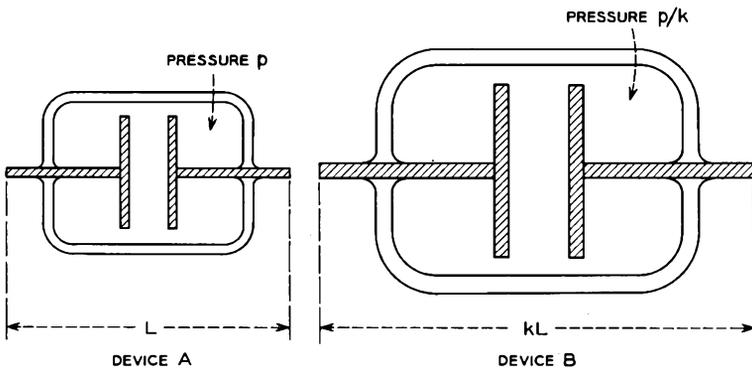


FIG. 15.1-2 The linear dimensions of device $B$ are $k$ times those of device $A$, and the pressure of the gas filling in device $B$ is $1/k$ times that in device $A$.

pressure is changed by $1/k$, the breakdown voltage to a first approximation remains unchanged. This is so because the average velocity gained by an electron in the interval between collisions is the same for the same applied voltage, and the average number of collisions which an electron makes in traveling between the electrodes is unchanged. (Or, stated another way, $E/p$ is the same.)

The same scaling law applies to glow discharges. Figure 15.1-2 shows two gas-discharge devices $A$ and $B$. If the linear dimensions of device $B$ are $k$ times those of device $A$ and if the pressure of the gas filling in device $B$ is $1/k$ times that of device $A$, the number of mean free paths between the cathode and anode of each device is the same. Consequently electrons traveling from the cathode to the anode of the two devices make the same average number of collisions on the way. If glow discharges are established in the devices, we would expect that the voltage difference per mean free path along a line joining the cathode and anode would be the same in each case (remembering that the discharges are self-sustaining). This means that the voltage at corresponding points between the electrodes would be the same, and the sustaining voltages would be the same. The latter point is, in fact, experimentally verified.

Now from Poisson's Equation we know that the net charge density $\rho$ is proportional to $\nabla^2 V$. But since $V$ is the same at corresponding points between the electrodes of the two devices, and since the dimensions of device $B$ are $k$ times those of device $A$, $\nabla^2 V$ in the interelectrode space of device $A$ is $k^2$ times that at corresponding points in device $B$. Consequently, the charge density $\rho$ in the interelectrode space of device $A$ is $k^2$ times that in device $B$. And since the voltage gained per mean free path is the same, the mean drift velocities of the charged particles at corresponding points between the electrodes is the same. Consequently, the cathode current density $J = \rho u$ of device $A$ is $k^2$ times that of device $B$, and $J/p^2$ is the same for the two devices. Since the cathode area of device $B$ is $k^2$ times that of device $A$, the cathodes are covered with glow at the same current.

Figure 15.1-3 shows how the sustaining voltage of a discharge tube with planar molybdenum electrodes and a neon filling varies with $pd$, where $d$ is the electrode spacing. The measurements were taken by varying $d$ with $p$ constant. For $5 < pd < 20$ mm of Hg $\times$ cm, the sustaining voltage is virtually constant. However, if $d$ is decreased so that $pd$ falls below 5 mm of Hg $\times$ cm, the sustaining voltage rises, apparently because the anode enters a region in which some of the excitation and ionization contributing to the discharge takes place. A discharge in a neon-molybdenum device with $pd < 5$ mm of Hg $\times$ cm is said to be "obstructed." (Since $pd_c$ for the cathode-fall region is only 1.08 mm of Hg $\times$ cm, it is evident that some of the metastables and ions reaching the cathode are generated well beyond
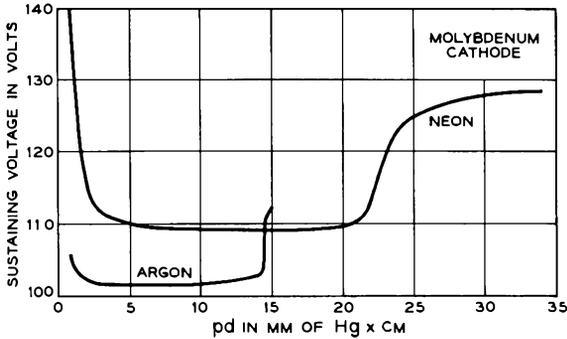
FIG. 15.1-3 The normal-glow-discharge sustaining voltage of devices with planar molybdenum electrodes and neon and argon fillings as a function of $pd$. In obtaining the curve for neon $p$ was held at 50 mm of Hg, and $d$ was varied. In obtaining the curve for argon $p$ was held at 25 mm of Hg, and $d$ was varied.

the cathode-fall region.) It would be reasonable to expect that the current density $J$ of the normal-glow discharge would be principally determined by events taking place at $pd < 5$ mm of Hg $\times$ cm. This is in fact the case, for if $d$ is varied with $p$ constant, $J$ remains unchanged provided $pd > 5$ mm of Hg $\times$ cm. When this result is combined with the scaling law for $J/p^2$ dis-
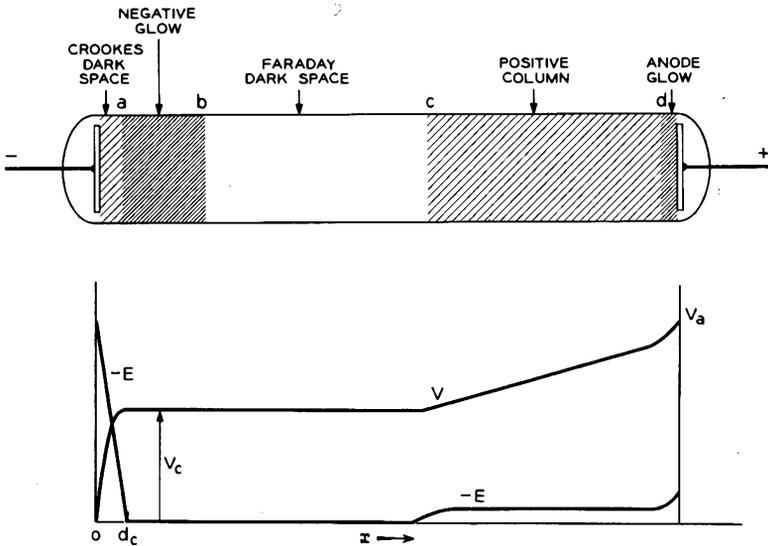


FIG. 15.1-4 The potential $V$ and electric field intensity $E$ in a long tube with $pd \gg 20$ mm of $Hg \times$ cm.

cussed above, it follows that $J/p^2$ for the normal-glow discharge is constant for all neon-molybdenum devices with planar electrodes and $pd > 5$ mm of Hg $\times$ cm.

If $pd$ is increased beyond about 20 mm of Hg $\times$ cm the sustaining voltage $V_a$ again rises. At somewhat higher values of $pd$, a region of glow appears in front of the anode and oscillations in the discharge can be detected. Figure 15.1-4 shows a tube in which $pd$ is assumed to be much greater than 20 mm of Hg $\times$ cm together with a plot of potential $V$ and the electric field intensity $E$. Several regions of visible glow can be distinguished in such a tube. The cathode-fall region extends from the cathode out to $a$. Relatively little light is emitted from this region, and it is frequently called the *cathode dark space*, or the *Crookes dark space*. Between $a$ and $b$ a much brighter region of glow known as the *negative glow* can be seen. In devices with $pd < 20$ mm of Hg $\times$ cm most of the visible glow comes from this region. Excitation and ionization in the region are thought to be largely due to fast electrons arriving from the cathode fall. In a neon-molybdenum device the outer edge of the negative glow corresponds to a $pd$ of about 5 mm of Hg $\times$ cm. Between $b$ and $c$ there is a region, known as the *Faraday dark space*, from which very little radiation takes place, apparently because most of the free electrons do not have enough energy to excite the gas molecules. Between $c$ and $d$ a region of diffuse glow known as the *positive column* can be seen; and finally, between $d$ and the anode, there is a somewhat more intense glow known as the anode glow.

The negative glow region, the Faraday dark space, and the positive column are all plasmas, or regions in which the ion and electron densities are approximately equal. Because the electron velocities in the plasma are much greater than the ion velocities, the current of electrons reaching the edge of the plasma in front of the glass walls is much greater than the ion current. The walls therefore become negatively charged, and a "sheath" forms between the plasma and the glass. The voltage drop in the sheath serves to reduce the electron current reaching the glass walls to a value equal to the ion current. (Since the walls are insulating, the electron and ion currents to the walls must be equal.) Some discussion of plasmas and sheaths is given in the first section of Chapter 16.

Fast electrons from the cathode fall expend most of their energy in the negative-glow region where most of the excitation and ionization needed to support the cathode emission take place. In the Faraday dark space few electrons have sufficient energy to cause excitation. Conduction through the Faraday dark space results largely from diffusion of electrons from the negative-glow region to the cathode end of the positive column. Ions enter the Faraday dark space both from the negative glow and from the positive column.

The electron current reaching the sheath at the edge of the plasma is generally hundreds of times greater than the ion current, so that electrons reaching the edge of the plasma, for the most part, are reflected by the sheath. However ions reaching the edge of the plasma are drawn into the sheath and neutralized upon reaching the walls. A nearly uniform gradient of potential extends over the length of the positive column giving the electrons sufficient energy to generate ion-electron pairs needed to make up for ion and electron losses to the walls. The nature and appearance of the positive column depend considerably on the product of the filling pressure and tube radius as well as on the current density. A variety of types of positive column can be obtained by varying these factors,[3] but a discussion of the possible forms of positive column would be outside the scope of the present treatment. The positive column is responsible for the radiation observed from "neon" advertisement signs and gas lasers and is indirectly responsible for the radiation emitted from fluorescent lamps.

If the current drawn by a cold-cathode tube is increased beyond the point where the cathode is completely covered with glow, the voltage drop across the tube rises, and the current density of the discharge increases linearly with the total current. Such a discharge is said to be an abnormal glow discharge. The fact that the discharge is restricted to a cathode of fixed size area implies a further boundary condition for the abnormal glow discharge in addition to the conditions listed earlier for the normal glow discharge. Probe measurements of the cathode-fall region of the abnormal glow discharge[4] show that $d_c$ decreases with increasing current, as would be expected from Equation (14.1-11).

## (b)  Ionization Time

The ionization time of a cold-cathode device is defined as the time that elapses between the application of a voltage greater than breakdown between the anode and cathode of the tube and the initiation of the glow discharge. We shall assume we have a device in which $pd$ is a little greater than the $pd$ value which gives minimum breakdown. If a voltage just above breakdown is applied to the anode, an electron leaving the part of the cathode surface closest to the anode on the average will generate sufficient excited atoms and ions to release slightly more than one electron from the cathode. A current buildup results which soon produces sufficient space charge to initiate the glow discharge.

The ionization time is made up of two parts. One part is the delay which occurs between the application of voltage to the anode and the release from the cathode of the electron or electrons which initiate the current buildup.

[3]Reference 15.6, p. 155.
[4]Reference 15.1.

The second part is the time required for the current to increase to the point where the glow discharge is established. In general, both parts decrease as the excess of the applied voltage over the breakdown voltage increases.

Some means to initiate the release of electrons from the cathode is generally provided in cold-cathode tubes. Often a spot of radioactive material is painted on the inside of the envelope close to the anode-to-cathode gap. This ensures frequent ionizing events in which high-energy $\alpha$ or $\beta$ particles are released into the filling gas causing ionization of the gas. A few of the ions and electrons generated in this manner diffuse through the gas to the anode-to-cathode gap. When voltage is applied to the anode, the electrons are drawn to the anode, perhaps causing further ionization on the way, and the ions are drawn to the cathode, where some may release electrons by the $\gamma$ process. Because only a few charged particles reach the interelectrode space, several ionizing events may take place before the current buildup is initiated.

If the radioactive material used is a compound of radium, a small quantity of radioactive radon gas is generated by the radium. The radon mixes with the filling gas and ensures some radioactivity throughout the interior of the tube.

In some tubes a radioactive isotope of krypton, known as krypton 85, is added to the filling gas. The nucleus of this isotope undergoes a radioactive decay in which it releases a $\beta$ particle which in turn causes ionization of the filling gas.

Tubes in which radioactive material has been placed for the purpose of obtaining short ionization times are said to have a radioactive keep-alive. The ionization time of these tubes is, of course, a function of the amount of radioactive material, the proximity of the decaying particles to the anode-to-cathode gap, and the excess of the supply voltage over the breakdown voltage, a quantity referred to as the "overvoltage." If the overvoltage is 10 volts, average ionization times of a few hundred microseconds can readily be obtained with a radioactive keep-alive. The time required for the current buildup with such an overvoltage may be several tens of microseconds. Because of the random nature of the radioactive decay, the ionization time is a statistical quantity and is best described by a distribution of individual measurements.

In some tubes a low-current discharge between auxiliary electrodes is operated continuously as a keep-alive mechanism. Such a discharge is called a dc keep-alive. Generally, the current drawn by the dc keep-alive is of the order of a few microamperes, although smaller currents are sometimes sufficient. Ions and electrons generated by the auxiliary discharge find

their way to the gap between the main anode and cathode and are re-
sponsible for initiating the current buildup in the main gap.

A third keep-alive mechanism, known as a photoelectric keep-alive, in-
volves the use of a surface which emits photoelectrons and which is held
close to cathode potential. If light falls on such a surface, a current of
photoelectrons is emitted. When voltage is applied to the anode, the photo-
electrons are drawn to the anode causing ionization on the way and initiat-
ing the discharge.

If no keep-alive mechanism is provided in a cold-cathode tube, the dis-
charge is eventually initiated by residual radioactivity within the tube or by
a cosmic ray event. In such cases ionization times of the order of a second or
more may be encountered.

Finally, if a very high voltage is applied across a short gap, the discharge
may be initiated by field emission from the cathode. Two common gas-
discharge devices in which the discharge is initiated by field emission are the
spark plug and the carbon-block lightning protector used on telephone
lines to prevent voltage surges from reaching central office or subscriber
equipment.

### (c)   The Arc Discharge

If the current density of the abnormal glow discharge is increased suffi-
ciently, the sustaining voltage $V_a$ reaches a maximum, after which it de-
creases either continuously or abruptly. The drop in $V_a$ is accompanied by
a transition to a form of discharge known as an arc. The arc is a self-
sustaining discharge which is characterized by a high current density
($\sim 10^3$ to $10^6$ amps/cm$^2$) and a low sustaining voltage ($\sim 10$ to 20 volts).
Usually the area of the cathode surface covered with glow from the dis-
charge is quite small. The visible light from the discharge always contains
strong spectral lines of the cathode material.

The low sustaining voltage implies that electrons are emitted from the
cathode by some supplementary mechanism in addition to the $\gamma$ process. It
has been shown that thermionic emission plays an essential role in arcs
formed with carbon and tungsten cathodes, the intense heat needed to
maintain the thermionic emission being generated by ion bombardment of
the cathode. However the electron-emission mechanism involved in arcs
formed with copper or liquid mercury cathodes seems less certain. Field
emission is a possibility.[5] (It can be shown that the temperatures needed

---

[5]Reference 15.6, p. 140; References 15.7 and 15.8. See also discussion of other possible
mechanisms by A. Von Engel, *Ionized Gases*, Oxford University Press, Oxford, England,
1955, Chapter 9 Sections 5 and 6, and discussion by J. M. Meek and J. D. Craggs,
*Electrical Breakdown of Gases*, Oxford University Press, Oxford, England, 1953, Chapter 12.

for appreciable thermionic emission from copper or liquid mercury cathodes would be sufficient to vaporize the cathode material.) If field emission is the mechanism involved, the large fields needed to release electrons from the cathode are probably generated by a high density of ions drifting toward the cathode.

## 15.2   Three Examples of Cold-Cathode Tubes

In this section we shall describe three cold-cathode tubes manufactured by the Western Electric Company, each of a different construction and each used for a different purpose. The first is the 423C, a two-electrode device used as a voltage-reference tube in power supplies. The second is the 427A, a three-electrode tube used as a voltage regulator in power supplies. The third is the 426A, a three-electrode tube used as a switching element in a telephone circuit which permits selective ringing of subscribers on party lines.

### The 423C

Figure 15.2-1 illustrates the construction of the 423C voltage reference tube. The electrodes are molybdenum and the gas filling is argon, a combination that has been found to give sustaining voltages which are par-
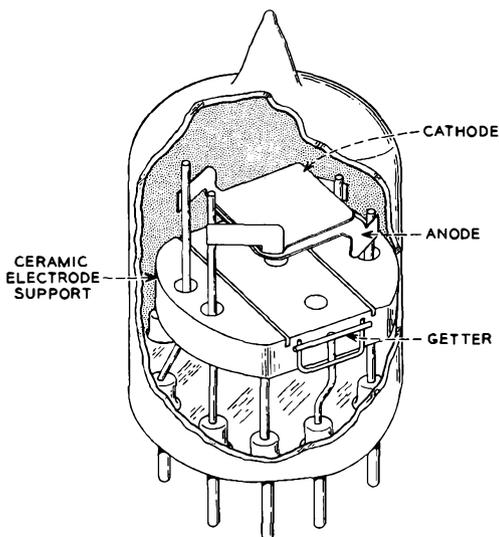


FIG. 15.2-1 The 423C voltage reference tube. The overall height of the tube is 4.4 cm.

ticularly stable with time. The back of the cathode is coated with powdered nickel which has a higher work function and hence lower $\gamma$ than molybdenum. This serves to restrict the glow to the inner face of the electrode. A small amount of radioactive krypton gas is added to the argon filling gas as a radioactive keep-alive. The getter shown in the figure is used to evaporate a thin film of barium on the inside of the glass envelope during processing of the tube. The barium "getter flash" serves to adsorb molecules of the non-noble residual gases within the tube.

Principal electrical characteristics of the tube are shown in Table 15.2-1.

TABLE 15.2-1 WE 423C OPERATING CHARACTERISTICS

*Typical Values*

Breakdown voltage, volts..........................135
Sustaining voltage, volts..........................100
Operating current, ma............................4 to 8

The pressure of the gas filling is 32 mm of Hg, and the electrode spacing is 0.05 cm. The product $pd$ is therefore 1.6 mm of Hg $\times$ cm. This choice of $pd$ represents a compromise between a desire to obtain minimum breakdown voltage and a desire to keep out of the obstructed-discharge region. Figure 14.6-3 shows that minimum breakdown voltage for a device with planar molybdenum electrodes and an argon filling is obtained at a $pd$ of 1.0 mm of Hg $\times$ cm. However, Figure 15.1-3 indicates that the discharge is obstructed at this $pd$. Minimum breakdown voltage is desirable from the standpoint of the power supply applications in which the tube is used, but since the sustaining voltage of an obstructed discharge is thought to be less stable on a long-term basis, a somewhat larger $pd$ is used.

The cathode area of the tube is 0.46 cm, and the normal glow discharge covers the cathode at a current of 8 milliamps. The parameter $J/p^2$ is therefore 17 microamps/cm$^2$ $\times$ (mm of Hg)$^2$. Using this value of $J/p^2$ and assuming $\gamma' = 0.12$, the quantity $pd_c$ can be calculated from Equations (15.1-11) and (15.1-12). It is found that $pd_c = 0.45$ mm of Hg $\times$ cm. Since $pd = 1.6$ mm of Hg $\times$ cm, the cathode-fall region of the discharge extends about 0.3 of the way from the cathode to the anode.

Tubes of the 423C type are found to drift less than a volt in sustaining voltage in a year of continuous operation. Part of this stability is achieved through careful processing to ensure that the electrodes are particularly clean and hence that $\gamma$ remains constant throughout the life of the device. The processing includes heating the electrodes inductively with rf to bright red temperatures while the tubes are being pumped. This cleans the electrode surfaces and drives contaminant gas from the interior of the metal. After pumping is completed, the tubes are filled with the argon-krypton 85

mixture and sealed off. They are then subjected to operation at relatively high currents including pulsed operation at currents of the order of several amperes. This processing, known as aging, causes molybdenum to be sputtered from the surface of the cathode onto the anode and the surrounding walls, leaving the cathode particularly clean. The sputtering results from positive ions striking the cathode and imparting their kinetic energy to molybdenum atoms or groups of atoms at the surface. The high current densities at which the tubes are aged (and which place the discharge far into the abnormal-glow region) are found to increase the sputtering rate greatly. At normal operating currents of 4 to 8 ma, the cathode sputtering is very slight.

The 423C is capable of almost indefinite life, since at normal operating currents there is nothing which "wears out" in the tube.

*The 427A*

Figure 15.2-2 shows the construction of the 427A voltage regulator tube. The electrodes are molybdenum and the gas filling is a mixture of argon and a small amount of krypton 85 at a pressure of 21 mm of Hg. The electrode configuration shown in the figure gives electrical characteristics which are similar to those of devices with planar, parallel electrodes. By folding the cathode around the anode, a relatively large cathode area can be enclosed in a relatively small envelope. Like the 423C, the outside of the cathode is coated with a material which restricts the glow to the inner surface of the cathode.
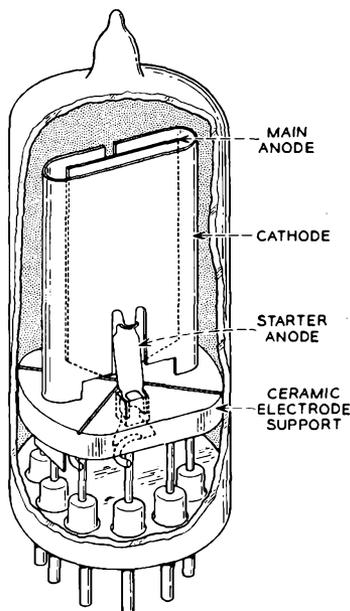


Fig. 15.2-2 The 427A voltage regulator tube. The overall height of the tube is 5.5 cm.

Important electrical characteristics of the tube are shown in Table 15.2-2.

TABLE 15.2-2 WE 427A OPERATING CHARACTERISTICS

*Typical Values*

Main-anode breakdown voltage, volts...............165
Main-anode sustaining voltage, volts................100
Starter-anode breakdown voltage, volts..............125
Recommended operating current, ma................5 to 40

The product $pd$ for the main-anode gap is equal to 3.0 mm of Hg × cm.

The tube is processed in a manner similar to the 423C, and its main-anode sustaining voltage is almost as stable as that of the 423C. Like the 423C it is capable of almost indefinite life.

A typical circuit in which the tube might be used is illustrated in Figure 15.2-3. As the load increases, the current drawn by the 427A decreases. If the current drawn by the 427A falls in the range from 5 to 40 ma, the sustaining voltage of the tube is very nearly constant, and the voltage drop across the load remains nearly constant.
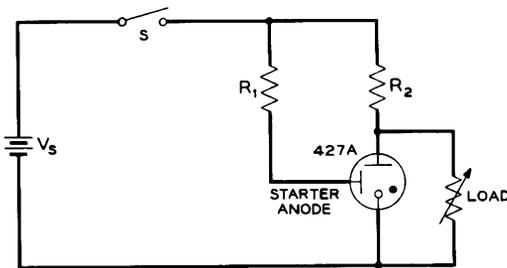


FIG. 15.2-3 A circuit in which the 427A is employed as voltage regulator.

The supply voltage $V_s$ in Figure 15.2-3 must at least equal the starter-anode breakdown voltage if conduction is to take place when the switch $S$ is closed. If $V_s$ is greater than the starter-anode breakdown voltage, closing the switch initiates a discharge between the starter anode and the main cathode. If the current of this discharge is greater than about a tenth of a milliampere, the discharge transfers immediately to the main anode. For smaller starter currents, transfer of the discharge to the main anode depends on the applied anode voltage.

*The 426A*

The construction of the 426A switching triode is illustrated in Figure 15.2-4. The tube is used in an application which requires a low sustaining voltage and a high main-anode breakdown voltage. The stability of the sustaining voltage is not at all critical, and the total life of the tube does not need to be more than 100 hours at a current of 30 ma. Two features of the design contribute toward a low sustaining voltage:

1. The cathode consists of nickel sheet coated with a mixture of barium and strontium oxides. This gives a low work function and a high $\gamma$.

2. The gas filling is a neon-argon mixture consisting of approximately 95 per cent neon and 5 per cent argon. The neon-argon mixture gives a

high $\eta$ because neon metastables ionize argon molecules. (The gas pressure is 40 mm of Hg. A very small amount of krypton 85 is added to the filling gas as a radioactive keep-alive.)

High main-anode breakdown voltage is obtained by locating the anode far from the cathode. This in turn leads to oscillations in the discharge, but
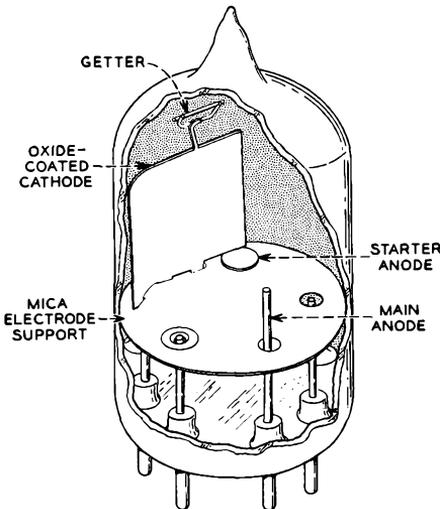


FIG. 15.2-4 The 426A switching triode. The overall height of the tube is 4.4 cm.
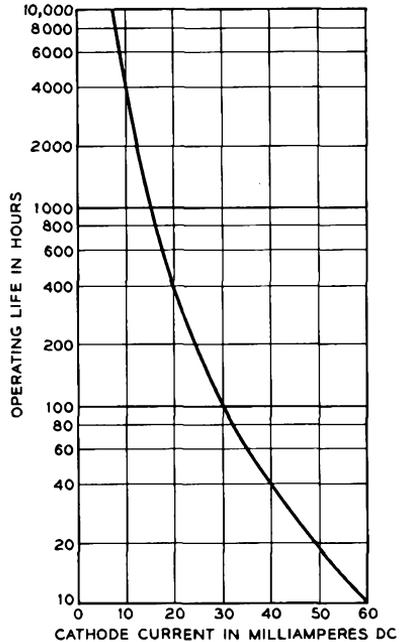
FIG. 15.2-5 Expected life of the 426A in hours as a function of the cathode current.

the oscillations are not a concern in the applications in which the tube is used.

Principal electrical characteristics of the 426A are shown in Table 15.2-3.

TABLE 15.2-3 WE 426A OPERATING CHARACTERISTICS

*Typical Values*

| | |
|---|---|
| Main-anode breakdown voltage, volts | >180 |
| Main-anode sustaining voltage, volts | 69 |
| Starter-anode sustaining voltage, volts | 72 |
| Operating current, ma | 30 |
| Ionization time, starter gap, milliseconds | 10 |

The starter-gap ionization time is measured with an applied voltage which exceeds the starter-gap breakdown voltage by 15 volts.

The life of cold-cathode tubes with oxide-coated cathodes is found to decrease rapidly with increasing current because of a corresponding increase in the sputtering of the cathode coating. When most of the oxide coating is sputtered away, the sustaining voltage starts to rise. End-of-life of tubes of this type is frequently measured in terms of a rise in sustaining voltage above a preselected value. Figure 15.2-5 shows how the expected life of the 426A varies with current. The cathode is covered with glow at about 10 milliamps.

## 15.3   The Hollow-Cathode Discharge and the Stepping Tube

If a cold-cathode device is constructed in which the cathode has a re-entrant or "hollow" part, a quite different form of glow discharge can sometimes be obtained. The phenomenon is known as the hollow-cathode discharge. Figure 15.3-1 shows several possible forms of hollow-cathode. The
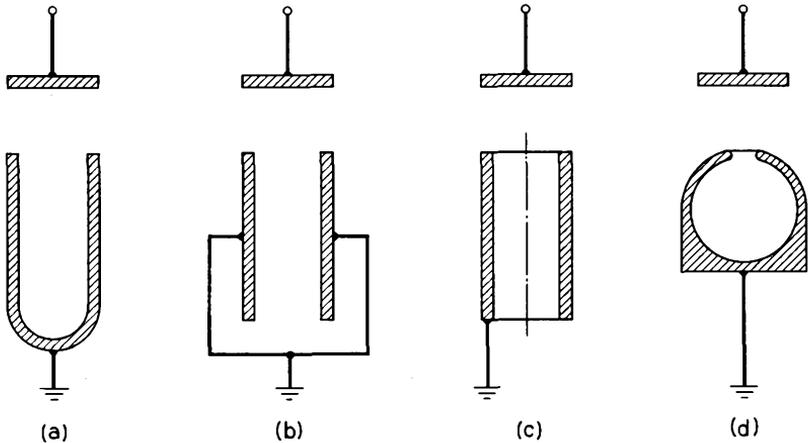


Fig. 15.3-1 Several possible forms of hollow cathode.

effect can be obtained with a U-shaped cathode, two parallel planar cathodes, a cylinder, or a spherical cavity. In order to obtain the hollow-cathode discharge, the product of the filling pressure $p$ and the distance $a$ across the hollow must be such that electrons leaving opposite inner sides of the hollow enter a common negative-glow region in the central portion of the hollow. In the case of a neon-molybdenum device with two parallel, planar cathodes, $a$ is taken to be the distance between the cathodes, and the

product $pa$ must be less than about 10 mm of Hg × cm and greater than some lower limit, which probably is less than 1 mm of Hg × cm.

Characteristics of the hollow-cathode discharge are:

1. Except at very low discharge currents, the sustaining voltage of the hollow-cathode discharge is lower, for a given discharge current, than the sustaining voltage of a device with a planar cathode having the same surface area as the hollow cathode. This means that when the discharge is first initiated, the current emission and glow build up on the inside of the hollow, drawing the anode voltage below the sustaining voltage for a discharge on the outside of the cathode. Consequently, over a wide range of discharge currents, very little glow is observed on the outside. (However, if the discharge current is increased sufficiently, the glow eventually spreads to the outside of the hollow.)

2. Depending on the filling pressure and the dimensions of the cathode, the current needed to cover the inner sides of two parallel planar cathodes with glow is sometimes as much as several hundred times that needed to cover an equal area of a planar cathode. The same applies to devices with U-shaped cathodes or cylindrical cathodes.

The mechanisms responsible for the anomalous behavior of the hollow-cathode discharge are not well understood at present, and we shall not attempt an explanation of them.
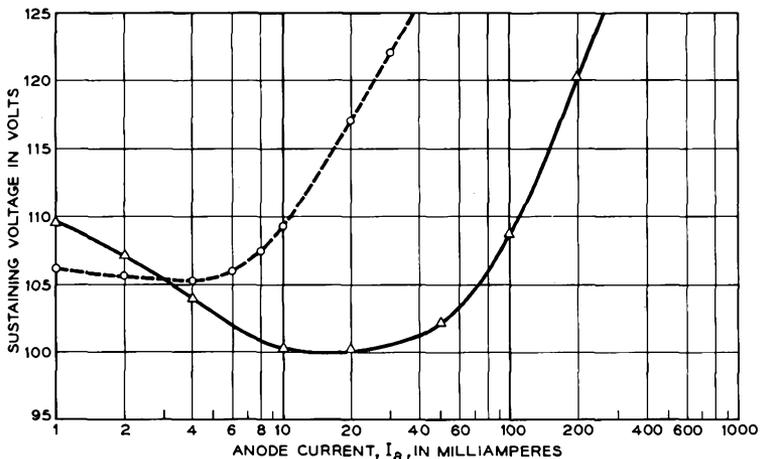


FIG. 15.3-2 (———) Anode voltage vs. anode current curve for a device with a hollow cathode of the type shown in Figure 15.3-1(d). (– – –) Anode voltage vs. anode current curve for a device with planar, parallel electrodes. Both tubes have molybdenum cathodes and neon fillings.

In Fig. 15.3-2 the solid curve shows a plot of sustaining voltage vs. discharge current measured for a cathode of the type shown in Figure 15.3-1(d) The cavity diameter was 0.75 mm, and the tube was filled with neon to a pressure of 98 mm of Hg. The product of pressure and diameter was therefore close to 7.5 mm of Hg × cm, and the area of the inside of the cavity was close to 0.017 cm². A planar cathode of equal area would be covered with glow at a current of about 0.8 milliamp at the same filling pressure.

The broken curve in the Fig. 15.3-2 shows a plot of the sustaining voltage vs. discharge current measured for a neon-molybdenum device with a planar cathode of area 0.45 cm², a planar, parallel anode, and a filling pressure of 60 mm of Hg. The cathode was covered with glow at about 8 ma. (The product $pd$ for the device was 7.5 mm of Hg × cm, where $d$ is the anode-to-cathode spacing.) The back of the cathode was coated with a material of lower $\gamma$ than that of molybdenum so that the glow would remain only on the face toward the anode. The higher sustaining voltage of the cavity cathode at very low currents can probably be attributed to the constriction of the discharge by the small opening to the cavity.
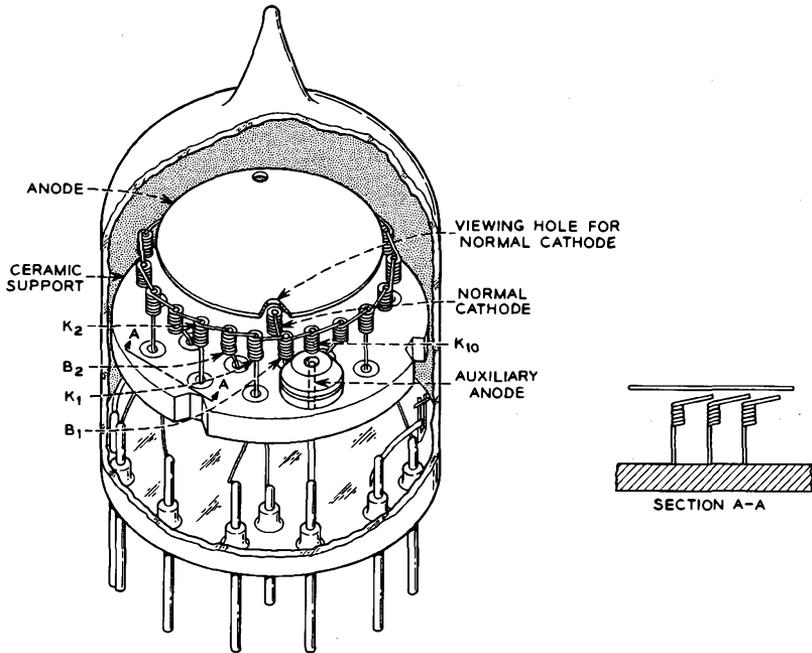


FIG. 15.3-3 The construction of the 439A stepping tube. The overall height of the tube is 5.6 cm.

The hollow-cathode effect is used in the 439A counting or stepping tube illustrated in Fig. 15.3-3. The tube has 20 cathodes arranged in a ring around a common disc-shaped anode. Alternate cathodes, called output cathodes, are connected to external leads and identified as $K_1$, $K_2$, ... $K_{10}$. The remaining cathodes are called stepping cathodes and identified as $B_1$, $B_2$, ... $B_{10}$. The stepping cathodes are joined together internally in two groups of five, each group having an external lead. For most circuit applications the two groups of five are connected together externally.

The tube is operated so that the discharge is normally on one of the $K$ cathodes. A negative pulse applied to the stepping cathodes causes a discharge on $K_1$ to transfer to $B_2$ during the pulse and to $K_2$ at the end of the pulse. A second pulse causes the discharge to transfer to $K_3$, and so on. The electrode arrangement permits both electrical and visual counting of the pulses. A normal cathode inside the ring of 20 can be used to "zero" the position of the discharge before a counting operation. An auxiliary anode located under $K_{10}$ can be used to obtain an output every tenth pulse so that several tubes can be used in a "decade counter." In a three-tube decade counter, one tube counts every pulse, a second tube counts every tenth pulse, and the third tube counts every hundredth pulse.

The key to the transfer mechanism lies in the shape of the cathodes. Each cathode consists of a molybdenum wire wound into a coil forming a hollow region with a straight portion of the wire, called a pick-up tab, at the upper end. The discharge is normally in the hollow part of the cathode, since it is more efficient there, and the sustaining voltage is lower. The pick-up tab extends over the end of the hollow part of an adjacent cathode and serves as a preference mechanism which causes the discharge to transfer in one direction only. The discharge current is usually about 2 ma. (The tube is filled with neon to a pressure of 105 mm of Hg.)
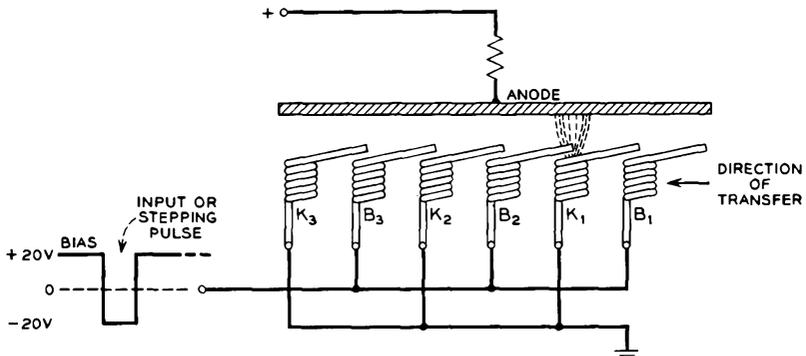


FIG. 15.3-4 A number of $B$ and $K$ cathodes arranged in a row.

Figure 15.3-4 shows a number of $B$ and $K$ cathodes arranged in a row. Alternate $B$ cathodes are connected together and serve to step the discharge from one $K$ cathode to another. The discharge is shown to be on $K_1$ in the figure, and the pick-up tab of $B_2$ extends into the region of ionization of the discharge. The $B$ cathodes are biased 20 volts positively with respect to the $K$ cathodes. A negative pulse of $-40$ volts superimposed on the bias voltage makes the voltage difference between $B_2$ and the anode greater than the sustaining voltage for a discharge on the pick-up tab of $B_2$. The discharge therefore quickly builds up on the pick-up tab of $B_2$, drawing the anode voltage down so that it is less than the sustaining voltage for the discharge on $K_1$. The glow on $K_1$ therefore extinguishes. Transfer of the discharge to the pick-up tab of $B_2$ is rapidly followed by a buildup of current in the hollow of $B_2$, since the sustaining voltage of the discharge in the hollow is less than on the pick-up tab. This causes the anode voltage to fall still further, and the glow on the tab extinguishes.

Thus the negative pulse applied to the $B$ cathodes causes the glow to transfer from the hollow part of $K_1$ to the hollow part of $B_2$. Removal of the pulse makes the $B$ cathodes more positive than the $K$ cathodes and causes the glow to transfer from the hollow of $B_2$ to the hollow of $K_2$. Successive pulses applied to the $B$ cathodes cause the discharge to advance one more $K$ cathode to the left for each pulse.

By applying a sufficiently large negative pulse to the normal cathode, the discharge can be transferred from any of the $K$ cathodes to the normal cathode, since the anode voltage falls to a value insufficient to maintain a discharge on any of the other cathodes. The stepping cathode $B_1$ has two pick-up tabs which enable it to transfer the discharge to $K_1$ from either the normal cathode or $K_{10}$.

In most circuit applications the auxiliary anode is connected through a series resistance to the anode supply. Because the auxiliary anode is shielded from all cathodes except $K_{10}$, it draws current only when the discharge is on $K_{10}$. When the glow transfers to $K_{10}$, the auxiliary anode potential falls giving a negative output pulse which can be used to drive a second stepping tube. Positive output pulses can be obtained from any of the $K$ cathodes by placing a resistance of suitable size in series with the cathode load.

With suitable external circuitry the 439A can be used to count pulses at repetition rates as high as five kilocycles.

## REFERENCES

Several texts on gas tubes and gas-discharge phenomena are listed at the end of Chapter 14. A particularly good description of the glow discharge is given by G. Francis, *Handbook of Physics*, Vol. 22, p. 53, Julius Springer, Berlin, 1956.

Other references covering specific subjects discussed in this chapter are:

15.1. R. Warren, *Phys. Rev.* **98,** 1650, 1955.

15.2. R. Warren, *Phys. Rev.* **98,** 1658, 1955.

15.3. R. P. Stein, *Phys. Rev.* **89,** 134, 1953.

15.4. P. F. Little and A. Von Engel, *Proc. Roy. Soc. (London)* **A 224,** 209, 1954.

15.5. W. H. Ernst, *Helv. Phys. Acta* **8,** 381, 1935.

15.6. M. J. Druyvesteyn and F. M. Penning, *Rev. Modern Phys.* **12,** 87, 1940.

15.7. W. S. Boyle and F. E. Haworth, *Phys. Rev.* **101,** 935, 1956.

15.8. K. G. Hernquist, *J. Appl. Phys.* **27,** 1226, 1956.

# Chapter 16

## HOT-CATHODE TUBES

In the normal glow discharge described in the previous chapter the current of discharge products striking the cathode amounts to several ions, metastables, and photons per electron released from the cathode. This relatively large ratio of incident discharge products to emitted electrons is made necessary by the low efficiency of the $\gamma$ and photoelectric emission processes. Since the discharge is self-sustaining, the emitted electrons must gain sufficient energy to generate the current of discharge products striking the cathode, and consequently the voltage drop in the cathode-fall region must be at least several times the ionization potential of the gas.

If a thermonically emitting cathode is used in place of the cold cathode, and if the thermionic emission current is at least equal to the current flowing in the external circuit, the voltage drop across the tube is much smaller, in many cases less than the ionization potential of the gas. Hot-cathode tubes filled with mercury vapor often conduct currents of several amperes at a voltage drop of 6 to 9 volts, despite the fact that the ionization potential of mercury is 10.4 volts.

The high-current, low-voltage properties of the hot-cathode discharge make possible the design of hot-cathode rectifier tubes for use in high-voltage power supplies, the power losses in the tubes being small compared with the power delivered to the load. The hot-cathode discharge is also used in some switching tubes where a low anode-to-cathode voltage drop is desirable. Both rectifier and switching tubes usually show a nearly constant voltage drop as the current is varied from a relatively small value up to the cathode thermionic emission current At higher currents the anode-to-cathode voltage rises.

In the conducting state most of the interelectrode space of a hot-cathode gas tube is filled with a plasma in which the ion and electron densities are approximately equal, and in which potential gradients are small. Often

the potential gradient in the direction from the cathode to the anode is of
the order of a few tenths of a volt per centimeter. Between the plasma and
the cathode there is a region of nonuniform charge, called a sheath, in which
a voltage drop of several volts takes place. Hot-cathode tubes are generally
operated with the cathode current space-charge-limited so that there is a
small potential minimum just outside the cathode. Ions flow from the
plasma toward the cathode, and electrons drawn from the potential min-
imum flow toward the plasma. Often the electron current in the cathode
sheath is several hundred times the ion current.

Sheaths also form between the plasma and the walls of the tube and
between the plasma and the anode. The walls are at a negative potential
with respect to the plasma, while there is often a small rise in potential in
going from the plasma to the anode.

The cathode of a hot-cathode rectifier or switching tube frequently con-
sists of a nickel ribbon or mesh coated with barium and strontium oxides or
other emissive material. The cathode is heated by passing a current
through it. Other tubes have indirectly heated, oxide-coated nickel
cathodes. Because power must be supplied to heat the cathode and hence
to maintain the discharge, the discharge is not self-sustaining.

Generally quite low filling pressures are used, in some cases a few tenths
of a millimeter of Hg, but often as small as a few thousandths of a millimeter
of Hg. The use of low filling pressures serves two purposes. It reduces the
cooling of the cathode by the gas, and it makes the product of the pressure
and the maximum distance between the electrodes much smaller than that
which gives minimum breakdown voltage. As a result, relatively high in-
verse voltages (anode negative with respect to the cathode) can be applied
without cold-cathode breakdown occurring. This is particularly desirable in
high-voltage rectifier applications where the anode may be several thousand
volts negative with respect to the cathode during part of the ac cycle. The
gas fillings are usually noble gases, or mercury vapor in equilibrium with
liquid mercury, or a combination of the two. In some special-purpose tubes
hydrogen is used. However, gases with diatomic or polyatomic molecules
lead to much higher voltage drops and consequently are not generally em-
ployed.

In all gas tubes there is a tendency for the discharge to cause a small
amount of the filling gas to become embedded or entrapped in the electrodes
and walls of the tube. This gas "cleanup" is thought to result from ions
impinging on the inner surfaces of the tube where they become entrapped
by material sputtered from the cathode. In cold-cathode tubes where the
filling pressure is relatively high, gas cleanup is seldom of importance
since the fractional loss of pressure is usually quite small. However, in hot-
cathode tubes the fractional loss of pressure during the life of the tube can

be appreciable because of the much smaller initial filling pressure. Tubes filled with noble gases at too low an initial filling pressure are likely to fail during operation life because of gas cleanup.

The use of mercury vapor in equilibrium with liquid mercury to provide the filling gas has the advantage that gas cleanup merely results in some of the condensed mercury being converted to vapor without any change in pressure. The mercury condenses on the coolest part of the tube, and the temperature of the condensed mercury serves to determine the pressure of the vapor in equilibrium with the liquid. Generally the mercury vapor pressure lies between $10^{-3}$ and $10^{-1}$ mm of Hg.

Most of the visible light from the discharge comes from the plasma. If a tube is constructed with a large electrode spacing, and the tube geometry is such that the plasma region is long and narrow, as in the case of the fluorescent lamp, the plasma is said to be a positive column. The light radiation from the positive column is used for illumination purposes in several large classes of tubes, including fluorescent lamps and mercury-vapor lamps. The cathodes of these tubes are heated by ions incident upon the cathode and are said to be ionically heated cathodes. The discharge in this case is a self-sustaining discharge, since no additional power is expended to maintain conduction.

In the first part of this chapter we shall describe the hot-cathode dis-
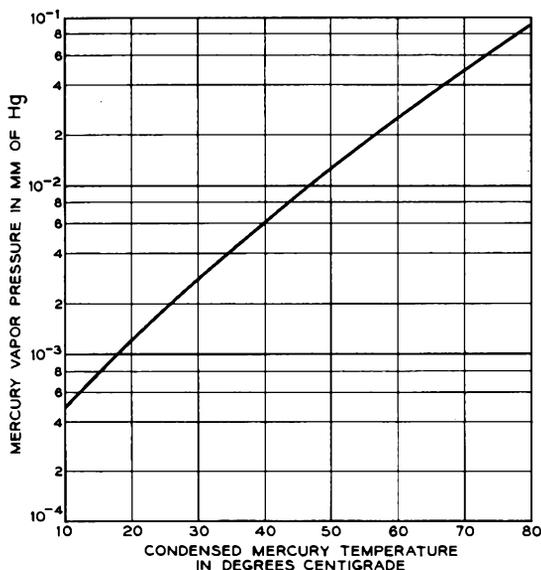


Fig. 16.1-1 Mercury-vapor pressure vs. condensed mercury temperature.

charge in mercury vapor. Later we shall discuss a number of specific types of hot-cathode devices.

## 16.1 The Hot-Cathode Discharge in Mercury Vapor; Plasmas and Sheaths

The hot-cathode discharge in mercury vapor and in other gases is discussed in references 16.1 to 16.11 listed at the end of this chapter.

Like the noble gases, the molecules of mercury vapor are single atoms. Their effective diameter for collision is not far from that of neon or argon molecules, but their mass is about 10 times that of the neon molecule.

Figure 16.1-1 shows the variation of mercury-vapor pressure as a function of the temperature of the condensed mercury. Many hot-cathode, mercury-vapor tubes are rated for condensed mercury temperatures falling in a range between 20°C and 80°C corresponding to vapor pressures of approximately $10^{-3}$ to $10^{-1}$ mm of Hg. If the condensed mercury temperature falls below 20°C, the voltage drop across the tube rises because collisions between plasma electrons and gas molecules become too infrequent. At condensed mercury temperatures appreciably greater than 80°C, the inverse breakdown voltage may fall below the tube rating, since the inverse breakdown voltage decreases rapidly with increasing vapor pressure.

Figure 16.1-2 shows the collision probability $P_c$ for electrons in mercury vapor as a function of the square root of the electron energy.[1] The figure is similar to Fig. 14.3-2. The electron mean free path is given by $L = 1/pP_c$ cm where $p$ is the vapor pressure in millimeters of Hg. At a pressure of $10^{-2}$ mm of Hg, the mean free path of a 10-electron-volt electron in mercury vapor is approximately 1.4 cm, and the mean free path of a 2-electron-volt electron is approximately 0.4 cm.

Figure 16.1-3 shows a schematic illustration of a hot-cathode tube with a

FIG. 16.1-2 The quantity $P_c$ plotted as a function of the square root of electron energy for electrons in mercury vapor. (From R. B. Brode, *Revs. Modern Phys.* 5, 257, 1933)

[1] Reference 16.12.
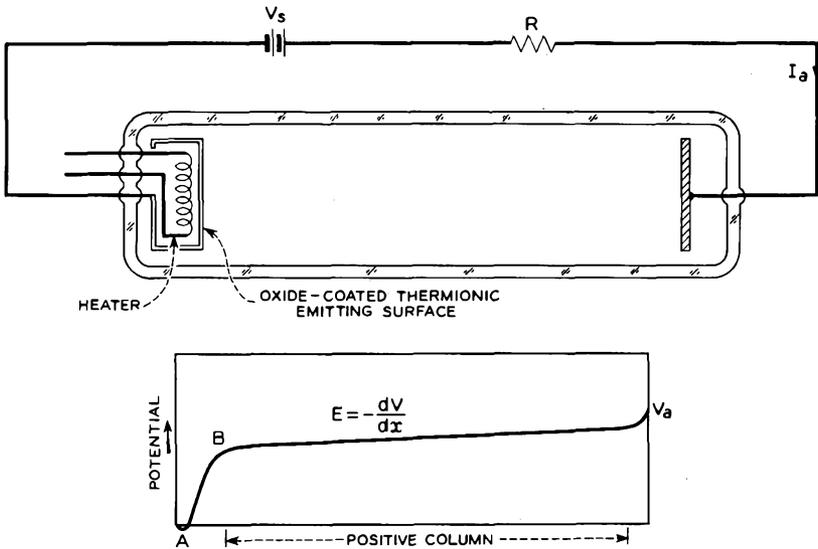
FIG. 16.1-3 A hot-cathode discharge tube with planar electrodes.

planar oxide-coated cathode and a planar anode. A plot of the potential along a line running from cathode to anode is also shown in the figure. Close to the cathode there is a potential minimum, the current drawn from the cathode being space-charge-limited. Just beyond the potential minimum the potential rises steeply, accounting for much of the potential difference between the electrodes. The remainder of the interelectrode space, except for thin sheaths close to the anode and glass walls, is filled with a plasma in which the ion and electron densities are approximately equal. If the length of the tube were increased, the voltage drops close to the cathode and anode would remain essentially unchanged, but the positive column would become longer, and the same potential gradient would continue over its full length. Tubes used in rectifier and switching applications usually have relatively small electrode spacings so that the voltage drop in the plasma will be small.

It will be convenient to describe the plasma and sheaths under separate headings below. Our objective will be twofold: (1) To present orders of magnitude for the principal physical quantities involved, and (2) to make plausible the current-voltage relationship of the hot-cathode discharge.

## (a) The Plasma

If a voltage well below the ionization potential of the gas is applied to a device such as that shown in Figure 16.1-3, only a very small current is drawn from the cathode and practically no ionization takes place. How-

ever, if the applied voltage is raised somewhat above the ionization poten-
tial of the gas, the rate of generation of ions increases rapidly. Both the ions
and electrons tend to diffuse in the direction perpendicular to the axis of the
tube, but because the electrons are very much faster, the walls become
negatively charged, and an excess positive charge accumulates in the in-
terior of the tube. Electrons are drawn toward the region of excess positive
charge with the result that the plasma is formed. Once the plasma is es-
tablished, the voltage drop of the tube generally falls below the value
needed to initiate the discharge. Since the walls are insulating, the electron
current striking the walls in the steady-state condition equals the ion
current.

The approximate uniformity of potential within the plasma can be at-
tributed to the much greater mobility of the electrons compared with that
of the ions. The electrons are drawn toward a region in which the potential
is maximum, and they flow away from a potential minimum. The net result
is that potential gradients within the plasma are small, and the density of
ions and electrons is very nearly equal. The electrons are almost entirely
responsible for conduction through the plasma, the drift velocity of the ions
being extremely small compared with that of the electrons. The fact that
the potential gradients in the direction of conduction are small means that
the plasma is effectively a good conductor.

The electrons cross and recross the plasma many times, experiencing
frequent collisions with neutral molecules, ions, and other electrons. Be-
cause the potential falls away at the edge of the plasma, the electrons, for
the most part, are reflected back toward the plasma by the sheaths. Meas-
urements of the distribution of electron velocities in the plasma indicate
that it is very nearly a Maxwell-Boltzmann distribution. The random nature
of the electron velocities is thought to result from interactions between the
electrons themselves rather than interactions with ions or neutral mole-
cules.[2] Two factors are important here:

1. In electron-molecule or electron-ion collisions, the energy exchange is
extremely small, provided the collisions are elastic, because of the much
greater mass of the molecule or ion. However, in electron-electron inter-
actions, an appreciable energy transfer can take place since the masses of
the colliding particles are equal.

2. The electron-electron interactions result from interactions of the fields
of the particles and, hence, are not of a hard-sphere nature like electron-

---

[2]Another mechanism which may contribute to the randomizing of the electron veloc-
ities in low-pressure discharges is discussed in Reference 15.13. The mechanism in-
volves space-charge oscillations in the sheaths.

molecule or intermolecular collisions. The potential caused by the field of an electron at a distance $r$ angstroms from its center is $14.4/r$ volts, so that the electron-electron interactions can take place at appreciably larger inter-particle distances than electron-molecule or intermolecule interactions. Consequently the "cross section" for electron-electron collisions is much larger than for electron molecule collisions.

The Maxwell-Boltzmann distribution of electron velocities is generally characterized by a relatively high temperature, often one to a few tens of thousands of degrees Kelvin. Measurements of the electron temperature $T_e$ for a positive column in mercury vapor are shown in Figures 16.1-4 and
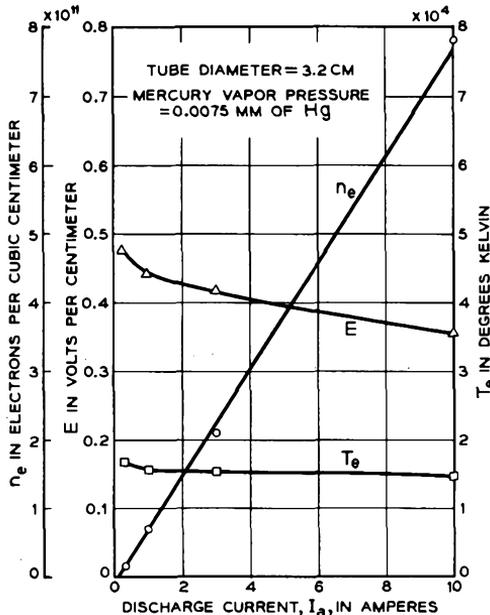


FIG. 16.1-4 The electron density $n_e$, the electron temperature $T_e$, and the axial electric field intensity $E$ vs. the discharge current. (From B. Klarfeld, *Tech. Phys. USSR* **5**, 913, 1938)

16.1-5. The average electron kinetic energy is given by $(\frac{3}{2})kT_e = T_e/7700$ electron volts. For $T_e = 15,000°K$ the average electron kinetic energy is about 1.9 electron volts. It is the few very high-energy electrons in the Maxwell-Boltzmann distribution that are responsible for the excitation and ionization taking place in the plasma. In a mercury vapor discharge at a
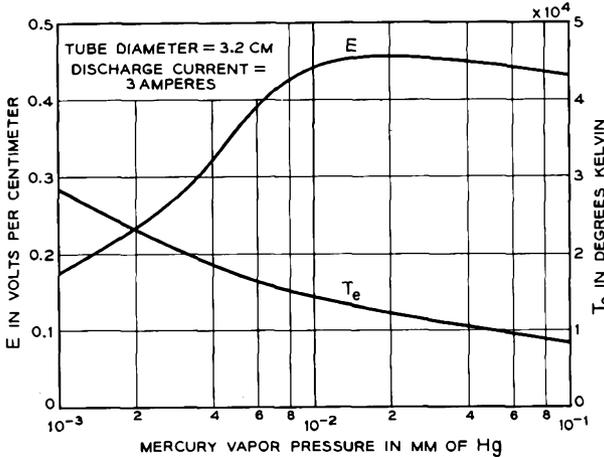
FIG. 16.1-5 The axial electric field intensity $E$ and the electron temperature $T_e$ vs. mercury-vapor pressure for a discharge in mercury vapor. (From B. Klarfeld, *Tech. Phys. USSR* **5**, 913, 1938)

condensed mercury temperature of 40°C, most of the ionization results from two-stage processes in which the molecules are raised to an excited state in a first collision and ionized in a second collision.[3] The excited states principally involved are metastable states with excitation energies of 5.46 and 4.66 electron volts.

Also shown in Figure 16.1-4 are the electron density $n_e$ and the potential gradient $E$ along the axis of the positive column. The electron density $n_e$ is an average of the electron density measured on the axis and at the edge of the plasma. The ratio of $n_e$ to the density of gas molecules is shown in Fig. 16.1-6. For the discharge conditions indicated in the figure only about one molecule in a thousand is ionized.

The ion kinetic energies are much smaller than those of the electrons, the average ion kinetic energy being at most a few times the thermal energy of the gas molecules. Interactions with other ions and neutral molecules lead to a randomizing of the ion velocities and an approximately Maxwell-Boltzmann distribution of velocities.

Ions reaching the edges of the plasma are drawn into the sheath and accelerated toward the walls of the tube, causing a depletion of the ion density toward the edge of the plasma. As a result, the potential along the axis of

---

[3]References 16.5 and 16.6.
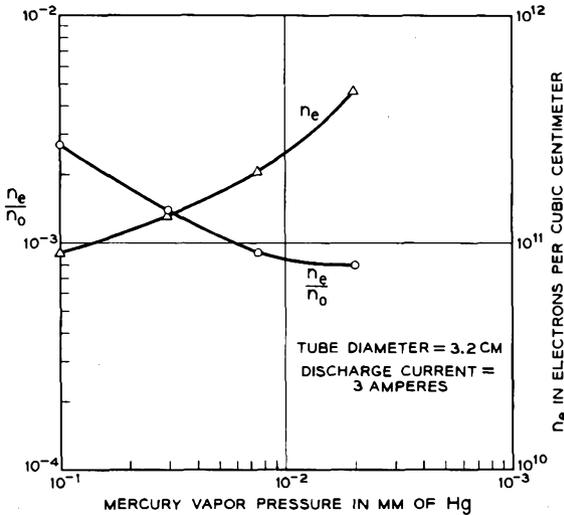
FIG. 16.1-6 The electron density $n_e$ and the ratio of $n_e$ to the density of gas molecules $n_o$ as a function of mercury-vapor pressure. (From B. Klarfeld, *Tech. Phys. USSR* 5, 913, 1938)

the positive column of a device like that shown in Figure 16.1-3 is likely to be several volts higher than at the edges of the plasma, and the radial electric field tends to accelerate the ions toward the sheaths.

It is convenient to regard the plasma ions as being characterized by an average lifetime equal to the average time spent by the ions within the plasma. Over a range of filling pressures $p$ and positive-column radii $R$, the average ion lifetime is a function of the product $pR$, since constant $pR$ means a constant number of mean free paths from the tube axis to the tube walls. High filling pressures and large radii are associated with long lifetimes, whereas low filling pressures and small radii lead to short lifetimes. To a first approximation the average lifetime depends only on the product $pR$ and is independent of the ion density or the current conducted along the positive column. The metastable molecules generated by the discharge are similarly characterized by an average lifetime.

The concept of an average ion lifetime is helpful in describing the balance of events taking place in the plasma:

1. The electron and ion currents to the walls are equal, and the rate of generation of ion-electron pairs per unit length of the positive column equals the current of electrons or ions striking unit length of the glass walls.

2. Since the ion and electron densities are approximately equal, each plasma electron must on the average generate one ion-electron pair once
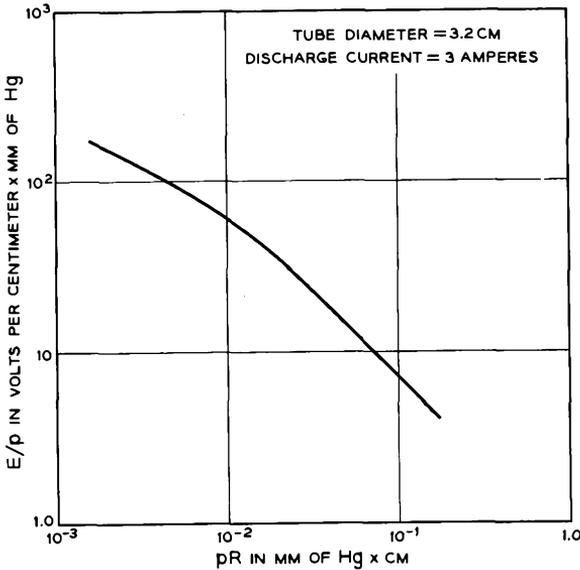
FIG. 16.1-7 $E/p$ vs. $pR$ for a discharge in mercury vapor. (From B. Klarfeld, *Tech. Phys. USSR* **5,** 913, 1938)

every average ion lifetime in order to maintain the supply of ions in the plasma.

In the steady-state condition the electric field intensity $E$ parallel to the axis of the positive column becomes that value which enables the electrons to gain sufficient energy to meet Condition 2 above. The average electron kinetic energy is, of course, a function of $E/p$ rather than $E$ alone. Figures
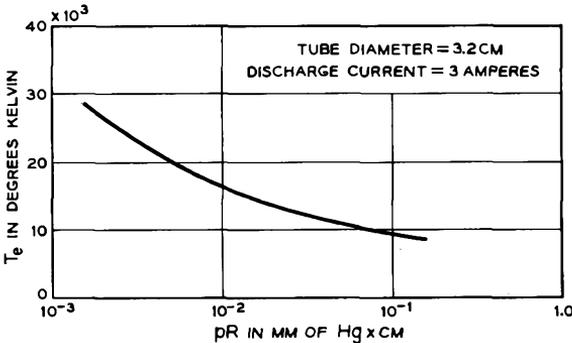


FIG. 16.1-8 The electron temperature $T_e$ vs. $pR$ for a discharge in mercury vapor. (From B. Klarfeld, *Tech. Phys. USSR* **5,** 913, 1938)

16.1-7 and 16.1-8 show plots of $E/p$ and $T_e$ vs. $pR$ for a positive column in mercury vapor. As $pR$ increases, corresponding to longer lifetimes, both $T_e$ and $E/p$ decrease.

The electron drift velocity parallel to the axis of the tube, $\bar{u}_e$, is also determined by $E/p$ and, hence, by Condition 2 above. If the ion lifetimes are short, $E/p$ is large, and $\bar{u}_e$ is large. If $A$ is the cross-sectional area of the positive column, and $n_e$ is the average electron density over the area $A$, the total current conducted by the positive column can be approximately expressed as

$$I_a \approx |\, n_e \bar{u}_e A e \,| \qquad (16.1\text{-}1)$$

where $e$ is the charge on the electron, and where the drift velocity of the ions parallel to the axis of the tube is assumed to be small compared with $\bar{u}_e$. To the extent that $E$, and hence $E/p$, are independent of $I_a$, $\bar{u}_e$ does not change with $I_a$, and *the electron density $n_e$ is directly proportional to $I_a$*, as is evident from Figure 16.1-4. Since the ion density is nearly equal to $n_e$ and since the average ion lifetime is to a first approximation independent of $I_a$, *the rate of generation of ion-electron pairs per unit volume also increases linearly with $I_a$*. We shall make use of this result later when we discuss the nature of the current-voltage relationship for the discharge.

(Actually both $E$ and $T_e$ decrease slightly with increasing $I_a$ [Figure 16.1-4] because the frequency of two-stage ionizing events tends to increase in proportion to the product of $n_e$ and the density of excited atoms.)

The product of $E$ and $I_a$ can be equated to the sum of the power radiated as light energy per unit length of the positive column and the power expended per unit length as heat energy.

As a final point we shall use Equations (14.1-4) and (16.1-1) to determine the ratio of the average electron velocity $u_{\text{avg}}$ to the electron drift velocity $\bar{u}_e$ for the discharge conditions indicated in Figure 16.1-4. From the data given in the figure, $A$ is equal to $(\pi/4) \times (3.2)^2 = 8$ cm$^2$, and $I_a/n_e \cong (\frac{4}{3}) \times 10^{-11}$ amp cm$^3$/electron. From Equation (16.1-1) we find that $\bar{u}_e = 1.04 \times 10^7$ cm/sec. The average electron velocity characteristic of the Maxwell-Boltzmann distribution can be obtained from equation (14.1-4). If $T_e = 15{,}000°$K, $u_{\text{avg}} = 7.6 \times 10^7$ cm/sec, or 7.3 times the electron drift velocity.

(b) *The Sheaths*

In considering the nature of the sheaths it will first be convenient to imagine that a planar conducting probe is inserted part way into a plasma. If the probe is held at a negative potential with respect to the plasma, it becomes surrounded by a thin layer of ions that are drawn from the plasma and accelerated toward the probe. Upon striking the probe, the ions lose

their charge and become unexcited molecules. The thickness of the sheath adjusts itself to account for the potential difference between the probe and plasma. Field lines extend from ions in the sheath to negative charges on the surface of the probe, and the integral of the electric field intensity from the probe to the plasma accounts for the potential drop in the sheath. If the discharge conducts a large current, the rate of arrival of ions at the edge of the plasma is large, and the electric field intensity in the sheath is large. The thickness of the sheath is therefore small. If the discharge current is reduced or if the potential difference between the probe and plasma is increased, the sheath thickness increases. In the case of a probe in a mercury vapor plasma, the sheath thickness is a small fraction of a millimeter except at very small discharge currents or high sheath voltages ($\gg$ 10 volts).
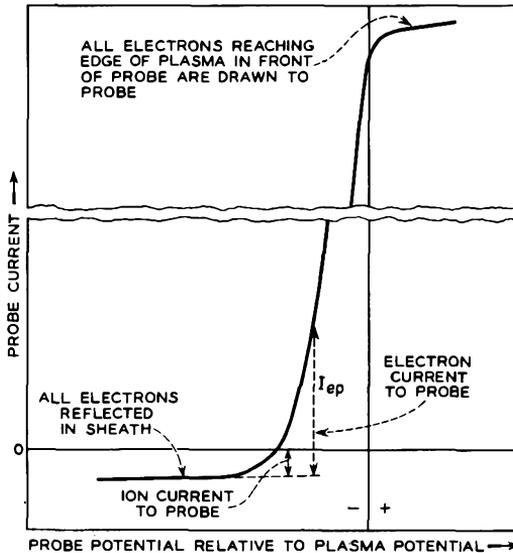


FIG. 16.1-9 Current to a probe in a plasma vs. potential of probe relative to that of the plasma.

Figure 16.1-9 shows qualitatively a plot of current to a probe in a plasma as a function of the probe potential relative to that of the plasma. At high negative probe potentials, the current to the probe is almost independent of the probe potential and equals the ion current drawn into the sheath. However, as the probe potential is made more positive, some of the high-energy electrons reaching the edge of the plasma are able to overcome the retarding field in the sheath and strike the probe. If the electrons have a

Maxwell-Boltzmann distribution of velocities and if the electron mean free path is large compared with the sheath thickness, the fraction of electrons able to overcome a potential $V$ is $\epsilon^{-eV/kT_e}$, where $T_e$ is the electron temperature. The current of electrons reaching the probe can therefore be expressed as

$$I_{ep} = I_e \epsilon^{-eV/kT_e} \qquad (16.1\text{-}2)$$

where $I_e$ is the electron current reaching the edge of the sheath in front of the probe, and $V$ is the voltage drop in the sheath. A plot of the electron current to a probe as a function of the probe voltage is shown in Figure 16.1-10. The linear part of the curve can be expressed as

$$\ln I_{ep} = \alpha - \frac{V}{1.2} \qquad (16.1\text{-}3)$$

where $V$ is in volts, and where $V$ is taken to be the potential rise from the probe to the plasma. The temperature of the plasma electrons is therefore $1.2(e/k) = 1.2 \times 11{,}600 = 13{,}900°K$. When the probe potential



FIG. 16.1-10 Measurements of electron current to a probe in a mercury-vapor plasma vs. probe potential. (From B. Klarfeld, *Tech. Phys. USSR* **4**, 44, 1937)

equals or exceeds the plasma potential, all the electrons reaching the edge of the plasma in front of the probe are drawn to the probe, and the curve of $I_{ep}$ vs. $V$ bends to the right.

The ion current drawn to a probe biased negatively with respect to a plasma has been studied analytically by Allen, Boyd, and Reynolds.[4] Curves given in their paper show the potential distribution within the sheath and near the edge of the plasma.

When a mercury-vapor discharge is established in a device like that shown in Figure 16.1-3, the voltage drop in the sheath between the plasma and the glass walls is of the order of 8 volts.[5] This sheath voltage serves to reduce the current of electrons reaching the walls to a value equal to the ion current.

Between the plasma and the anode of the tube shown in Figure 16.1-3 there is probably a small rise in potential. Ions generated in the sheath

[4]Reference 16.14.
[5]Reference 16.11.

flow toward the positive column and serve to match boundary conditions with respect to ion flow at the anode end of the positive column. (Within the positive column the ion drift velocity has a component parallel to the axis of the tube and directed toward the cathode. There is therefore effectively a flow of ions along the positive column in the direction of the cathode. Ions generated in the anode sheath serve to match this flow.) If the anode is made very small, the voltage rise in the anode sheath may be appreciable, and a region of more intense glow surrounds the anode. In other tubes in which a cylindrical anode surrounds a cylindrical hot cathode, the electron current reaching the edge of the plasma in front of the anode may exceed the current flowing in the external circuit, in which case there is probably a small fall in potential in going from the plasma to the anode.

Because the work function of the anode is usually a few volts greater than that of the cathode, the anode-to-cathode voltage drop of a hot-cathode tube as measured in the external circuit is a few volts greater than it would be if the electrodes were of the same work function.

The cathode sheath represented by the region $AB$ in Figure 16.1-3 is sometimes called a double sheath. Electrons flow from $A$ toward $B$, and ions flow in the opposite direction. Electrons also enter the sheath from the plasma but for the most part are reflected back to the plasma. Field lines extend from ions in the $B$ side of the sheath to electrons in the $A$ side. Since the electric field intensity at $A$ and $B$ is very nearly zero, there is approximately one ion in the sheath for every electron.

Now the velocity of an electron or ion of given kinetic energy is inversely proportional to the square root of the mass of the particle. Consequently, if the potential in the sheath were symmetrical about the midpoint between $A$ and $B$, the ratio of the time required for an ion to travel from $B$ to $A$ to the time required for an electron to travel from $A$ to $B$ would be $(m_i/m_e)^{1/2}$, where $m_i$ is the ion mass, $m_e$ is the electron mass, and where it is assumed that the particles start with zero velocity and experience no collisions on the way. For mercury ions, this ratio is 604. Consequently, the electron current flowing from $A$ to $B$ in a mercury discharge is probably several hundred times the ion current flowing from $B$ to $A$.

The voltage drop in the cathode sheath is such as to assure that the electrons arriving at the plasma from $A$ will have sufficient energy to generate the ions needed to make up for ion losses from the cathode end of the positive column. If the electron and ion currents in the cathode sheath are denoted by $I_e$ and $I_p$, respectively, then $I_e + I_p = I_a$. Since $I_e/I_p$ is to a first approximation determined by factors other than the discharge current $I_a$, both $I_e$ and $I_p$ are directly proportional to $I_a$. And *since the ionization taking place in the plasma is directly proportional to $I_a$, the current*

*$I_p$ of ions arriving at the cathode end of the positive column is proportional to
$I_a$, and the voltage drop in the cathode sheath does not change with $I_a$. Be-
cause the voltage drop in the plasma also shows little change with $I_a$, the voltage
drop in mercury-vapor rectifier tubes is nearly independent of current up to
a current equal to the cathode thermionic emission current.* At higher currents
the anode-to-cathode voltage rises.

The thickness of the cathode sheath varies with the discharge current
$I_a$ and is such as to assure that the necessary electron current is drawn from
the potential minimum at $A$. When $I_a$ is large, the electric field intensity
at the midpoint between $A$ and $B$ is large, and the sheath thickness is small.

## 16.2    Three Examples of Hot-Cathode Tubes

In this section we shall describe three hot-cathode tubes, two used in
rectifier applications and one used as a switching element. The first is the
249B, a two-electrode mercury-vapor tube used in high-voltage rectifier
applications. The second is the 393A, a three-electrode tube filled with
both mercury vapor and argon and used as a rectifier in regulated power
supplies. The third is the 2D21, a four-electrode tube filled with xenon and
used as a switching element.

### The 249B

Figure 16.2-1 illustrates the construction of the 249B. The cathode con-
sists of a nickel mesh filament coated with barium and strontium oxides.
The filament is surrounded by a nickel heat shield which helps to reduce the
power needed to heat the filament and which is connected electrically to
the center of the filament. During conduction a thin sheath forms between
the heat shield and the plasma, with the result that the heat shield has
little effect on conduction through the plasma. The anode consists of a
nickel disc on which carbon has been deposited to increase heat radiation.

During assembly of the tube a small glass pellet containing liquid
mercury is attached to the lower part of the cathode supporting structure
by means of a nickel mesh which completely encloses the pellet and which
is welded to the cathode supporting structure. Ater the tube has been
pumped and before it is sealed off from the pump station, the nickel mesh is
heated with rf causing the glass pellet to soften and admit liquid mercury
to the inside of the tube. The nickel mesh serves to retain the ruptured
glass pellet.

Principal electrical characteristics and ratings for the tube are shown
in Table 16.2-1. The filaments of hot-cathode tubes are designed to operate
at low voltages so that the voltage drop from end to end of the filament
will be small compared with the anode voltage drop. The condensed mer-
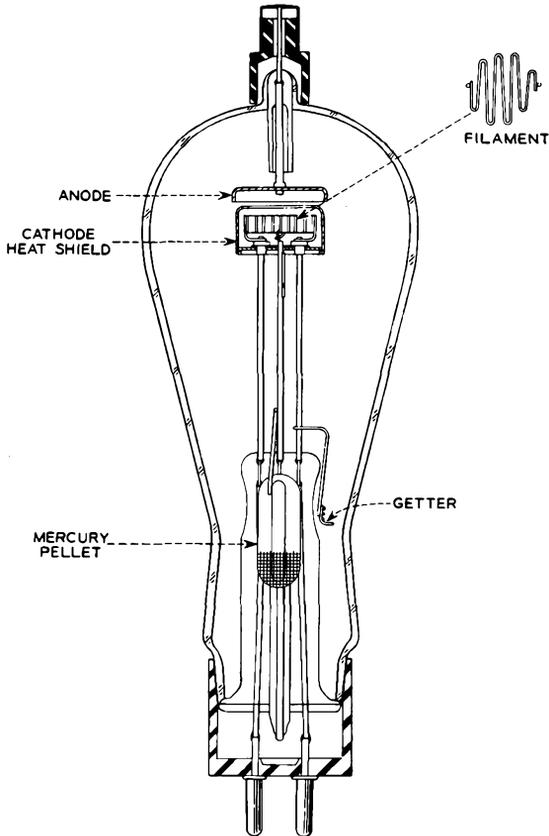
FIG. 16.2-1 The construction of the 249B diode rectifier tube. The overall height of the tube is 19.4 cm.

cury temperature is assumed to be the temperature of the lowest part of the glass envelope, just above the base.

Figure 16.2-2 shows a plot of anode voltage drop vs. anode current for a particular 249B. The plot shows a nearly constant anode voltage drop over a wide range of currents, as would be expected from our earlier discussion. The curve should be compared with the $I_a$-$V_a$ curve for a space-charge-limited vacuum diode in which $I_a \propto V_a^{3/2}$.

Figure 16.2-3 shows a plot of anode voltage drop vs. condensed mercury temperature for a particular 249B. Below a condensed mercury temperature of 20° to 25°C, the anode voltage drop rises because collisions between the electrons and gas molecules become too infrequent. It has
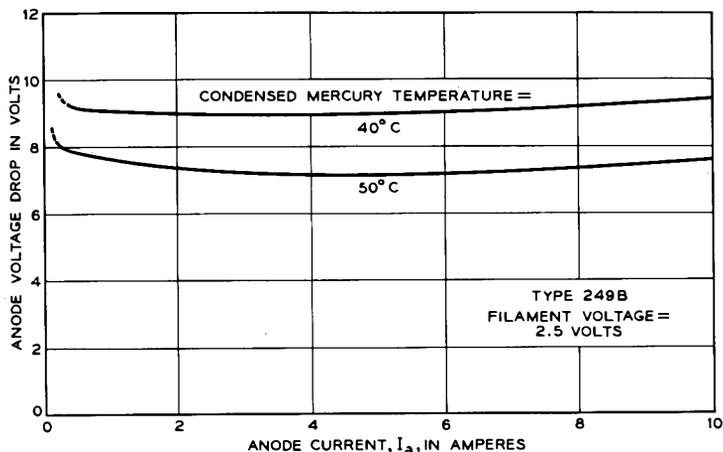
FIG. 16.2-2 Anode voltage drop vs. anode current $I_a$ for a particular 249B.

TABLE 16.2-1 WE 249B OPERATING CHARACTERISTICS

Maximum Ratings

Peak inverse anode voltage, volts............. 7500
Peak anode current, amps.................... 2.5
Average anode current, amps................ 0.64
Condensed mercury temperature, °C.......... +20 to +70

Electrical Data

Anode voltage drop, volts........................7 to 11*
Filament voltage, volts ac.......................2.5
Filament current, amps ac.......................7.5

*Typical values for new tubes at a condensed mercury temperature of 40°C.

been found that the sputtering of the cathode coating in hot-cathode mercury-vapor tubes increases rapidly with increasing anode voltage drop above about 25 volts. Consequently, operation for extended periods at condensed mercury temperatures below 15° or 20°C leads to greatly reduced life.

Other commercially available mercury-vapor diode rectifiers are capable of conducting average currents as high as 20 amps.

*The 393A*

Figure 16.2-4 illustrates the construction of the 393A rectifier tube. The tube has a third electrode, called a grid, which serves to control initiation of the discharge. Hot-cathode tubes with such a control elec-
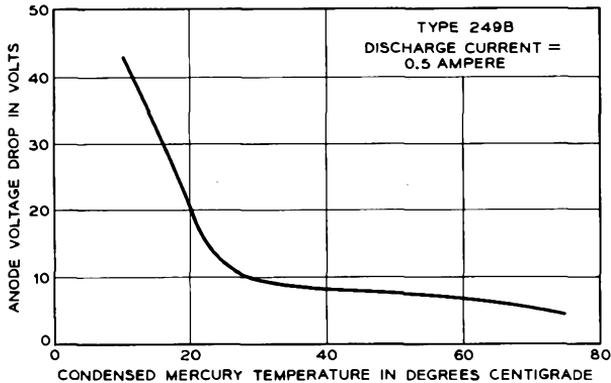
FIG. 16.2-3 Anode voltage drop vs. condensed mercury temperature for a particular 249B.

trode are called thyratrons. Like the 249B, the cathode is a nickel mesh filament coated with barium and strontium oxides, and the anode is a carbonized nickel disc. The grid electrode, also made of carbonized nickel, very nearly surrounds the cathode. The tube is filled with both argon at a pressure of 0.1 mm of Hg and mercury vapor in equilibrium with liquid mercury.

If the grid is biased to about $-10$ volts with respect to the cathode and if the anode voltage is then raised from zero to $+1000$ volts, essentially no anode-to-cathode conduction takes place. The negative bias on the grid in this case is sufficient to cause electrons emitted from the cathode to be returned to the cathode and the breakdown voltage between the anode and grid is considerably in excess of 1000 volts. However, if the grid bias is gradually reduced with the same applied anode voltage, a point is reached, perhaps at a grid bias of $-5$ volts, at which a small current of electrons is drawn through the slot at the upper end of the grid. The electrons passing through the slot enter a region of relatively strong electric field intensity in the grid-anode space and are accelerated toward the anode. On the way some cause ionization. If sufficient current flows to initiate a plasma, the tube quickly transfers to the high conduction state in which the anode voltage drop is of the order of 12 volts, and the anode current is limited by the resistance in series with the supply voltage. In this second state the grid is surrounded by a sheath which accounts for the voltage drop between the grid and plasma and which is thin compared with the width of the slot in the grid. Increasing the negative voltage applied to the grid in this case has little effect on conduction through the tube, except at extremely small anode currents.

Figure 16.2-5 shows a control characteristic measured for a particular
393A (solid curve) at a condensed mercury temperature of 40°C.  Points to
the left of the curve correspond to operating conditions in which the tube
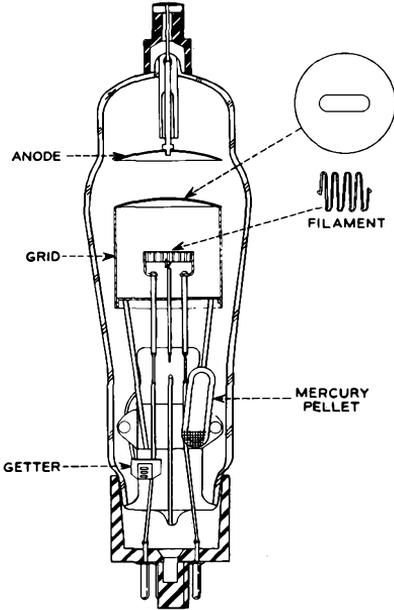


FIG. 16.2-4   The construction of the
393A thyratron.  The overall height
of the tube is 16.8 cm.
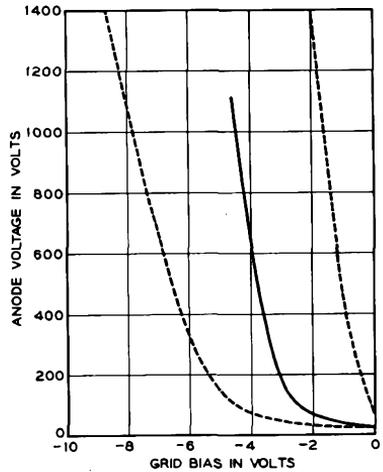


FIG. 16.2-5   Control characteristic
for the 393A.

does not conduct provided the grid bias is applied before the anode voltage.
However, if the grid bias is reduced so that the operating point moves to the
right and crosses the curve shown in the figure, conduction is initiated at the
point of crossing.  The broken curves in the figure show the published
limits for the tube.  At higher condensed mercury temperatures the control
characteristic moves to the left, and at lower condensed mercury tempera-
tures it moves to the right.  The spread between the two broken curves
includes variations within the rated condensed mercury temperatures,
variations during life, and variations from tube to tube.

If the grid potential is raised above that of the plasma, the grid becomes
an anode, and practically the full current of the discharge is drawn to it.
The voltage drop of the tube with the grid tied to the anode is often 6 to 9
volts at a condensed mercury temperature of 40°C.  Figure 16.2-6 shows the
variation of grid current with grid voltage, the grid voltage being measured

relative to the cathode potential. In most applications a resistance is placed in series with the grid to limit the grid current.

During operation of the tube the grid is heated by radiation from both the cathode and anode. Since some of the cathode coating is generally
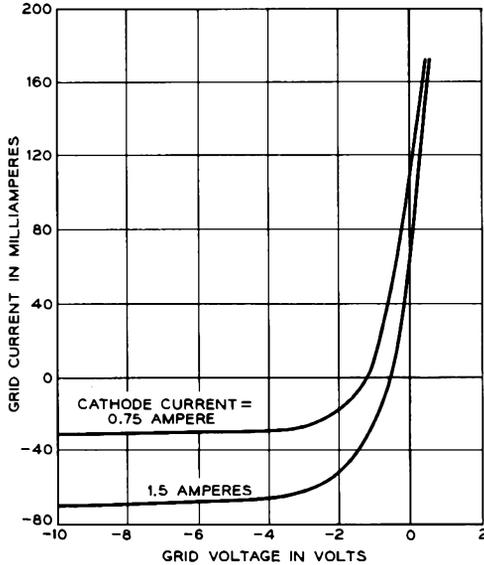


FIG. 16.2-6 Grid current vs. grid voltage for a 393A.

sputtered onto the grid, sufficient heating of the grid will result in electron emission from the grid. Part of the emission current is drawn to the anode and, if the current flow is sufficient, the control action of the grid is lost. The average current handling capabilities of the tube are therefore limited by the need to avoid overheating the grid.

The use of a filling of both argon and mercury vapor permits operation of the tube at ambient temperatures as low as −55°C. If the condensed mercury temperature is below 15° or 20°C when conduction is first initiated, the discharge starts off as an argon discharge, the mercury vapor density being insufficient to support the discharge. Because the ionization potential of argon is higher than that of mercury (15.8 volts compared with 10.4 volts), and perhaps because the lifetimes of the lighter argon ions are shorter than those of the mercury ions, the anode voltage drop of the argon discharge is a few volts higher than that of the mercury discharge, and the electron temperature is similarly higher. As conduction continues, heat is dissipated in the electrodes and conducted and radiated to the envelope. This causes

the condensed mercury temperature to rise and increases the mercury vapor pressure. When the condensed mercury temperature reaches 20° or 25°C, the mercury vapor density is sufficient to support the discharge, and the anode voltage drop falls. Since the electron temperature is lower when the mercury discharge is established, very little ionization of the argon molecules takes place. It is found that cleanup of the argon occurs primarily while the argon discharge is established, and that it is greatly reduced when the mercury discharge takes over.

Principal electrical characteristics and ratings for the 393A are summarized in Table 16.2-2. The condensed mercury temperature limits apply only to starting conditions. The tube must be operated in an environment which will permit the equilibrium condensed mercury temperature to reach at least 20°C, since operation at lower condensed mercury temperatures for prolonged periods leads to cleanup of the argon and shortens the life of the tube.

TABLE 16.2-2 WE 393A OPERATING CHARACTERISTICS

Maximum Ratings
    Peak anode voltage, forward or reverse, volts......1250
    Peak anode current, amps......................6
    Average anode current, amps...................1.5
    Average electron current to grid, ma.............10
    Condensed mercury temperature, °C..........−55 to +80

Electrical Data
    Anode voltage drop, volts......................10 to 14*
    Filament voltage, volts ac.....................2.5
    Filament current, amps ac.....................7.0

*Typical values for new tubes when the grid is held at a negative potential with respect to the plasma, and when the condensed mercury temperature is 40°C.

The deionization time or recovery time of a thyratron is the time required for the grid to regain control of the discharge after conduction has been stopped by removing the anode voltage. Often several hundred microseconds elapse before the ion and electron densities in the interelectrode space become sufficiently small that the grid regains control. The deionization time increases with the vapor density, with the discharge current before interruption, and with decreasing grid bias. At a condensed mercury temperature of 40°C and an anode current of 1.5 amps, the deionization time of the 393A is of the order of a few hundred microseconds.

Other commercially available mercury-vapor thyratrons conduct average currents as high as 30 amps. Inverse anode voltages of some mercury-vapor thyratrons range as high as 15,000 volts.
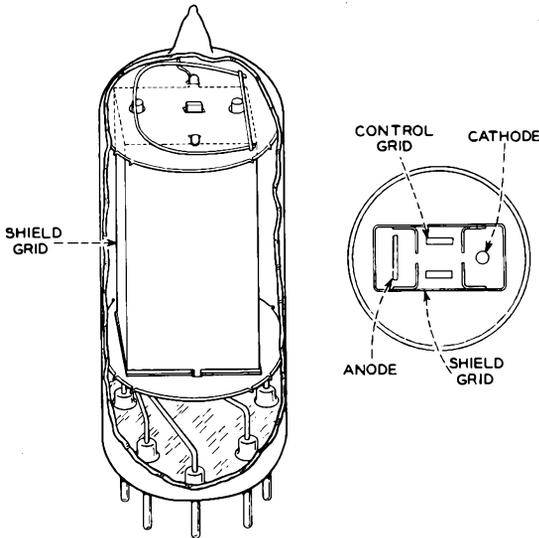
FIG. 16.2-7 The construction of the 2D21 shield-grid thyratron. The overall height of the tube is 5.4 cm.

## The 2D21

Figure 16.2-7 shows the construction of the 2D21 shield-grid thyratron. The tube has an indirectly heated cathode, a control grid, an anode, and an additional electrode called a shield grid. The shield grid virtually surrounds the other electrodes, and baffles in the shield grid serve to separate the remaining electrodes from one another. The tube is filled with xenon to a pressure of 0.16 mm of Hg.

The use of a shield grid has several advantages:

1. It reduces heat radiation to the control grid from the cathode and anode, and it reduces the amount of cathode coating sputtered onto the control grid. Grid emission is thus greatly reduced.

2. The shield grid also reduces capacitive coupling between the grid and anode. Without the shield grid or other circuit protection, the grid-anode capacitance of a thyratron is sometimes sufficient to cause the grid to be driven positive by a fluctuation or surge of anode voltage, thereby permitting faulty initiation of the discharge.

3. The shield grid provides an additional means for controlling initiation of the discharge. Figure 16.2-8 shows the control characteristic of the tube for several values of shield-grid voltage. By making the shield grid a few volts positive or negative with respect to the cathode, the control char-
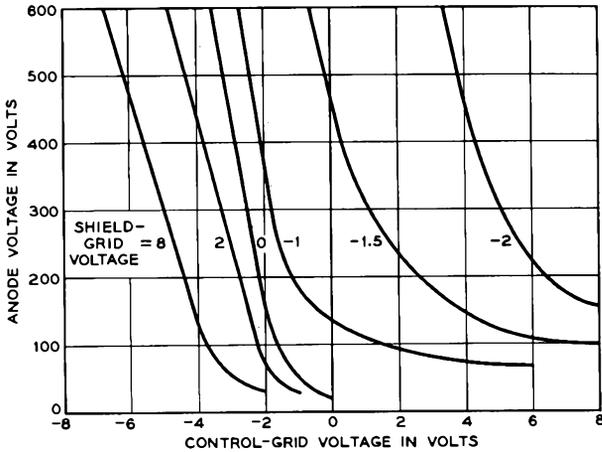
FIG. 16.2-8 Control characteristic for the 2D21 for several values of shield-grid voltage.

acteristic can be moved to the left or right, respectively. The shield grid can thus be used to switch the grid control circuit in and out of operation.

The 2D21 is principally used in switching applications. Because the grid emission and grid-anode capacitance are both small, the tube can be driven by high impedance sources.

Principal electrical characteristics and ratings for the 2D21 are listed in Table 16.2-3.

TABLE 16.2-3 2D21 OPERATING CHARACTERISTICS

Maximum Ratings

Peak inverse anode voltage, volts................1300
Peak anode current, ma.........................500
Average anode current, ma......................100
Ambient temperature, °C....................−75 to +90

Electrical Data

Anode voltage drop, volts.......................8
Heater voltage, volts ac .......................6.3
Heater current, amp ac .........................0.6

## 16.3   Other Types of Hot-Cathode Tubes

### (a)   The Tungar

This is a hot-cathode diode frequently used as a half-wave rectifier in storage-battery chargers. The tube is filled with argon to a pressure be-

tween 50 and 100 mm of Hg and has a thoriated tungsten filament. The filament operates at about 2200°C, or about 200°C higher than thoriated tungsten filaments used in transmitting tubes. The higher filament operating temperature is made possible by the high argon filling pressure, which reduces the rate of evaporation of thorium atoms from the cathode. Tungars typically conduct average currents of 2 to 15 amps. The filament voltage is often 2.0 or 2.5 volts, and the filament current may be as high as 15 to 25 amps. The high argon pressure leads to relatively low peak inverse voltages, of the order of 300 volts.

### (b)  Hydrogen Thyratrons

A class of hydrogen filled thyratrons has been developed for use in power supplies which deliver driving power to magnetrons and klystrons in radar applications. The hydrogen thyratron is generally connected in series with a pulse transformer across a charged delay line, and the microwave tube is connected to the output of the pulse transformer. A positive pulse applied to the control grid of the thyratron serves to discharge the delay line through the transformer and thyratron. Between periods of conduction the delay line is recharged with energy for the next pulse. Pulse currents as high as 1 to 3 thousand amps are conducted by some hydrogen thyratrons.

The use of a hydrogen filling leads to shorter deionization times than can be obtained with other fillings.[6] Deionization times of hydrogen thyratrons are typically 2 to 10 microsec with some very high-current tubes ranging as high as 50 microsec. The short deionization times permit operation of the magnetron or klystron at pulse rates of several kilocycles.

The anode voltage drops of hydrogen thyratrons are much higher than those of mercury-vapor or noble-gas filled thyratrons, often between 90 and 150 volts. The relatively high voltage drops result partly from the short lifetimes of the hydrogen ions, and partly from the fact that most of the inelastic collisions between plasma electrons and gas molecules result in excitation of the vibrational and rotational states of the molecules and in disassociation of the molecules rather than excitation of electronic states or ionization.

Hydrogen thyratrons almost always have indirectly heated, oxide-coated, nickel cathodes. The filling pressure frequently lies between 0.2 and 1.0 mm of Hg. Because gas cleanup is appreciable in high-current tubes, the filling is often obtained by means of a "hydrogen reservoir" consisting of a

---

[6](1) Because the lighter ion mass leads to shorter ion lifetimes in the decaying plasma at the end of a pulse; (2) it is thought that recombination of ions and electrons in the decaying plasma may be appreciable; (3) there are no hydrogen metastables. Probably (1) is the most important reason.

quantity of titanium hydride in contact with an auxiliary heater. When heated, the titanium hydride evolves hydrogen until the surrounding gas reaches an equilibrium pressure at which the rate of evolution of hydrogen equals the rate of reabsorption. Higher reservoir temperatures lead to higher equilibrium pressures, and when the reservoir heater is turned off, the gas filling is reabsorbed, and the tube pumps down to very nearly a vacuum.

### (c)  Tubes with Ionically Heated Cathodes

A number of gas tubes used for illumination purposes have cathodes that are heated to thermionic emitting temperatures by the current of ions incident upon the cathode rather than by passing a current through the cathode. The discharge in this case is in many respects similar to the thermionic arc. Familiar examples of tubes using such ionically heated cathodes are the fluorescent lamp and mercury-vapor lamps used for street-lighting. These tubes frequently have two ionically heated cathodes, one at either end of a discharge tube, and are operated from an ac supply.

The cathodes consist of a coil of fine tungsten wire coated with oxides suitable for thermionic emission and wound on a somewhat heavier tungsten support wire. Often the coil is itself wound into a larger coil, and sometimes this in turn is coiled once more giving a "triple coil" structure. Usually the discharge heats only a portion of the cathode to thermionic emitting temperatures. Sometimes the heated region slowly moves along the cathode, and other times it remains stationary. The multiple-coil structure provides a large amount of cathode area with closely controlled thermal and electrical resistance to the current carrying support wire. Starting the discharge is accomplished by passing a current through the support wire for a short period or by applying a sufficiently high voltage between the electrodes that the discharge starts as a cold-cathode discharge.

Fluorescent lamps usually contain argon or krypton at a pressure of about 3 mm of Hg and mercury vapor in equilibrium with liquid mercury. The noble-gas filling aids in starting the discharge and serves to increase the lifetimes of ions and metastables once the discharge is established. However the noble-gas ions do not otherwise take part in the discharge once it is established. The discharge converts as much as 60 per cent of the input power to ultraviolet radiation of wavelength 2537 angstroms. Phosphors coated on the inside of the bulb serve to transform part of the ultraviolet radiation to visible light, hence the term "fluorescent."

Mercury-vapor lamps generate visible light directly, the efficiency being increased by operation at high mercury-vapor pressures, often several atmospheres.

More detailed descriptions of gas-discharge lamps are given in References 16g, 16h, and 16i.

## REFERENCES

Several texts on gas-discharge phenomena are listed at the end of Chapter 14. The performance of hot-cathode tubes, particularly diode rectifiers and thyratrons, is discussed in the following references:

16a. W. G. Dow, *Fundamentals of Engineering Electronics*, Chapter 18, John Wiley and Sons, Inc., New York, 1952.

16b. J. Millman and S. Seely, *Electronics*, 2nd Ed., Chapter 11, McGraw-Hill Book Co., Inc., New York, 1951.

16c. D. S. Peck, *Electrical Engineers Handbook*, 4th Ed. (H. Pender and K. Mc-Ilwain, Eds.) pp. 4–58ff., John Wiley and Sons, Inc., New York, 1950.

16d. D. E. Marshall, *Industrial Electronics Reference Book* (Westinghouse Electric Corp.), Chapter 6, John Wiley and Sons, Inc., New York, 1948.

16e. J. D. Cobine, *Gaseous Conductors*, Chapter 11, McGraw-Hill Book Co., Inc., New York, 1941.

16f. F. A. Maxfield and R. R. Benedict, *Theory of Gaseous Conduction and Electronics*, Chapter 11, McGraw-Hill Book Co., Inc., New York, 1941.

Gas-discharge lamps are described in the following references:

16g. W. E. Forsythe and E. Q. Adams, *Fluorescent and Other Gas Discharge Lamps*, McGraw-Hill Book Co., Inc., New York, 1948.

16h. J. Millman and S. Seely, *Electronics*, 2nd Ed. pp. 324ff., McGraw-Hill Book Co., Inc., New York, 1951.

16i. *General Electric Glow Lamp Manual*, General Electric Co., East Cleveland, Ohio.

The hot-cathode discharge in mercury vapor is described in the following references:

16.1. I. Langmuir, *Phys. Rev.* **33,** 954, 1929.

16.2. I. Langmuir, *J. Franklin Inst.* **275,** 1932.

16.3. T. J. Killian, *Phys. Rev.* **35,** 1238, 1930.

16.4. B. Klarfeld, *Tech. Phys. USSR* **4,** 44, 1937.

16.5. B. Klarfeld, *Tech. Phys. USSR* **5,** 913, 1938.

16.6. R. M. Howe, *J. Appl. Phys.* **24,** 881, 1953.

The hot-cathode discharge in noble gases at higher pressures of the order of a few mm of Hg is described in the following references:

16.7. M. J. Druyvesteyn and F. M. Penning, *Rev. Modern Phys.* **12,** 87, 1940.

16.8. L. Malter, E. O. Johnson, and W. M. Webster, *RCA Rev.* **12,** 415, 1951.

16.9. W. M. Webster, E. O. Johnson, and L. Malter, *RCA Rev.* **13,** 163, 1952.

The discharge in mixtures of mercury vapor and a noble gas is described in the following references:

16.10. C. Kenty, *J. Appl. Phys.* **21,** 1309, 1950.

16.11. J. F. Waymouth and F. Bitter, *J. Appl. Phys.* **27,** 122, 1956.

Other references covering specific subjects discussed in this chapter are:

16.12. R. P. Brode, *Rev. Modern Phys.* **5,** 257, 1933.

16.13. D. Gabor, E. A. Ash, and D. Dracott, *Nature* **176,** 916, 1955.

16.14. J. E. Allen, R. L. F. Boyd, and P. Reynolds, *Proc. Phys. Soc. (London)*, B, **70,** 297, 1957.

# Chapter 17

## GAS LASERS

A relatively recent development in the field of gas-discharge devices is the gas laser. This device generates a highly directional beam of coherent light with perhaps a few milliwatts of power and with a line width, or frequency stability, which in principle can be made as small as a few cycles per second or a few tens of cycles per second. Before the invention of the laser, all available light sources, such as incandescent lights, fluorescent lights, and gas discharge sources, provided only incoherent light, that is, light in which the photons are emitted at random instants of time and with random phase. The light generated by the laser is a sinusoidal electromagnetic wave with a frequency stability which is determined largely by the dimensional stability of the device.

The brightness of a gas laser beam is enormously greater than that of non-laser light sources. Thus the sun, which radiates much like a black body with a temperature of 6000°K, has a total radiation at all wavelengths of about seven kilowatts per square centimeter of its surface. However, if it were possible to filter out a narrow band of light one kilocycle wide in the region of the spectrum where the sun has its peak output (at a wavelength of 4800 angstroms), the total power radiated within this narrow band would be only $10^{-8}$ watt per square centimeter of the sun's surface. The radiation per unit solid angle in the direction normal to the sun's surface would be $1/\pi$ times this power (see Equation (2.4-13)). In contrast, a gas laser might generate a beam with a few milliwatts of power in a band which in principle could be only a few cycles per second wide. The angular spread of the beam can be made less than one minute of arc. The gas laser there-fore provides an entirely new and powerful tool in the field of optics, and it may well find a number of applications in the communications field.

The word "laser" is made up from the first letters of the words *light amplification by stimulated emission of radiation*. Stimulated emission refers to an interaction between an excited atom and an electromagnetic field in

which the excited atom undergoes a transition to a state of lower energy and imparts the lost energy to the electromagnetic field. The process is the reverse of the more familiar phenomenon of light absorption by matter and, in fact, involves precisely the same physical concepts. For stimulated emission to occur, the frequency of the electromagnetic wave multiplied by Planck's constant $h$ must equal the energy given up by the excited atom. Consider an atom which is excited to a state from which it can decay to a lower state by spontaneous emission of a photon $h\nu$. If an electromagnetic field is established in the region of the atom and if the frequency of the field is $\nu$, the atom can decay to the lower state either by spontaneous emission of a photon $h\nu$ or by imparting the same amount of energy to the energy stored in the electromagnetic field. The latter process is called *stimulated emission*.

Suppose an electromagnetic wave propagates through a gas in which a fraction of the atoms are in either of two excited states, one with an energy $h\nu$ above the other and such that atoms in the upper state can decay to the lower state by spontaneous emission of a photon $h\nu$. The electromagnetic field causes some of the atoms in the upper state to decay to the lower state by stimulated emission. Similarly atoms in the lower state can *absorb* an amount of energy $h\nu$ from the electromagnetic field and become excited to the higher state. The probability per unit time of an atom in the lower state absorbing an amount of energy $h\nu$ and becoming excited to the higher state is equal to the probability per unit time of an atom in the upper state being stimulated by the field to undergo a transition to the lower state. This probability is proportional to the square of the field intensity of the electromagnetic wave.

In a noble gas at room temperature essentially all the atoms are in their lowest, or ground, state. Excitation to higher states can occur in several ways, for example, by establishing a dc discharge in the gas, by applying a high-frequency rf field of sufficient intensity to maintain a glow discharge in the region of the gas, or by irradiating the gas with light of a suitable wavelength. Suppose that by one of these mechanisms two excited states 1 and 2 are populated so that there are densities $n_1$ and $n_2$ atoms in these states per cubic centimeter of the gas. Suppose further that state 2 is of higher excitation energy than state 1, that atoms of state 2 can decay to state 1 by spontaneous emission of a photon $h\nu$, and that the conditions of the discharge are such that[1] $n_2 > n_1$. Then an electromagnetic wave of fre-

---

[1]As discussed subsequently in Section 17.1, atoms in a given excited state are characterized by a total angular momentum which may have several possible orientations relative to a particular direction of observation. These individual possible orientations of the total angular momentum are called magnetic substates. In the absence of a magnetic field, all magnetic substates have the same energy, and all are equally populated. In the above discussion the quantities $n_1$ and $n_2$ must be taken to be the populations of the individual magnetic substates of excited states 1 and 2.

quency $\nu$ passing through the gas will cause more atoms per unit volume and per unit time in state 2 to decay to state 1 by stimulated emission than it will cause atoms in state 1 to absorb $h\nu$ from the field and become excited to state 2. *There will therefore be a net addition of energy to the electromagnetic wave, and the wave is amplified in passing through the gas.*

Suppose next that optical reflecting surfaces are placed at either end of the discharge tube. If the surfaces are of such a spacing that there is an integral number of half wavelengths at the frequency $\nu$ between them, a standing-wave pattern can be established in which the waves are reflected forward and back across the discharge. Such a reflecting system for light waves is called a Fabry-Perot interferometer and is the optical equivalent of a cavity resonator for microwaves. When a standing wave pattern of wavelength $c/\nu$ is established in the region of the discharge, and when $n_2 > n_1$, energy is added to the electromagnetic wave by the discharge. The power level then builds up until the upper state is sufficiently depopulated by the electromagnetic wave that an equilibrium is reached between power added to the wave by gas discharge and power lost in reflection, diffraction, and absorption in the optical system.

Such a system is effectively an oscillator of frequency $\nu$ and is called a laser. If one of the reflectors transmits a small portion of the incident light ($\sim$1 per cent) rather than totally reflecting it, the transmitted beam is a sinusoidal electromagnetic wave of frequency $\nu$ and of frequency stability determined largely by the stability of the optical path length between the reflecting surfaces. Gas laser action has been observed to occur in a number of gases and between a number of pairs of excited states of these gases (see Table 17.3-1). The wavelengths corresponding to some of the transitions are in the visible spectrum, whereas others are in the infrared or ultraviolet regions.

If the light output from a laser is passed through a second discharge tube with the same gas filling and discharge conditions, so that $n_2 > n_1$, the light is *amplified* in passing through the discharge.

The first gas laser[2] used a gas filling consisting of a mixture of helium and neon. It happens that the excitation energies of two helium metastable states[3] very nearly coincide with the excitation energy of two excited states in neon. When a dc or rf discharge is established in the gas mixture, a high density of helium metastables is generated. These metastables collide with unexcited neon atoms causing excitation of the neon atoms to the excited neon states of nearly equal energy and causing the helium atoms to return

---

[2] Reference 17.1.

[3] A metastable state is an excited state from which spontaneous decay to a lower state is forbidden by the quantum-mechanical selection rules for electric-dipole transitions. (This does not necessarily preclude de-excitation by other processes or different types of transitions having a much smaller probability per unit time.)

to their ground state. The excited neon atoms can be stimulated to decay
to a number of lower excited states, permitting laser action at a number of
wavelengths ranging from the visible to the infrared region.

In the present chapter we shall first discuss the excited energy levels of
atoms, particularly helium and neon, and we shall describe the balance of
events occurring in laser action between two excited states of an atom.
Then we shall describe a particular helium-neon laser and its performance,
and we shall summarize briefly the range of performance and operating
conditions presently obtained with other lasers.

## 17.1    Energy Levels in the Atoms

From quantum mechanics we learn that there are only certain discrete
energies and angular momenta which the electrons that are bound to an
atom can have. An atom in its ground state, or unexcited state, has all its
electrons in the lowest possible energy states. As discussed in Section 2.1,
for many purposes an electron in a given energy state can be thought of as
causing a cloud of charge about the nucleus. The probability of finding an
electron in a given volume element about the nucleus is proportional to the
density of the cloud at the volume element. It is found that electrons in
certain energy states tend to contribute their maximum charge density at
approximately the same distance from the nucleus, and consequently there
are said to be shells of electrons about the nucleus. Helium in its ground
state has two electrons in a single shell about the nucleus. Neon in its
ground state has two shells of electrons with two electrons in the inner
shell and eight electrons in the outer shell. Argon in its ground state has
three shells with two electrons in the inner shell, eight electrons in the next
shell, and eight electrons in the outer shell.

Because electrons in the innermost shell of an atom are on the average
closer to the nucleus than electrons in the next shell, greater energy is re-
quired to remove these electrons from the atom. Consequently electrons in
the innermost shell are said to be in states of lower energy than electrons in
the next shell. Similarly, in atoms with three or more shells, electrons in the
second shell are more tightly bound to the atom than electrons in the third
shell, and so on. However, an experimental law of physics, known as the
Pauli exclusion principle, prohibits all the electrons in the atom from going
into the states of lowest energy and hence into the shell which is closest to
the nucleus. Only two electrons can go into the inner shell, whereas up to
eight electrons can go into the second shell, and up to eighteen can go into
the third shell.

The fact that there are shells of electrons about the nucleus is directly
related to the quantization of the electron motions. It is customary in the

field of spectroscopy to describe an electronic state by means of two quantum numbers $n$ and $l$. The quantum number $n$ is related to the average radial distance of the electron orbit from the nucleus and hence to the shell that the electron is in. The number $n$ may have integral values 1, 2, 3 . . . . An electron in a state for which $n = 1$ is in the innermost shell of the atom, an electron in a state for which $n = 2$ is in the next shell, and so on. The quantum number $l$ indicates the oribital angular momentum of the electron about the nucleus. This angular momentum is given by $\sqrt{l(l+1)}\ h/2\pi$, where $h$ is Planck's constant. $l$ can have values 0, 1, 2, . . . $n - 1$. The atomic spectroscopists further use the letters $s$, $p$, $d$, and $f$ to correspond to $l = 0, 1, 2,$ and 3. Thus the electronic configuration of an unexcited helium atom is given by $(1s)^2$, where $1s$ indicates a state with $n = 1$ and $l = 0$, and the superscript 2 indicates that two electrons are in this state. The electronic configuration of the ground state of neon is written as $(1s)^2\ (2s)^2\ (2p)^6$, indicating that there are two electrons in $1s$ ($n = 1$, $l = 0$) states, two electrons in $2s$ ($n = 2$, $l = 0$) states, and 6 electrons in $2p$ ($n = 2$, $l = 1$) states. An excited neon atom might have one of the $2p$ electrons raised to a $3s$ level The electronic configuration for the atom in this case would be $(1s)^2\ (2s)^2\ (2p)^5\ 3s$. The excited electron might also be raised to any of the following states: $3p$, $3d$, $4s$, $4p$, $4d$, $4f$, $5s$, $5p$, etc. Radiative transitions can occur only between states whose values of $l$ differ by $\pm\ 1$.

The energy levels are not determined uniquely by the quantum numbers $n$ and $l$. Electrons also have an intrinsic spin angular momentum, and interactions between this angular momentum and the orbital angular momenta lead to a splitting of the excitation energies into closely spaced groups of states. Finally, each of these states is characterized by a total angular momentum which is a vector sum of the spin and orbital angular momenta of all the electrons of the atom. The total angular momentum is characterized by a quantum number $J$ such that the total angular momentum is equal to $\sqrt{J(J+1)}\ h/2\pi$ , where $h$ is Planck's constant. If there are an even number of electrons in the atom, $J$ is an integer, and if there are an odd number of electrons in the atom, $J$ has a value equal to one half an integer. The total angular momentum can have $2J + 1$ possible components in any direction of observation. For example, if $J$ is 2, the total angular momentum along a direction of observation can be $-2h/2\pi$, $-1h/2\pi$, 0, $+1h/2\pi$, or $+2h/2\pi$. These individual possible orientations of the total angular momentum are referred to as *magnetic substates*. In an applied magnetic field, the energies of the magnetic substates are separated by an amount proportional to the magnetic field intensity. In the absence of a magnetic field, all magnetic substates have the same energy, and all are equally populated. In the introductory part of this chapter we referred to population densities $n_1$ and $n_2$ for two excited states, 1 and 2. As noted[1], these population densities must be

taken to mean the density of atoms in any one magnetic substate of excited states 1 and 2. Thus, if $N_1$ is the *total* density of atoms in state 1 and if there is no externally applied magnetic field, $n_1 = N_1/(2J + 1)$.

It is not possible to make an exact quantum mechanical calculation of the energy levels in an atom with more than one electron because the forces acting on an individual electron result not only from the nucleus but also from all the other electrons, and the motion of each electron affects the motions of all the other electrons. However, calculations involving a number of approximations have been made, and the energy levels can be measured with a high degree of accuracy by spectroscopic observations.

Figure 17.1-1 shows a graphical presentation of the energy levels of the lower excited states in helium and neon. It is assumed in this plot that only one electron is excited. Each excited level in neon is actually broken into a group of closely spaced states which result from different possible orientations of the electron spin and orbital angular momenta. The lowest excited states in neon are four $3s$ states in the neighborhood of 16.7 electron volts. An excited atom in one of these states has the configuration $(1s)^2 (2s)^2 (2p)^5$
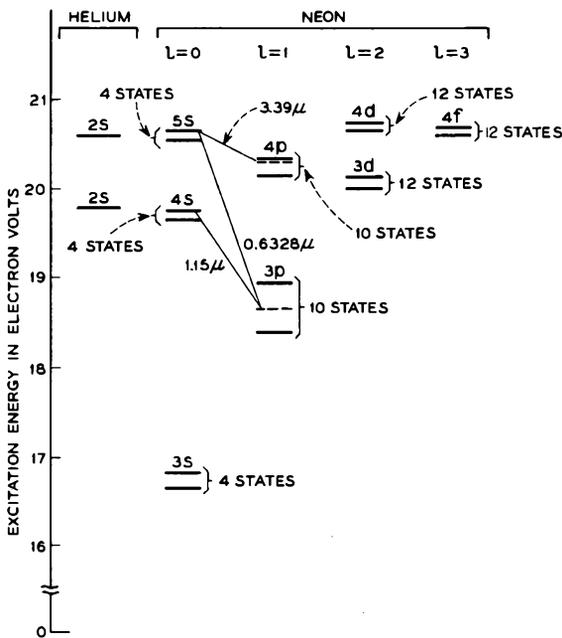


FIG. 17.1-1 The lower excited states in helium and neon. The excitation energy is measured relative to the ground state.

3s. The splitting of the 3s state into four states results from different orientations of the spin angular momenta relative to the orbital angular momenta. Each of the four 3s states is characterized by a particular value of the total angular momentum quantum number $J$. This in turn leads to $2J + 1$ magnetic substates for each of the four 3s states. The next higher states are ten 3p states in the neighborhood of 18.7 electron volts. The electronic configuration for these states is $(1s)^2 (2s)^2 (2p)^5 3p$.

The lowest excited states in helium are two 2s states which differ in energy because the electron spin angular momenta are parallel in one case (the state of lower energy) and opposed in the other case. Both these states are metastable. Their excitation energies very nearly coincide with those of the 4s and 5s states in neon. In the operation of the helium-neon laser, the helium 2s metastables are destroyed largely by collisions with ground-state neon atoms. In these collisions the excitation energy of the metastable helium atom is transferred to a neon atom, thereby exciting it to the 4s or 5s state. Laser action can then occur in neon between 5s and 4p states, between 5s and 3p states, and between 4s and 3p states. Three wavelengths frequently observed with helium-neon lasers are shown in Figure 17.1-1 The wavelengths are measured in microns (symbol $\mu$). $1\mu = 10^{-6}$ meter = $10^4$ angstroms.

## 17.2  Operation of a Laser

It will be helpful to express in equation form the balance of events that occur in laser action between two excited states of an atom. Our objective will be to derive an expression for the power output of light from one end of the laser in terms of certain parameters of the discharge. Figure 17.2-1 shows an energy level diagram in which laser action is assumed to take place between two excited states, 1 and 2. We shall use the following quantities:

$n_1$ and $n_2$ are the population densities, or numbers of atoms per cubic centimeter, in states 1 and 2.

$\dot{n}_1$ and $\dot{n}_2$ are the time rates of change of the population densities $n_1$ and $n_2$.

$S_1$ and $S_2$ are the rates at which atoms in states 1 and 2 are produced by the discharge per cubic centimeter of gas. $S_1$ is assumed to account for all means of generating atoms in state 1 except spontaneous decay of atoms in state 2 to state 1. We shall consider spontaneous decay from state 2 to state 1 separately.

$A_1$ and $A_2$ are the probabilities per unit time of atoms in states 1 and 2 decaying by spontaneous emission.

$A_{21}$ is the probability per unit time of an atom in state 2 decaying to state 1 by spontaneous emission.
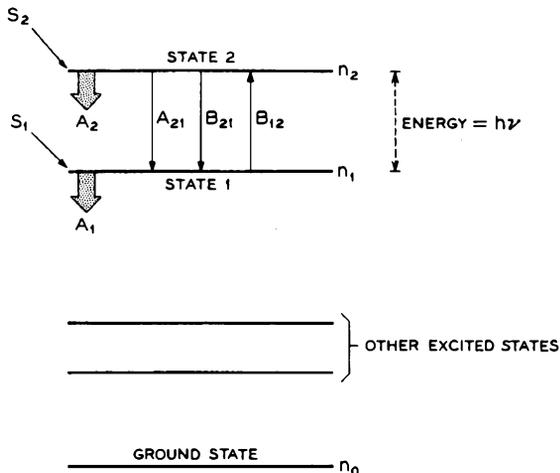
FIG. 17.2-1 An energy level diagram showing the excitation energies of a hypothetical atom measured relative to the ground state.

$P$ is the electromagnetic power crossing a square centimeter of a plane normal to the direction of propagation of the wave.

$B_{12}$ and $B_{21}$ are the stimulated emission coefficients for transitions from states 1 to state 2 and for transitions from state 2 to state 1. The coefficient $B_{12}$ is such that $B_{12}P$ is the probability per unit time that an atom in state 1 will be stimulated to absorb energy $h\nu$ from the electromagnetic field and become excited to state 2. $B_{21}P$ is the probability per unit time that an atom in state 2 will be stimulated by the electromagnetic field to decay to state 1. The coefficients $B_{12}$ and $B_{21}$ are equal.

The quantities $A_1$, $A_2$, $A_{21}$, $B_{12}$, and $B_{21}$ are functions of the electronic configuration of the individual states.

The time rates of change of the population densities $n_1$ and $n_2$ are given by

$$\dot{n}_1 = S_1 + n_2(A_{21} + B_{21}P) - n_1(A_1 + B_{12}P) \qquad (17.2\text{-}1)$$

and

$$\dot{n}_2 = S_2 - n_2(A_2 + B_{21}P) + n_1 B_{12}P \qquad (17.2\text{-}2)$$

In the steady-state condition,

$$\dot{n}_1 = \dot{n}_2 = 0 \qquad (17.2\text{-}3)$$

Combining the above three equations and solving for $n_2 - n_1$, we obtain

$$n_2 - n_1 = \frac{S_2(A_1 - A_{21}) - S_1 A_2}{A_1 A_2 + B_{21}P(A_1 + A_2 - A_{21})} \qquad (17.2\text{-}4)$$

where we have made use of the equality $B_{12} = B_{21}$.

The power $P$ increases with distance $x$ in the direction of propagation if $n_2 > n_1$ and decreases if $n_2 < n_1$. We can write

$$\frac{dP}{dx} = h\nu B_{21} P(n_2 - n_1) = \frac{\alpha P}{1 + \eta P} \qquad (17.2\text{-}5)$$

where

$$\alpha = \frac{h\nu B_{21}}{A_1 A_2}[S_2(A_1 - A_{21}) - S_1 A_2] \qquad (17.2\text{-}6)$$

and

$$\eta = B_{21}\left(\frac{1}{A_1} + \frac{1}{A_2} - \frac{A_{21}}{A_1 A_2}\right) \qquad (17.2\text{-}7)$$

In the small-signal case in which a weak electromagnetic wave passes through the gas, $\eta P \ll 1$, and $dP/dx = \alpha P$, or

$$P = P_o \epsilon^{\alpha x} \qquad (17.2\text{-}8)$$

where $P_o$ is the electromagnetic power crossing a square centimeter of a plane normal to the direction of propagation at $x = 0$. The power of the electromagnetic wave therefore builds up exponentially with distance in the direction of propagation. However, as $\eta P$ approaches unity, $n_2 - n_1$ decreases, and $dP/dx$ is no longer proportional to $P$.

Suppose that reflecting mirrors are placed at opposite ends of the discharge tube and that a fraction $t$ of the light incident upon the mirrors is transmitted through the mirrors. We shall assume that $t$ is small compared with unity ($\sim$1 per cent) and that a fraction $1 - t$ of the light incident upon the mirrors is reflected without loss. The particular laser we are considering therefore generates a beam of equal intensity from both ends of the discharge tube. In the steady-state condition, the standing wave field pattern between the mirrors can be resolved into two traveling waves, one traveling to the left and one traveling to the right between the mirrors. The power gain of one of these waves in traveling the distance $L$ between the mirrors must equal the power lost in reflection, and hence the power transmitted through the mirror. The power transmitted through the mirror is $tP$ per square centimeter of the mirror, where $P$ is the light power incident upon a square centimeter of the mirror. The quantity $tP$ is small compared with $P$, so that the power gain in traveling the distance between mirrors is also small compared with $P$. Accordingly we can use Equation (17.2-5) to express this power gain as

$$\frac{dP}{dx}L = \frac{\alpha P L}{1 + \eta P} \qquad (17.2\text{-}9)$$

Equating the right-hand side of the equation to $tP$ and solving for $tP$, we obtain

$$tP = \frac{t}{\eta}\left(\frac{\alpha L}{t} - 1\right)$$  (17.2-10)

This expression gives the light power output per square centimeter from one end of the laser in terms of the quantities $\alpha$, $\eta$, and $t$. The quantities $\alpha$ and $\eta$ are related to $S_1$, $S_2$, $A_1$, $A_2$, and $A_{21}$ through Equations (17.2-6) and (17.2-7)

## 17.3  A Helium-Neon Gas Laser

Figure 17.3-1 shows the construction of a particular helium-neon gas laser. The helium-neon gas mixture is contained within a discharge tube which has windows at either end and which has a relatively long positive-
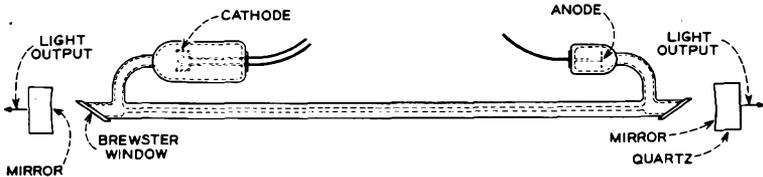


FIG. 17.3-1 The construction of a helium-neon laser.

column region. The cathode and anode are mounted to one side of the tube axis. Precision mirrors coated to reflect red light are mounted beyond the windows at the ends of the discharge tube in such a manner that they reflect light forward and back along the tube. The mirrors are only partially reflecting ($\sim$99 per cent), so that a small amount of light ($\sim$1 per cent) is transmitted through them. The transmitted light provides the power output of the device. The light output is visible and has a wavelength of 0.6328 micron.

The discharge tube is filled with a mixture of helium and neon, consisting of seven parts of helium mixed with one part of neon. The filling pressure is 1.8 mm of Hg. The inside diameter of the long glass tube is 2 mm, and the length of the straight portion of the tube is 30 cm. The discharge is operated at a current of approximately 10 ma, which corresponds to an anode-to-cathode voltage drop of 1300 volts. The cathode consists of a directly heated nickel mesh with an oxide coating. The anode is made of nickel.

At both ends of the discharge tube there are fused quartz windows, called Brewster windows. Each window is inclined relative to the tube axis such that the normal to the window makes an angle of 55.5 degrees with the tube axis. At this angle the reflection of 0.6328-$\mu$ light polarized in a plane

containing the tube axis and the normal to the window is a minimum. The angle is known as the Brewster angle[4]. The light transmission through the windows for this wavelength and polarization is approximately 99.9 per cent for each passage through a window. The transmission of light polarized in a direction perpendicular to the plane containing the tube axis and the normal to the windows is much less, and consequently laser action does not take place for such light. The windows are about 2 mm thick and polished on both faces to a flatness of better than 800 angstroms.

Each mirror consists of a highly polished fused quartz surface onto which is evaporated alternate layers of a material of high index of refraction (ZnS) followed by a material of low index of refraction (ThOF$_2$). Each layer is a quarter wavelength thick at 0.6328 $\mu$, the outer layer being of the material of a high index of refraction. Reflections from each interface beneath the surface add in phase to the wave reflected from the outer surface and increase the total light reflection. The mirrors used in the laser shown in Figure 17.3-1 consist of 7 layers of ZnS and six layers of ThOF$_2$. Such a mirror transmits about 0.7 per cent of the incident light, whereas less than 0.5 per cent is absorbed or scattered, and the remainder is reflected. The surface of the mirrors is slightly concave with a radius of curvature of 3 meters. The curvature serves to focus the reflected light in the direction of the opposite mirror and reduces the loss of radiation by diffraction effects.

Light is reflected between the mirrors at either end of the discharge tube, and for wavelengths at which the optical distance between the mirrors is an integral number of half wavelengths, a standing-wave pattern is established. The two mirrors therefore serve as a sort of cavity resonator. When such a system is constructed using plane mirrors, it is called a *Fabry-Perot interferometer*. At the resonant wavelengths, successive reflections from the mirrors add in phase. If the frequency of one or more of the resonant modes lies within the frequency range of a neon transition between two excited states, and if the upper state has a sufficient excess population over the lower state, laser action can occur.

Let $L$ be the optical spacing between the mirrors. The resonant frequencies of the Fabry-Perot interferometer are such that

$$\frac{2L}{\lambda} = \frac{2L}{c}\nu = n \qquad (17.3\text{-}1)$$

or

$$\nu = \frac{nc}{2L} \qquad (17.3\text{-}2)$$

[4]The Brewster angle $\theta(\lambda)$ for a given wavelength of radiation $\lambda$ is such that $\tan \theta(\lambda)$ equals the index of refraction of the window material for the wavelength $\lambda$.

where $n$ is an integer, $\nu$ and $\lambda$ are the frequency and wavelength of the radiation, and $c$ is the velocity of light. In the laser shown in Figure 17.3-1 the distance $L$ is approximately 37 cm. With such a spacing between planar mirrors, the resonant frequencies are separated by $c/2L = 400$ Mc.

The frequency stability of the laser shown in Figure 17.3-1 is determined principally by the stability of the optical path $L$ between the mirrors. Differentiating Equation (17.3-2) with respect to $L$, we obtain

$$d\nu = -\frac{ncdL}{2L^2} = -\frac{\nu}{L}dL \qquad (17.3\text{-}3)$$

or

$$\frac{d\nu}{\nu} = -\frac{dL}{L} \qquad (17.3\text{-}4)$$

For a change in optical path $dL = 1$ ansgtrom $= 10^{-8}$ cm (about a quarter of an atomic diameter) in a length $L$ of 37 cm, we obtain $d\nu/\nu = 10^{-8}/37 = 2.8 \times 10^{-10}$. Frequency stabilities of this magnitude or better are attainable using various frequency control schemes or by careful isolation of the laser from all mechanical disturbances. In principle it should be possible to stabilize the frequency to within a range of a few cycles per second. On the other hand, the frequency can be modulated in accordance with Equation (17.3-4) by attaching one of the mirrors to an electromechanical transducer which causes the distance $L$ to vary in response to an applied electrical signal.

*Operation of the Laser Shown in Figure 17.3-1*

At a discharge current of 10 ma, both helium and neon ions take part in the current conduction, and the anode-to-cathode voltage drop is found to be intermediate between that of a pure neon discharge with the same partial pressure and a pure helium discharge with the same partial pressure. The discharge leads to the generation of a relatively high density of helium metastables in the two helium metastable states ($\sim 3 \times 10^{11}/$cm$^3$). The metastables are formed by electron collisions with helium atoms in which the atom is excited either directly to the metastable level or to a higher level from which it decays by one or more radiative transitions to the metastable level.

The metastables in turn diffuse through the gas mixture until they are destroyed by one of the following processes: (1) collisions with neon atoms in which the neon atom is excited to a $4s$ or $5s$ state, (2) collisions with the walls of the discharge tube, and (3) collisions with free electrons in which the metastable is excited to a higher state. Primarily the metastables are lost by the first of these processes, and consequently neon atoms are excited to the $4s$ and $5s$ levels at a relatively high rate.

In the absence of an electromagnetic field, excited neon atoms in a $5s$ state decay to $4p$ and $3p$ states by spontaneous emission of a photon. Atoms in $4p$ states in turn decay to $3d$, $4s$, $3s$ states, or the ground state, by spontaneous emission of a photon, and atoms in $3p$ states decay to $3s$ states or the ground state by spontaneous emission of a photon.[5] The probability per unit time of an excited atom in a $4p$ or $3p$ state decaying by spontaneous emission of a photon is sufficiently large that in the absence of an electromagnetic field the population of excited neon atoms in $4p$ or $3p$ states is less than the population of excited neon atoms in $5s$ states. The laser shown in Figure 17.3-1 actually exhibits laser action simultaneously between the $5s$ and $3p$ levels in neon leading to the generation of $0.6328$-$\mu$ light, and between the $5s$ and $4p$ levels leading to the generation of $3.39$-$\mu$ light.

In the emission spectrum of neon the half-power width of the $0.6328$-$\mu$ line is found to be approximately 1.5 Gc, expressed as a frequency. This width results primarily from the well-known Doppler effect. In the discussion following Equation (17.3-2) it was noted that the resonant frequencies of the cavity illustrated in Figure 17.3-1 are separated by 400 Mc. Thus there are approximately three resonant frequencies within the width of the $0.6328$-$\mu$ line. In consequence of this, the laser illustrated in Figure 17.3-1 exhibits laser action at three closely spaced wavelengths within the Dopper width of the $0.6328$-$\mu$ line. From Equation (17.3-2) it is evident that the frequency spacing of these "modes" of oscillation increases as the optical distance $L$ between the mirrors is reduced. Thus lasers have been constructed with about one third the length of the device shown in Figure 17.3-1 which give laser action in only one mode. In these devices the frequency spacing between modes is approximately equal to the Doppler width of the $0.6328$-$\mu$ line.

The quantity $\alpha$ in Equation (17.2-8) has been measured for $0.6328$-$\mu$ light and for the discharge conditions of the device illustrated in Figure 17.3-1. The frequency of the light was at the center of the Doppler width of the line. The quantity $\alpha$ was found to be approximately 0.14 per meter, corresponding to a gain of about 15 per cent per meter of travel through the gas. The corresponding value of $n_2 - n_1$ for small $P$ has been calculated to be[6] $3.9 \times 10^9$ per cm³. For the $5s - 4p$ transition leading to the generation of $3.39$-$\mu$ light, $\alpha$ has been measured in a discharge tube with an inside diameter of 6 mm and found to be $\geqslant 4$ per meter, corresponding to a power gain of $\geqslant 52$ per meter. The corresponding value of $n_2 - n_1$ has been calculated to be $\geqslant 9 \times 10^8$ for small $P$.

---

[5]Not all $3p$ and $4p$ states have quantum-mechanically allowed transitions to the ground state.

[6]Reference 17.2.

One way to prevent laser action at a particular wavelength while maintaining it at another wavelength is to use prisms between the Brewster windows and the mirrors, as illustrated in Figure 17.3-2(a). By orienting the prisms at an appropriate angle, operation at the desired wavelength can be obtained. The function of the prism and mirrors also can be combined, as shown in Figure 17.3-2(b).



(a)

(b)   DIELECTRIC
      COATING
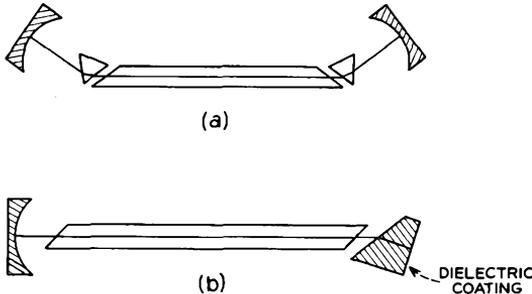
Fig. 17.3-2 By the use of prisms, the laser action can be limited to a single wavelength.

The total light power output of the laser shown in Figure 17.2-1 for the three modes in the neighborhood of 0.6328-$\mu$ is approximately 3 mw at a discharge current of 10 ma. Many design parameters besides the length $L$ of the discharge tube affect this power output. These include: the diameter of the discharge tube, the filling pressure, the ratio of helium to neon, as well as the discharge current. However consideration of these parameters is beyond the scope of the present discussion.

The beam spread of the light transmitted through the mirrors is of the order of several minutes of arc. However this spread can be reduced with the aid of a lens to approximately $\lambda/d$, where $d$ is the diameter of the discharge tube. For $\lambda = 0.6328 \times 10^{-4}$ cm and $d = 0.2$ cm, $\lambda/d$ is approximately one minute of arc.

## 17.4   Other Gas Lasers

For laser action to occur in a gas one must establish an "inverted population" for two excited states between which spontaneous emission can occur. By "inverted population" we mean that the magnetic substate population of the upper state must exceed that of the lower state. In the helium-neon laser the $5s$ neon state is populated largely by excitation of neon atoms to the $5s$ state in collisions with helium metastables.

Inverted populations also can be obtained in dc discharges in the pure noble gases as well as in molecular gases. Finally, inverted populations have been obtained in cesium by irradiating a tube containing cesium vapor with light from an adjacent helium discharge tube. It turns out that the photon energy of one of the stronger spectral lines observed in the helium discharge falls within the Doppler width of the excitation energy of a cesium state, and by irradiating the cesium with this helium light, excitation of the cesium atoms occurs and laser action for two transitions has so far been observed.

In Table 17.4-1 we have attempted to summarize the numbers of lines in the various gases for which laser action, either pulsed or continuous, has been observed. The minimum and maximum wavelengths for these

TABLE 17.4-1

| Gas | Number of Lines Observed | Range of Wavelengths, Microns |
|-----|--------------------------|-------------------------------|
| He | 1 | 2.06 |
| Ne | 117 | 0.59–133 |
| A | 31 | 1.62–26.9 |
| $A^+$ | 10 | 0.45–0.53 |
| Kr | 25 | 1.69–7.06 |
| Xe | 27 | 2.03–18.5 |
| O | 1 | 0.85 |
| C | 2 | 1.07–1.45 |
| N | 2 | 1.36–1.45 |
| $N_2$ | Several tens of transitions | 0.3–1.2 |
| S | 2 | 1.05–1.06 |
| Hg | 2 | 1.53–1.81 |
| Cl | 2 | 1.97–2.02 |
| Br | 4 | 0.85 |
| I | 2 | 3.24–3.43 |
| Cs | 2 | 3.20–7.18 |
| CO | 20 | 0.56–0.66 |
| $H_2O$ | 9 | 23.3–78.8 |
| $CO_2$ | 21 | 9–11 |

lines for each gas is also listed. At the time of writing, investigation of laser action in gases is proceeding at such a rapid rate that the listing will be out of date by the time this textbook has been printed. However the table will serve to indicate the extent of the published investigations as of early 1964.

The lines observed in $A^+$ are transitions between excited states of singly ionized argon. Laser action has been observed for atomic nitrogen (N) transitions and for molecular nitrogen ($N_2$) transitions. It turns out that the lower state for the $N_2$ and CO transitions is metastable, so that as laser

action proceeds, the population of the lower state increases, and the inverted population soon vanishes. Laser action is therefore observed using a pulsed discharge and is found to occur for only a fraction of a microsecond at the start of the pulse.

## REFERENCES

A more extensive description of gas laser principles is given in the following reference:

17a. W. R. Bennett, Jr., *Appl. Optics*, Supplement 1: Optical Masers, p. 24, 1962.
17b. Bela A. Lengyel, *Lasers*, John Wiley and Sons, Inc., New York, 1962.

Other references covering specific topics discussed in this chapter are given below.
17.1 A. Javan, W. R. Bennett, Jr., and D. R. Herriott, *Phys. Rev. Letters* 6, 106, 1961.
17.2 W. L. Faust and R. A. McFarlane, private communication.

# APPENDIX I

## MKS UNITS AND DIMENSIONS

| Quantity | Unit | Dimensions |
|---|---|---|
| Length | Meter | $L$ |
| Mass | Kilogram | $M$ |
| Time | Second | $T$ |
| Charge | Coulomb | $Q$ |
| Velocity | Meter /second | $LT^{-1}$ |
| Force | Newton | $LMT^{-2}$ |
| Energy, work | Joule | $L^2MT^{-2}$ |
| Power | Watt | $L^2MT^{-3}$ |
| Electric potential | Volt | $L^2MT^{-2}Q^{-1}$ |
| Electric field intensity $E$ | Volt /meter | $LMT^{-2}Q^{-1}$ |
| Electric flux | Coulomb | $Q$ |
| Electric flux density $D$ | Coulomb /square meter | $L^{-2}Q$ |
| Capacity | Farad | $L^{-2}M^{-1}T^2Q^2$ |
| $\varepsilon_o$ (permittivity of free space) | Farad /meter | $L^{-3}M^{-1}T^2Q^2$ |
| Electric current | Ampere | $T^{-1}Q$ |
| Magnetic field intensity $H$ | Ampere /meter | $L^{-1}T^{-1}Q$ |
| Magnetic flux | Weber | $L^2MT^{-1}Q^{-1}$ |
| Magnetic flux density | Weber /square meter | $MT^{-1}Q^{-1}$ |
| Inductance | Henry | $L^2MQ^{-2}$ |
| $\mu_o$(permeability of free space) | Henry /meter | $LMQ^{-2}$ |
| Electric resistance | Ohm | $L^2MT^{-1}Q^{-2}$ |

# APPENDIX II

## TABLE OF PHYSICAL CONSTANTS* AND CONVERSION BETWEEN UNITS

| | | |
|---|---|---|
| Permittivity of free space | $\varepsilon_o = 8.854 \times 10^{-12}$ | farad /meter |
| Permeability of free space | $\mu_o = 4\pi \times 10^{-7}$ | henry /meter |
| Velocity of light | $c = 2.998 \times 10^8$ | meters /second |
| Charge of the electron | $e = 1.602 \times 10^{-19}$ | coulomb |
| Mass of the electron | $m = 9.108 \times 10^{-31}$ | kilogram |
| Ratio: electron charge /electron mass | $e/m = 1.759 \times 10^{11}$ | coulomb /kilogram |
| Boltzmann's constant | $k = 1.380 \times 10^{-23}$ | joule /degree |
| Planck's constant | $h = 6.625 \times 10^{-34}$ | joule-second |

### Conversion Between Units

$$1 \text{ angstrom (A)} = 10^{-10} \text{ meter}$$
$$1 \text{ micron } (\mu) = 10^{-6} \text{ meter}$$
$$1 \text{ gauss} = 10^{-4} \text{ weber /square meter}$$
$$1 \text{ oersted} = (1/4\pi) \times 10^3 \text{ amperes /meter}$$
$$1 \text{ pound} = 0.4536 \text{ kilogram}$$
$$1 \text{ liter} = 1000 \text{ cm}^3$$
$$1 \text{ torr} = 1 \text{ mm of Hg pressure}$$
$$= 13.595 \text{ kilograms /square meter}$$
$$1 \text{ electron volt} = 1.602 \times 10^{-19} \text{ joule}$$

---

*For adjusted best values of the physical constants as of 1955, see E. R. Cohen, K. M. Crowe, and J. W. M. Dumond, *Fundamental Constants of Physics*, Interscience Publishers, Inc., New York, 1957.

# APPENDIX III

## SOME RELATIONSHIPS PERTAINING TO ELECTRIC AND MAGNETIC FIELDS AND CURRENT FLOW

(a) STATIC FIELDS

| Electric Fields | Magnetic Fields |
|---|---|
| Electric field intensity . . . . . . . . . . . **E**, $E$ | Magnetic field intensity . . . . . . . . **H**, $H$ |
| Electric flux density . . . . . . . . . . . . . **D**, $D$ | Magnetic flux density . . . . . . . . . . . **B**, $B$ |
| Electric potential . . . . . . . . . . . . . . . $V$ | Magnetic potential . . . . . . . . . . . . . $\psi$ |
| Permittivity of free space . . . . . . . . $\varepsilon_o$ | Permeability of free space . . . . . . . . $\mu_o$ |
| Relative dielectric constant . . . . . . . $\varepsilon$ | Relative permeability . . . . . . . . . . . $\mu$ |
| Space charge density . . . . . . . . . . . . . $\rho$ | Current density vector . . . . . . . . . . **J** |
| Total charge . . . . . . . . . . . . . . . . . . . . $q$ | Total current . . . . . . . . . . . . . . . . . . . $I$ |

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

$$\int_{\text{closed surface}} \mathbf{D}\cdot\mathbf{n}dS = \int_{\text{volume}} \rho dv = q$$

* $$\mathbf{B} = \mu\mu_o\mathbf{H}$$

$$\oint_{\text{closed path}} \mathbf{H}\cdot d\mathbf{l} = I$$

$$\nabla \times \mathbf{H} = \mathbf{J}$$

* $$\nabla\cdot\mathbf{D} = \rho$$

*Energy stored per unit volume $= \frac{1}{2}\varepsilon\varepsilon_o E^2$

* $$\nabla\cdot\mathbf{B} = 0$$

*Energy stored per unit volume $= \frac{1}{2}\mu\mu_o H^2$

In the absence of current-carrying conductors:

$$V_{AB} = -\int_A^B \mathbf{E}\cdot d\mathbf{l}$$

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

$$\psi_{AB} = -\int_A^B \mathbf{H}\cdot d\mathbf{l}$$

$$\mathbf{H} = -\nabla\psi$$

In a region of uniform dielectric constant where there is no distributed charge density:

$$\nabla^2 V = 0$$

In a region of uniform permeability *and in the absence of current carrying conductors:*

$$\nabla^2\psi = 0$$

In the presence of a distributed charge density $\rho$

$$\nabla^2 V = -\frac{\rho}{\varepsilon_o}$$

*At the interface between two dielectric materials, the normal component of **D** and the tangential component of **E** are continuous. (It is assumed here that there are no surface charges.)

*At the interface between two magnetic materials, the normal component of **B** and the tangential component of **H** are continuous. (It is assumed here that there are no surface currents flowing at the interface.)

---

*These relations also apply to time-varying fields.

## (b) TIME-VARYING FIELDS

Maxwell's Equations are:

$$\nabla \times \boldsymbol{\mathcal{E}} = -\frac{\partial \boldsymbol{\mathcal{B}}}{\partial t}$$

$$\nabla \times \boldsymbol{\mathcal{K}} = \boldsymbol{\mathcal{J}} + \frac{\partial \boldsymbol{\mathcal{D}}}{\partial t}$$

$$\nabla \cdot \boldsymbol{\mathcal{D}} = \rho$$

$$\nabla \cdot \boldsymbol{\mathcal{B}} = 0$$

Script letters are used here to indicate that the field components are time varying. If the field components vary with a single angular frequency $\omega$, we can set

$$\boldsymbol{\mathcal{E}} = \text{Re}\mathbf{E}\epsilon^{j\omega t}, \qquad \boldsymbol{\mathcal{B}} = \text{Re}\mathbf{B}\epsilon^{j\omega t}, \text{ etc.},$$

and

$$\frac{\partial \boldsymbol{\mathcal{B}}}{\partial t} = \text{Re}j\omega\mathbf{B}\epsilon^{j\omega t}, \text{ etc.}$$

Then

$$\nabla \times \mathbf{E} = -j\omega\mathbf{B}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + j\omega\mathbf{D}$$

$$\nabla \cdot \mathbf{D} = \rho$$

$$\nabla \cdot \mathbf{B} = 0$$

## (c) CURRENT FLOW

Ohm's Law can be written as

$$\mathbf{J} = \sigma\mathbf{E}$$

or

$$V = IR$$

where $\sigma$ is the conductivity of the medium, and $R$ is the resistance between the terminals where the voltage $V$ is measured.

The equation of continuity can be written as

$$\nabla \cdot \boldsymbol{\mathcal{J}} = -\frac{\partial \rho}{\partial t}$$

If the current density $\boldsymbol{\mathcal{J}}$ is time varying at a single angular frequency $\omega$, we can set $\boldsymbol{\mathcal{J}} = \text{Re}\mathbf{J}\epsilon^{j\omega t}$. Then

$$\nabla \cdot \mathbf{J} = -j\omega\rho$$

*A SUMMARY OF RELATIONS PERTAINING TO THE VELOCITY DIS-
TRIBUTION, ENERGY DISTRIBUTION, AND ANGULAR DISTRIBU-
TION OF THE ELECTRONS EMITTED FROM A THERMIONIC CATHODE*

*Symbols:*

$u_n$ = velocity of an emitted electron in the direction normal to the emitting surface in meters per second.

$u_t$ = velocity of an emitted electron in the "transverse" direction, or parallel to the emitting surface, in meters per second.

$u$ = total emission velocity in meters per second.

$W_n$ = kinetic energy of an emitted electron in the direction normal to the emitting surface in electon volts.

$W_t$ = kinetic energy of an emitted electron in the "transverse" direction, or parallel to the emitting surface, in electron volts.

$W$ = total emission energy in electron volts.

$k$ = Boltzmann's constant.

$T$ = absolute temperature of emitting surface in degrees Kelvin.

$W_T$ = electron-volt equivalent of $kT$.

$\theta$ = direction of emission velocity relative to the normal to the surface.

$|e|$ = a dimensionless positive constant numerically equal to the charge on the electron.

$m$ = mass of the electron.

$J_o$ = total emission current density in amperes/meter².

SOME RELATIONSHIPS BETWEEN THE ABOVE QUANTITIES:

$$W_n = \frac{mu_n^2}{2|e|}; \qquad W_t = \frac{mu_t^2}{2|e|}; \qquad W = W_n + W_t = \frac{mu^2}{2|e|}$$

$$W_T = \frac{kT}{|e|} = \frac{T}{11,600}\text{electron volts.}$$

THE DISTRIBUTION FUNCTIONS:

The probability that an electron is emitted with a component of velocity normal to the cathode surface in the range $u_n$ to $u_n + du_n$ is

$$dP(u_n) = \frac{mu_n}{kT}\epsilon^{-mu_n^2/2kT}du_n \qquad (1)$$

The probability that an electron is emitted with a component of velocity parallel to the cathode surface in the range $u_t$ to $u_t + du_t$ is

$$dP(u_t) = \frac{mu_t}{kT}\epsilon^{-mu_t^2/2kT}du_t \qquad (2)$$

The probability that an electron is emitted with kinetic energy normal to the cathode surface in the range $W_n$ to $W_n + dW_n$ is

$$dP(W_n) = \frac{1}{W_T}\epsilon^{-W_n/W_T}dW_n \qquad (3)$$

614

The probability that an electron is emitted with kinetic energy parallel to the cathode surface in the range $W_t$ to $W_t + dW_t$ is

$$dP(W_t) = \frac{1}{W_T}\epsilon^{-W_t/W_T}dW_t \tag{4}$$

The probability that an electron is emitted with total kinetic energy in the range $W$ to $W + dW$ is

$$dP(W) = \frac{W}{W_T}\epsilon^{-W/W_T}\frac{dW}{W_T} \tag{5}$$

The probability that the direction of the emission velocity makes an angle in the range $\theta$ to $\theta + d\theta$ with respect to the normal is

$$dP(\theta) = 2\sin\theta\cos\theta d\theta \tag{6}$$

The emission current density per unit solid angle at an angle $\theta$ with respect to the normal is

$$J_o\frac{dP(\theta)}{2\pi\sin\theta d\theta} = J_o\frac{\cos\theta}{\pi} \tag{7}$$

# APPENDIX V

## *IN AN AXIALLY SYMMETRIC FIELD THE POTENTIAL AT OFF-AXIS POINTS CAN BE EXPRESSED IN TERMS OF THE POTENTIAL ON THE AXIS AND ITS DERIVATIVES*

Here we consider the potential at radius $r$ from an axis of symmetry, which will be designated as the $z$ axis. Laplace's equation for an axially symmetric field is

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial z^2} = 0 \tag{1}$$

Let us suppose that the potential at radius $r$ can be expressed as a power series in $r$ of the form

$$V(z,r) = a_o(z) + a_2(z)r^2 + a_4(z)r^4 + \ldots \tag{2}$$

where the odd powers in $r$ are missing because of symmetry about the axis. Substituting this expression into Equation (1), we obtain for the coefficient of $r^n$

$$(n + 2)^2 a_{n+2} + \frac{d^2 a_n}{dz^2} \tag{3}$$

If the right-hand side of Equation (2) is to be a solution of Equation (1), the coefficient of $r^n$ must be zero. Hence

$$a_{n+2} = -\frac{1}{(n + 2)^2}\frac{d^2 a_n}{dz^2} \tag{4}$$

Now $a_o(z)$ is the potential on the axis, or $V(z,0)$. Hence $a_2(z) = -\frac{1}{4}V''(z,0)$, $a_4(z) = +\frac{1}{64}V''''(z,0)$, and so on, where the primes indicate differentiation with respect to $z$. The potential $V(z,r)$ is therefore given by

$$V(z,r) = V(z,0) - \frac{r^2}{4}V''(z,0) + \frac{r^4}{64}V''''(z,0) - \ldots \tag{5}$$

# APPENDIX VI

## SEVERAL RELATIONS BETWEEN THE OBJECT POSITION, THE IMAGE POSITION AND THE FOCAL LENGTHS OF AN ELECTRON LENS

### (a) The Relation $\dfrac{f_1}{u} + \dfrac{f_2}{v} = 1$

Figure VI-1 shows three electron trajectories $r_1(z)$, $r_2(z)$, and $r_3(z)$ which pass through a region of axially symmetric field. To the left of the lens the trajectory $r_2(z)$ is parallel to the axis but displaced unit distance from it, whereas to the right of the lens it passes through the focal point $F_2$. Similarly $r_1(z)$ passes through the focal point $F_1$ to the left of the lens and emerges parallel to the axis and unit distance from it to the right of the lens. The trajectory $r_3(z)$ crosses the axis $u$ units to the left of the first principal plane and $v$ units to the right of the second principal plane.
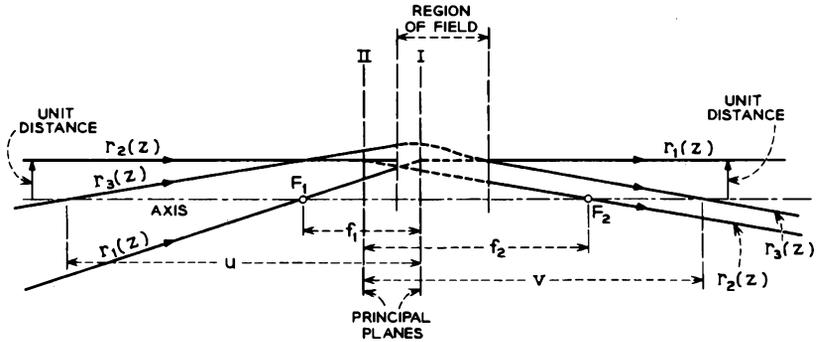


FIG. VI-1 Three electron trajectories which pass through a lens.

Since $r_1(z)$ and $r_2(z)$ are independent solutions of the paraxial ray equation, we can write that

$$r_3(z) = ar_1(z) + br_2(z) \tag{1}$$

where $a$ and $b$ are constants. At the point where $r_3(z)$ crosses the axis to the left of the lens,

$$r_3(z) = 0 = ar_1(z) + br_2(z) = -a\left(\frac{u - f_1}{f_1}\right) + b \tag{2}$$

from which

$$\frac{f_1}{u} = \frac{a}{a + b} \tag{3}$$

At the point where $r_3(z)$ crosses the axis to the right of the lens,

$$r_3(z) = 0 = a - b\left(\frac{v - f_2}{f_2}\right) \tag{4}$$

617

and

$$\frac{f_2}{v} = \frac{b}{a+b} \tag{5}$$

Adding Equations (3) and (5), we obtain

$$\frac{f_1}{u} + \frac{f_2}{v} = 1 \tag{6}$$

Thus, if we know the focal lengths of a lens and the positions of the principal planes, Equation (6) can be used to determine the focusing action of the lens on any trajectory which is close to the axis and nearly parallel to the axis. In the case of an einzel lens in which the electrodes and potentials are symmetrical about a geometrical mid-point of the lens, $f_1 = f_2 = f$, and Equation (6) becomes

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \tag{7}$$

(b) Magnification $\dfrac{h_2}{h_1} = \dfrac{f_1}{f_2}\dfrac{v}{u}$

With the aid of Figure VI-2 and geometrical considerations similar to those
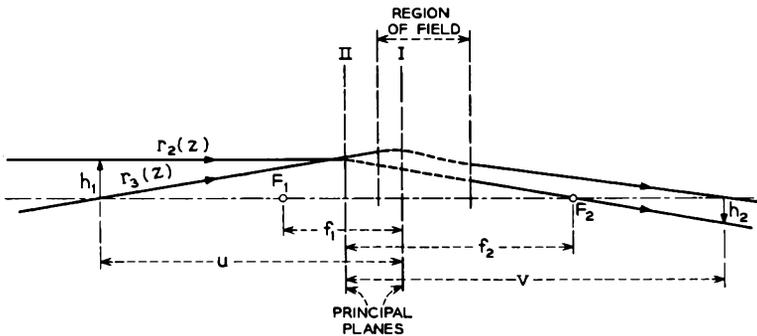


FIG. VI-2 Trajectories used to determine the expression for the magnification of a lens.

used in part (a) above, it is easily shown that the magnification of the lens, given by $h_2/h_1$, can be expressed as

$$\frac{h_2}{h_1} = \frac{v - f_2}{f_2} = \frac{f_1}{f_2}\frac{v}{u} \tag{8}$$

(c) The Relation $\dfrac{f_1}{f_2} = \left(\dfrac{V_1}{V_2}\right)^{1/2}$

Let us suppose that $r_1(z)$ and $r_2(z)$ are two solutions of the paraxial-ray equation, Equation (3.1-8). Then

$$r_1'' + \frac{V'(z,0)}{2V(z,0)}r_1' + \frac{V''(z,0)}{4V(z,0)}r_1 = 0 \tag{9}$$

and

$$r_2'' + \frac{V'(z,0)}{2V(z,0)}r_2' + \frac{V''(z,0)}{4V(z,0)}r_2 = 0 \tag{10}$$

Multiplying the first equation by $r_2$ and the second by $r_1$ and subtracting, we obtain

$$(r_1''r_2 - r_2''r_1) + \frac{V'(z,0)}{2V(z,0)}(r_1'r_2 - r_2'r_1) = 0 \tag{11}$$

which can be expressed as

$$\frac{1}{r_1'r_2 - r_2'r_1}\frac{d(r_1'r_2 - r_2'r_1)}{dz} = -\frac{V'(z,0)}{2V(z,0)} \tag{12}$$

Integrating both sides with respect to $z$, we obtain

$$\ln(r_1'r_2 - r_2'r_1) = -\tfrac{1}{2}\ln V(z,0) + \ln A \tag{13}$$

and hence

$$r_1'r_2 - r_2'r_1 = A[V(z,0)]^{-1/2} \tag{14}$$

where $A$ is a constant. Let us suppose that to the left of the lens $r_2(z) = 1$, $r_2'(z) = 0$, and $V(z,0) = V_1$, and to the right of the lens $r_1(z) = 1$, $r_1'(z) = 0$, and $V(z,0) = V_2$. The trajectories $r_1(z)$ and $r_2(z)$ are therefore as shown in Figure VI-1. From Equation (14), we can write that to the left of the lens

$$r_1' = AV_1^{-1/2} \tag{15}$$

and

$$f_1 = \frac{1}{r_1'} = \frac{V_1^{1/2}}{A} \tag{16}$$

To the right of the lens,

$$-r_2' = AV_2^{-1/2} \tag{17}$$

and

$$f_2 = -\frac{1}{r_2'} = \frac{V_2^{1/2}}{A} \tag{18}$$

Finally, combining Equations (16) and (18), we obtain

$$\frac{f_1}{f_2} = \left(\frac{V_1}{V_2}\right)^{1/2} \tag{19}$$

# APPENDIX VII

## *A STEADY-STATE SOLUTION OF POISSON'S EQUATION FOR A SPACE-CHARGE-LIMITED DIODE IS UNIQUE*

The potential in the interelectrode space of a diode satisfies Poisson's Equation,

$$\nabla^2 V(x,y,z) = -\frac{\rho(x,y,z)}{\varepsilon_o} \tag{1}$$

*We shall assume that the electrons leave the cathode surface with zero velocity.*

Let us first consider the case in which the anode voltage $V_{ao}$ is zero. Clearly, $V(x,y,z) = \rho(x,y,z) = 0$ is the only possible solution in this case. If $\rho$ were not zero, electrons would be present between the electrodes, and the potential in the interelectrode space would be depressed below cathode potential. Field lines then would extend from induced charges on the cathode to the charge in the interelectrode space, and the potential gradient $dV/dn$ at the cathode surface would be negative. However, since the electrons leave the cathode surface with zero velocity, all the electrons would be returned to the cathode immediately upon emission. This contradicts our assumption that there is charge present in the interelectrode space. Thus we must conclude that for $V_{ao} = 0$, $\rho(x,y,z) = 0$. Equation (1) then reduces to Laplace's Equation, which gives $V(x,y,z) = 0$ for the boundary condition $V_{ao} = 0$.

For $V_{ao} > 0$ and for space-charge-limited operation, a steady-state solution to Equation (1) must satisfy the following boundary conditions:

1. $V = \dfrac{dV}{dn} = 0$

at the cathode surface, where $d/dn$ is the derivative in the direction normal to the cathode surface. (Note that the assumption of zero emission velocity means that the potential minimum coincides with the cathode.)

2. $V = V_{ao}$ at the anode surface.

Let us suppose that there are two independent solutions of Equation (1) which meet these boundary conditions for $V_{ao} > 0$. Let the solutions be:

$V_1(x,y,z)$, corresponding to a charge distribution $\rho_1(x,y,z)$

and

$V_2(x,y,z)$, corresponding to a charge distribution $\rho_2(x,y,z)$

In such a case $V = V_1 - V_2$ would be a solution of Equation (1) which satisfies the boundary conditions $V = 0$ at the anode and $V = dV/dn = 0$ at the cathode and which corresponds to a charge distribution $\rho_1 - \rho_2$. But from the discussion in the previous paragraph we know that $V(x,y,z) = \rho(x,y,z) = 0$ is the only solution to Equation (1) which meets the boundary conditions for $V_{ao} = 0$. Thus we conclude that $V_1 = V_2$, and $\rho_1 = \rho_2$. Hence a steady-state solution of Poisson's Equation for a space-charge-limited diode is unique.

## APPENDIX VIII

### IF A TWO-DIMENSIONAL POTENTIAL IN FREE SPACE IS SYMMETRIC ABOUT AN AXIS, THE POTENTIAL AT OFF-AXIS POINTS CAN BE EXPRESSED IN TERMS OF THE POTENTIAL ON THE AXIS

The derivative of a complex function $f(z) = f(x + jy) = u(x,y) + jv(x,y)$ is defined as

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \tag{1}$$

where $\Delta z = \Delta x + j\Delta y$. It can be shown that necessary and sufficient conditions for the existence of a unique derivative of $f(z)$ are that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{2}$$

and

$$\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} \tag{3}$$

These equations are known as the Cauchy-Riemann conditions. (See, for instance, R. V. Churchill, *Introduction to Complex Variables and Applications*, McGraw-Hill Book Co., Inc., New York, 1948, p. 30.) A function $f(z) = u + jv$ is said to be *analytic* in a region of the $z$ plane if the derivative $f'(z)$ exists at every point in that region. Examples of analytic functions are $z$, $z^2 + 1$, $e^z$, and $\sin z$, where $z = x + jy$ in each case.

If we take the partial derivative of Equation (2) with respect to $x$ and the partial derivative of Equation (3) with respect to $y$, we obtain

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y} \quad \text{and} \quad \frac{\partial^2 v}{\partial x \partial y} = -\frac{\partial^2 u}{\partial y^2} \tag{4}$$

from which

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{5}$$

In a similar manner it can be shown that

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \tag{6}$$

Hence the real and imaginary parts of an analytic function of $z$ satisfy Laplace's Equation in two dimensions.

Now $u(x,y)$ can be expressed as

$$u(x,y) = \operatorname{Re} f(x + jy) = \tfrac{1}{2}[f(x + jy) + f(x - jy)] \tag{7}$$

If $y$ in this expression is replaced by $-y$, the value of $u$ is unchanged, so that $u(x,y)$ is symmetric about the $x$ axis. It follows, therefore, that the real part of an analytic function $f(z)$ defines a potential which satisfies Laplace's Equation in two dimensions and which is symmetric about the $x$ axis.

621
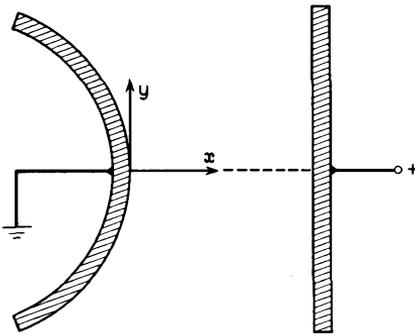
FIG. VIII-1 Two electrodes which are symmetric about an axis and which extend indefinitely above and below the page.

Figure VIII-1 shows two electrodes which are assumed to extend to infinity above and below the page. One is at ground potential, the other is at a positive potential. The coordinate axes shown in the figure are such that the $x$ axis lies in the plane of symmetry of the electrodes. Let us suppose the potential along the $x$ axis between the electrodes is given by $V = f_1(x)$, where $f_1(z) = f_1(x + jy) = u_1(x,y) + jv_1(x,y)$ is an analytic function. Then consider the potential given by

$$u_1(x,y) = \tfrac{1}{2}[f_1(x + jy) + f_1(x - jy)] \quad (8)$$

This satisfies Laplace's Equation in two dimensions, and when $y = 0$, it gives the correct potential along the $x$ axis. Furthermore, it is symmetric about the $x$ axis.

To show that no other function $u_2(x,y)$ also satisfies these conditions, and hence that $u_1(x,y)$ is in fact the potential in the interelectrode space of Figure VIII-1, we can make use of two other results of complex variable theory. These are:

1. Any function $u_2(x,y)$ which satisfies Laplace's Equation in two dimensions defines a function $v_2(x,y)$ such that $f_2(z) = u_2(x,y) + jv_2(x,y)$ is analytic. (See Churchill, p. 139.)

2. If $f_1(z)$ and $f_2(z)$ are analytic throughout a region in the $z$ plane, and if $f_1(z) = f_2(z)$ along a curve within the region, then $f_1(z) = f_2(z)$ throughout the region. (Churchill, page 189.)

It follows from (1) above that the function $u_2(x,y)$ defines an analytic function $f_2(z)$. But $f_2(z) = f_1(z)$ along the $x$ axis, and consequently $f_2(z) = f_1(z)$ at points off the $x$ axis, and $u_2(x,y) = u_1(x,y)$.

# APPENDIX IX

## *APPROXIMATE EXPRESSIONS FOR THE ELECTROSTATIC AMPLI-FICATION FACTOR OF A PLANAR TRIODE AND FOR THE FUNCTIONS $F_1$ AND $F_2$*

Here we make use of complex variable theory and that branch of complex variable theory known as conformal mapping to derive approximate expressions for the electrostatic amplification factor of a planar triode and for the functions $F_1$ and $F_2$. First it will be helpful to discuss a few concepts relating to conformal mapping.

Suppose that $w = u(x,y) + jv(x,y) = f(z) = f(x + jy)$ is an analytic function of $z$. (See Appendix VIII for a definition of an analytic function.) The function $f(z)$ "maps" each point $z_o$ in the $x$-$y$ plane into a corresponding point $w_o$ in the $u$-$v$ plane. Suppose that two curves $C_1$ and $C_2$ in the $x$-$y$ plane intersect at $z_o$, and the tangents to these curves at $z_o$ make an angle $\gamma$ with each other. The two curves will map on the $u$-$v$ plane into two curves $C_1'$ and $C_2'$ which intersect at $w_o$. Furthermore it can be shown that, provided $f(z)$ is analytic at $z_o$ and provided $f'(z_o) \neq 0$, the angle between the curves $C_1'$ and $C_2'$ at $w_o$ is also $\gamma$. (See, for instance, R. V. Churchill, *Introduction to Complex Variables and Applications*, McGraw-Hill Book Co., Inc., New York, 1948, pp. 135 and 136.) A mapping which preserves angles between pairs of curves in this manner is said to be conformal.

For example, the equation $u(x,y) = a$ defines a curve in the $x$-$y$ plane, and every point on that curve maps onto a point on the straight line $u = a$ in the $u$-$v$ plane. (See Figure IX-1.) Similarly, every point on the curve $v(x,y) = b$ in the $x$-$y$
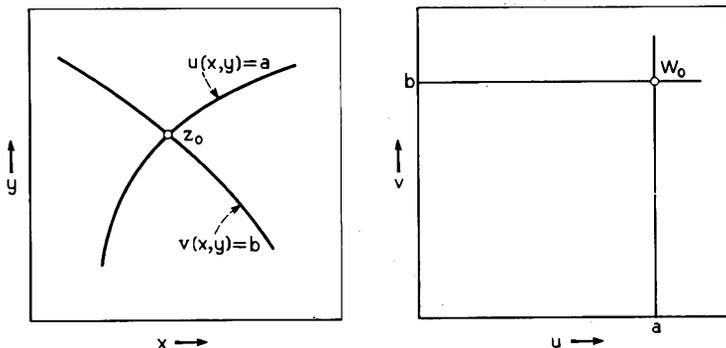


FIG. IX-1 The curves $u(x,y) = a$ and $v(x,y) = b$ in the $x$-$y$ plane map into the lines $u = a$ and $v = b$ in the $u$-$v$ plane.

plane maps onto the line $v = b$ in the $u$-$v$ plane. Suppose the two curves in the $x$-$y$ plane intersect at $z_o$. Consider the slope of the curve $u(x,y) = a$ at $z_o$. Taking the total derivative of $u(x,y) = a$, we obtain

$$du = \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial y}dy = 0 \qquad (1)$$

and

$$\frac{dy}{dx} = -\frac{\dfrac{\partial u}{\partial x}}{\dfrac{\partial u}{\partial y}} \tag{2}$$

If the partial derivatives in Equation (2) are evaluated at $z_o$, the expression gives the slope of the curve $u(x,y) = a$ at $z_o$. Similarly the slope of the curve $v(x,y) = b$ at $z_o$ is given by $-(\partial v/\partial x)/(\partial v/\partial y)$, where the partial derivatives are evaluated at $z_o$. By applying the Cauchy-Riemann conditions given by Equations (2) and (3) of Appendix VIII, it is easily shown that the slope of one curve is minus the reciprocal of the slope of the second curve and hence that the two curves are orthogonal to each other at $z_o$. Since the curves map into the lines $u = a$ and $v = b$ in the $u$-$v$ plane, the mapped curves are likewise orthogonal at the point of intersection, and the mapping preserves the angle between the curves.

*Similarly, a set of orthogonal curves giving the equipotential contours and the corresponding field lines for a particular boundary value problem in the $x$-$y$ plane would map into a second set of orthogonal curves in the $u$-$v$ plane.*

Suppose that $V(x,y)$ is a solution of Laplace's Equation in two dimensions for a particular boundary-value problem in the $x$-$y$ plane. Then

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \tag{3}$$

Let the potential $V(x,y)$ be generated by applying voltages $V_1, V_2, \ldots$ to electrodes $1, 2, \ldots$ in the $x$-$y$ plane. The curves defining the electrodes in the $x$-$y$ plane map into corresponding curves in the $u$-$v$ plane. Let us further suppose that the voltages $V_1, V_2, \ldots$ are applied to the "mapped electrodes." Now since each point $z_o$ in the $x$-$y$ plane defines a point $w_o$ in the $u$-$v$ plane, the potential $V(x,y)$ can be expressed as $V(u,v)$. Furthermore $V(u,v) = V_1$ at mapped electrode 1, it equals $V_2$ at mapped electrode 2, and so on. In addition, it is shown below that $V(u,v)$ satisfies Laplace's Equation in the $u$-$v$ plane, and since solutions to Laplace's Equation are unique, it follows that $V(u,v)$ gives the potential in the region surrounding the mapped electrodes when the voltages $V_1, V_2, \ldots$ are applied to them. Thus the equipotential contours and field lines in the $x$-$y$ plane map onto corresponding equipotential contours and field lines in the $u$-$v$ plane. This result is useful in solving two-dimensional potential problems because it may be possible to determine $V(u,v)$ more easily than $V(x,y)$. However, once we know $V(u,v)$, we can obtain $V(x,y)$ by a change of variables.

As a final point let us show that $V(u,v)$ satisfies Laplace's Equation in two dimensions. We can write

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial V}{\partial v}\frac{\partial v}{\partial x} \tag{4}$$

and

$$\frac{\partial^2 V}{\partial x^2} = \frac{\partial^2 V}{\partial u^2}\left(\frac{\partial u}{\partial x}\right)^2 + 2\frac{\partial^2 V}{\partial u \partial v}\frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial V}{\partial u}\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 V}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2 + \frac{\partial V}{\partial v}\frac{\partial^2 v}{\partial x^2} \tag{5}$$

A similar expression can be obtained for $\partial^2 V/\partial y^2$. Substituting these expressions in

Equation (3) and making use of Equations (2), (3), (5), and (6) of Appendix VIII, we obtain

$$\left(\frac{\partial^2 V}{\partial u^2} + \frac{\partial^2 V}{\partial v^2}\right)\left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2\right] = 0 \tag{6}$$

Since the quantity in the rectangular brackets is generally not zero, it follows that

$$\frac{\partial^2 V}{\partial u^2} + \frac{\partial^2 V}{\partial v^2} = 0 \tag{7}$$

Thus $V(u,v)$ satisfies Laplace's Equation.

Next let us see how these results can be applied to a planar triode in the absence of space charge. Figure IX-2(a) shows a portion of the planar triode. One grid



(a)                                                                 (b)
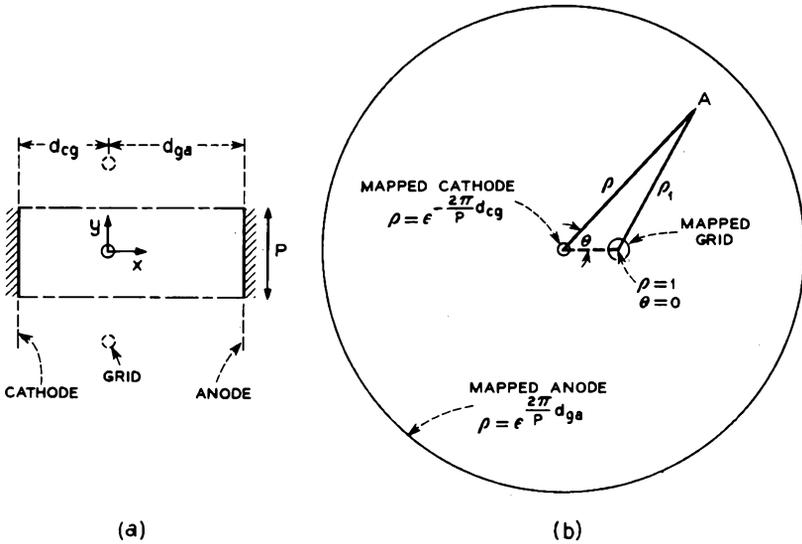
FIG. IX-2  A portion of a planar triode in the $x$-$y$ plane mapped onto the $\rho$-$\theta$ plane.

wire is shown with a solid ring, and the two adjacent wires are shown with broken rings. Two lines extend from the cathode to the anode midway between the central grid wire and the two adjacent wires. The pitch of the grid is $P$, so that the spacing between the two lines is $P$. We shall consider only the portion of the triode between the two lines. Clearly the amplification factor of this portion of the triode will be the same as for the whole device.

It will be convenient to replace $u + jv$ in the foregoing discussion by $\rho e^{j\theta}$. The transformation which enables us to solve the potential problem in the planar triode is given by

$$u + jv = \rho e^{j\theta} = f(z) = \epsilon^{2\pi z/P} = \epsilon^{2\pi(x+jv)/P} \tag{8}$$

From this we see that

$$\rho = \epsilon^{2\pi x/P} \tag{9}$$

and

$$\theta = \frac{2\pi}{P}y \tag{10}$$

Figure IX-2(b) shows how the triode maps onto the $\rho$-$\theta$ plane by means of this transformation. The anode maps onto a circle $\epsilon^{2\pi d_{ga}/P}$, where $d_{ga}$ is the grid-to-anode spacing; whereas the cathode maps onto a circle of radius $\epsilon^{-2\pi d_{cg}/P}$, where $d_{cg}$ is the cathode-to-grid spacing. Since $\epsilon^{2\pi} \approx 256$, it is evident that the radius of the mapped anode is very large compared with unity, and the radius of the mapped cathode is very small compared with unity. The center of the grid wire maps onto the point $\rho = 1$, $\theta = 0$, and the perimeter of the grid wire maps onto a closed curve about the point $\rho = 1$, $\theta = 0$. If the radius of the grid wire is very small compared with $P$, say less than $P/20$, the closed curve is nearly a circle and is small in diameter. Notice that the transformation maps the point $(x, y + nP)$, where $n$ is an integer, onto the same point in the $\rho$-$\theta$ plane as the point $(x,y)$. Consequently the remainder of the triode beyond the part shown in Figure IX-2(a) is mapped on top of the mapping shown Figure IX-2(b).

Suppose the mapped cathode carries an axial charge density $+\tau_c$ coulombs per unit length in the direction normal to the page, and the mapped grid carries an axial charge density $+\tau_g$ coulombs per unit length in the direction normal to the page. We shall assume that the radii of the mapped cathode and mapped grid are sufficiently small that these axial charge densities can be considered as line charges. An expression for the potential resulting from a line charge $\tau$ coulombs per unit length surrounded *coaxially* by a cylindrical conductor is obtained by integrating Equation (1.4-5) with respect to radius $r$. Thus

$$V = V_o - \frac{\tau}{2\pi\varepsilon_o}\ln r \tag{11}$$

where $V_o$ is a constant that adjusts for the level of potential in the region, and $r$ is the radial distance from the line charge to the point where the potential is determined. The potential in the interelectrode space of Figure IX-2(b) is a superposition of that arising from the axial charge density $\tau_c$ and that arising from the axial charge density $\tau_g$. The contribution resulting from the axial charge density $\tau_c$ is given by Equation (11), where $\tau$ becomes $\tau_c$, and $r$ becomes $\rho$. Since the radius of the mapped anode is much greater than unity, the axial charge density $\tau_g$ is *nearly* coaxial with the mapped anode. Hence to a good approximation Equation (11) also can be used to give the potential resulting from this axial charge, where in this case radius $r$ is measured from the point $\rho = 1$, $\theta = 0$. Thus an approximate expression for the potential at point $A$ is given by

$$V(\rho,\theta) = C - \frac{\tau_c}{2\pi\varepsilon_o}\ln\rho - \frac{\tau_g}{2\pi\varepsilon_o}\ln\rho_1$$

$$= C - \frac{\tau_c}{2\pi\varepsilon_o}\ln\rho - \frac{\tau_g}{4\pi\varepsilon_o}\ln(\rho^2 + 1 - 2\rho\cos\theta) \tag{12}$$

where $\rho_1 = (\rho^2 + 1 - 2\rho\cos\theta)^{1/2}$ is the distance from the point $\rho = 1$, $\theta = 0$ to point $A$, and $C$ is a constant which adjusts for the level of potential in the region.

Combining Equations (9), (10), and (12), the potential in the interelectrode space of the planar triode can be expressed as

$$V(x,y) = C - \frac{\tau_c x}{\varepsilon_o P} - \frac{\tau_g}{4\pi\varepsilon_o}\ln\left(\epsilon^{4\pi x/P} + 1 - 2\epsilon^{2\pi x/P}\cos\frac{2\pi}{P}y\right) \tag{13}$$

At the cathode $x = -d_{cg}$, and the exponential terms in the argument of the logarithm are extremely small compared with unity. The argument of the logarithm is therefore nearly 1, and the potential at the cathode is approximately given by

$$V_{co} = C + \frac{\tau_c d_{cg}}{\varepsilon_o P} \tag{14}$$

If we assume that the cathode potential is zero, then

$$C = -\frac{\tau_c d_{cg}}{\varepsilon_o P} \tag{15}$$

At the anode $x = d_{ga}$, and the first term in the argument of the logarithm is far larger than the second and third terms. Neglecting the second and third terms, and substituting for $C$ from Equation (15), the anode potential is found to be

$$V_{ao} = -\frac{\tau_c d_{cg} + \tau_c d_{ga} + \tau_g d_{ga}}{\varepsilon_o P} \tag{16}$$

Examination of the equipotentials given by Equation (13) indicates that in the neighborhood of the origin they are very nearly circles about the origin. If the grid radius $R$ is small compared with $P$, say $R \leq P/20$, the potential at $(R, 0)$ is therefore very nearly equal to the potential at $(0, R)$. Let us evaluate the potential of the grid by setting $x = 0$ and $y = R$. Then

$$V_{go} = -\frac{\tau_c d_{cg}}{\varepsilon_o P} - \frac{\tau_g}{2\pi\varepsilon_o} \ln 2 \sin \frac{\pi R}{P} \tag{17}$$

where we have made use of the relationship $1 - \cos 2\alpha = 2 \sin^2 \alpha$.

Now the electrostatic amplification factor of the triode is equal to minus the ratio of the anode voltage to grid voltage needed to give zero electric field at the cathode. Since $\tau_c = 0$ when the electric field at the cathode is zero, the electrostatic amplification factor is given by

$$\mu_{es} = -\frac{V_{ao}}{V_{go}}\bigg|_{\text{for } \tau_g=0} = -\frac{2\pi d_{ga}}{P \ln \left( 2 \sin \dfrac{\pi R}{P} \right)} \tag{18}$$

As a final point, Equations (16) and (17) can be solved for $\tau_c$ and $\tau_g$ and the resulting expressions can be substituted in Equation (13). In this way we can express $V(x,y)$ in the form

$$V(x,y) = V_{go} F_1 + V_{ao} F_2 \tag{19}$$

where

$$F_1 = \mu_{es} \frac{d_{cg} + x - (P/4\pi d_{ga})(d_{ga} + d_{cg}) \ln \left( \epsilon^{4\pi z/P} + 1 - 2\epsilon^{2\pi z/P} \cos \dfrac{2\pi}{P} y \right)}{d_{ga} + d_{cg} + \mu_{es} d_{cg}} \tag{20}$$

and

$$F_2 = \frac{d_{cg} + x + (P/4\pi d_{ga})\mu_{es} d_{cg} \ln \left( \epsilon^{4\pi z/P} + 1 - 2\epsilon^{2\pi z/P} \cos \dfrac{2\pi}{P} y \right)}{d_{ga} + d_{cg} + \mu_{es} d_{cg}} \tag{21}$$

At $x = -d_{cg}$, the argument of the logarithmic term is very nearly 1, and the derivative of the logarithmic term is essentially zero. If these approximations are taken into account, it is easily shown that

$$\mu_{es} = \frac{\dfrac{\partial F_1}{\partial x}}{\dfrac{\partial F_2}{\partial x}}\Bigg|_{x=-d_{cg}} \tag{22}$$

Note that the origin for the coordinate system is different here than in Equation (5.1-1).

# APPENDIX X

## IMPEDANCE OF A SPACE-CHARGE-LIMITED PLANAR DIODE

When a small ac signal is superimposed on the dc voltage applied to a space-charge-limited diode, the electron velocity exhibits an ac component, and an ac induced current flows in the external circuit connected to the diode. Here we derive an expression for the impedance given by the ratio of the applied ac voltage to the ac induced current in the external circuit for the case of a planar diode.

The planar diode is illustrated in Figure X-1. We assume that the virtual cathode coincides with the actual cathode so that, strictly speaking, our results apply only to this case. Let us gather together three relations which we shall use later in obtaining the impedance of the diode:
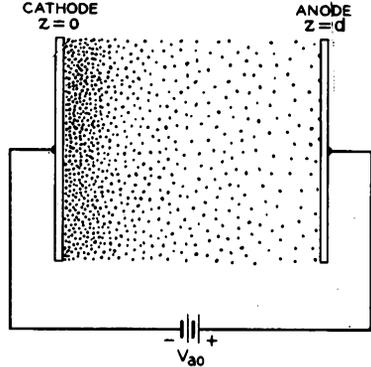


FIG. X-1 A space-charge-limited planar diode.

1. From Equations (4.1-8) and (4.1-10), the dc current density flowing between the electrodes of the diode is given by

$$J_o = \frac{4\varepsilon_o(2\eta)^{1/2}}{9}\frac{V^{3/2}}{z^2} = \frac{2\varepsilon_o}{9\eta}\frac{u_o^3}{z^2} = \frac{4\varepsilon_o(2\eta)^{1/2}}{9}\frac{V_{ao}^{3/2}}{d^2} \qquad (1)$$

where $\eta = e/m$, $V$ is the dc potential at distance $z$ from the cathode, $u_o = \sqrt{2\eta V}$ is the dc electron velocity at $z$, $V_{ao}$ is the dc anode voltage, and $d$ is the electrode spacing.

2. The electron transit time between the electrodes in the absence of an applied ac signal is given by

$$T_o = \int_0^d \frac{dz}{u} = \left[\frac{6\varepsilon_o d}{\eta J_o}\right]^{1/3} \qquad (2)$$

where we have substituted from Equation (1) for $u_o$.

3. The low-frequency "dynamic anode resistance" for a unit area of the electrodes is given by

$$r_a = \frac{dV_{ao}}{dJ_o} = \frac{2}{3}\frac{V_{ao}}{J_o} = \frac{T_o d}{2\varepsilon_o} \qquad (3)$$

Note that this has the dimensions of resistance *times* area.

Let us proceed now with the derivation of the impedance of the diode. Since the fields are one-dimensional, Poisson's Equation for this problem is

$$\frac{\partial E}{\partial z} = \frac{\rho}{\varepsilon_o} \qquad (4)$$

where $\rho$ is the space-charge density at the point where $E$ is determined. The equation of continuity, Equation (1.3-2), becomes

$$\frac{\partial \rho(u_o + u)}{\partial z} + \frac{\partial \rho}{\partial t} = 0 \qquad (5)$$

where $u$ is the ac component of the electron velocity. Combining Equations (4) and (5), we obtain

$$\frac{\partial}{\partial z}\left[\rho(u_o + u) + \varepsilon_o\frac{\partial E}{\partial t}\right] = 0 \qquad (6)$$

This equation states that the quantity $\rho(u_o + u) + \varepsilon_o(\partial E/\partial t)$ is independent of $z$ in the region between the electrodes. The term $\rho(u_o + u)$ is simply the current density resulting from the flow of charge between the electrodes. The second term also has the dimensions of a current density and is called the displacement current density.

To understand the displacement current density better, consider a parallel-plate capacitor in the absence of space charge. If a time-varying voltage $V$ is applied to it, a current $i = C(dV/dt) = \varepsilon_o A(dE/dt)$ flows in the leads to the capacitor. The quantity $\varepsilon_o(dE/dt)$ is the displacement current density. The total displacement current crossing a surface between the plates is equal to the conduction current in the leads. It is a real current in the sense that it gives rise to the same magnetic field that would be produced by a similar distribution of current in a conductor. In the parallel-plate capacitor, the displacement current density is uniform everywhere between the plates. However, in the space-charge-limited diode, the electric field lines extend between charge that is in transit between the electrodes and induced charges on the anode. Clearly $E$ is a function of $z$ in this case, so that the displacement current density also is a function of $z$.

We shall call the quantity $\rho(u_o + u) + \varepsilon_o(\partial E/\partial t)$ the total current density and denote it by $J_T$. The current flowing in the cathode and anode leads of the diode is given by $J_T A$, where $A$ is the area of the electrodes. With the aid of Equation (4), $J_T$ can be expressed as

$$J_T = \rho(u_o + u) + \varepsilon_o\frac{\partial E}{dt} = \varepsilon_o\left(\frac{\partial E}{\partial z}\frac{dz}{dt} + \frac{\partial E}{\partial t}\right) \qquad (7)$$

The quantity in the brackets is the time rate of change of the electric field experienced by a moving electron. The first term arises from the variation of $E$ with $z$ as the electron travels with velocity $dz/dt$ toward the anode, and the second term arises from the time variation of the field. The sum of the terms gives the total derivative of $E$ with respect to time. Hence

$$J_T = \varepsilon_o\frac{dE}{dt} \qquad (8)$$

The acceleration of the electron is given by

$$\frac{d^2z}{dt^2} = -\eta E \qquad (9)$$

Differentiating this with respect to time and substituting for $dE/dt$ from Equation (8), we obtain

$$\frac{d^3z}{dt^3} = -\eta\frac{dE}{dt} = -\frac{\eta}{\varepsilon_o}J_T \qquad (10)$$

Equation (6) showed that $J_T$ is independent of $z$. However, it is not independent of time if the voltage applied to the electrodes is changing with time. If a small sinusoidal voltage is superimposed on the dc voltage between the electrodes, $J_T$ will have a sinusoidal component and can be expressed as

$$J_T = -(J_o + J_1 \sin \omega t) \qquad (11)$$

where $J_o$ is the current density given by Equation (1), and $J_1$ is small compared with $J_o$. The minus sign implies that the current flows in the minus $z$ direction, or toward the cathode. Substituting in Equation (10) for $J_T$, we can perform successive integrations with respect to time to obtain expressions for the acceleration, velocity, and position of an electron with respect to time. Let the time the electron left the cathode be $t_o$. The boundary conditions at time $t_o$ are $z = dz/dt = d^2z/dt^2 = 0$. Integrating Equation (10) from $t_o$ to $t$ then gives

$$\frac{d^2z}{dt^2} = \frac{\eta}{\varepsilon_o}\left[ J_o(t - t_o) - \frac{J_1}{\omega}(\cos \omega t - \cos \omega t_o) \right] \qquad (12)$$

Integrating a second time gives

$$\frac{dz}{dt} = \frac{\eta}{\varepsilon_o}\left[ J_o\frac{(t - t_o)^2}{2} - \frac{J_1}{\omega^2}(\sin \omega t - \sin \omega t_o) + \frac{J_1}{\omega}(t - t_o) \cos \omega t_o \right] \qquad (13)$$

Finally,

$$z = \frac{\eta}{\varepsilon_o}\left[ J_o\frac{(t - t_o)^3}{6} + \frac{J_1}{\omega^3}(\cos \omega t - \cos \omega t_o) + \frac{J_1}{\omega^2}(t - t_o) \sin \omega t_o + \frac{J_1}{\omega}\frac{(t - t_o)^2}{2} \cos \omega t_o \right] \qquad (14)$$

By setting $z = d$ in the last of these equations, the time $t - t_o$ becomes the electron transit time $T$. If we then multiply both sides of the equation by $6\varepsilon_o/\eta J_o$, the equation can be rewritten in the form $T^3 = T_o^3 + \delta = T_o^3(1 + \delta/T_o^3)$, where $\delta$ is a summation of terms containing $J_1/J_o$ as a factor. If the ac voltage applied between the electrodes is small compared with the dc voltage, $J_1$ is small compared with $J_o$, and $\delta/T_o^3$ is small compared with unity. We can then write $T = T_o (1 + \delta/3T_o^3) = T_o + \delta/3T_o^2$. Since $\delta/3T_o^2$ is small compared with $T_o$, we can set $t - t_o = T_o$ and $t_o = t - T_o$ in the expression for $\delta/3T_o^2$. This is equivalent to neglecting terms containing the product of two or more small quantities. Thus we obtain

$$T = T_o - \frac{J_1}{J_o}\frac{2}{\omega^3 T_o^2}\left[ \cos \omega t + \omega T_o \sin \omega(t - T_o) \right.$$
$$\left. + \left(\frac{(\omega T_o)^2}{2} - 1\right) \cos \omega(t - T_o) \right] \qquad (15)$$

As a final part of our calculations we shall integrate the electric field $E$ from the cathode to the anode to obtain an expression for the instantaneous voltage between the cathode and anode. Thus

$$v_a = -\int_o^d E dz = \frac{1}{\eta}\int_o^d \frac{d^2z}{dt^2}dz = \frac{1}{\eta}\int_{t_o=t}^{t_o=t-T} \frac{d^2z}{dt^2}\frac{dz}{dt_o}dt_o \qquad (16)$$

where $d^2z/dt^2$ is the acceleration of an electron at distance $z$ from the cathode at time $t$, and we have substituted $E = -(1/\eta)d^2z/dt^2$ from Equation (9). The

derivative $dz/dt_o$ can be found by differentiating Equation (14) with respect to $t_o$. Thus

$$\frac{dz}{dt_o} = -\frac{\eta}{\varepsilon_o} \frac{(t - t_o)^2}{2}(J_o + J_1 \sin \omega t_o) \tag{17}$$

We can now substitute from Equations (12) and (17) into Equation (16) and carry out the integration. We shall neglect terms containing the product of two or more small quantities, that is, terms containing $J_1^2$ and higher powers of $J_1$. The time $T$ appears in the answer, and we can substitute for it from Equation (15). Thus we obtain

$$v_a = V_o + \frac{\eta J_o J_1}{\varepsilon_o^2 \omega^4}\Big\{[2(1 - \cos \omega T_o) - \omega T_o \sin \omega T_o] \sin \omega t$$
$$- \left[\frac{(\omega T_o)^3}{6} + \omega T_o(1 + \cos \omega T_o) - 2 \sin \omega T_o\right] \cos \omega t\Big\} \tag{18}$$

This is of the form

$$v_a = V_o + r J_1 \sin \omega t + x J_1 \cos \omega t \tag{19}$$

where $r$ and $x$ are, respectively, a resistance for a unit area and a reactance for a unit area and are given by

$$r = 12 r_a \left[\frac{2(1 - \cos \omega T_o) - \omega T_o \sin \omega T_o}{(\omega T_o)^4}\right] = r_a\left[1 - \frac{(\omega T_o)^2}{15} + \ldots\right] \tag{20}$$

$$x = -12 r_a\left[\frac{1}{6\omega T_o} + \frac{\omega T_o(1 + \cos \omega T_o) - 2 \sin \omega T_o}{(\omega T_o)^4}\right]$$
$$= -r_a\left[\frac{3}{10}\omega T_o - \frac{(\omega T_o)^3}{84} + \ldots\right] \tag{21}$$

and $r_a = \frac{4}{3}V_{ao}/J_o$ from Equation (3). The quantities $r/r_a$ and $x/r_a$ are plotted in Figure 7.1-6 as functions of $\omega T_o$.

An ac induced current $J_1 A \sin \omega t$ flows in the external circuit in response to the applied ac voltage given by Equation (19), where $A$ is the area of the electrodes. The diode therefore presents an impedance given by $Z = (r/A) + j(x/A)$ to the applied ac voltage. At low frequencies, or small $\omega T_o$, this impedance reduces to

$$Z = \frac{r_a}{A}\left(1 - j\frac{3}{10}\omega T_o\right) \tag{22}$$

In constructing an equivalent network for the device it is more useful to use the admittance given by the reciprocal of this, or

$$Y = \frac{1}{Z} \approx \frac{A}{r_a}\left(1 + j\frac{3}{10}\omega T_o\right) = g_o + j\omega\frac{3}{5}C_o \tag{22}$$

where $g_o = A/r_a = A\partial I_{ao}/\partial V_{ao}$, and $C_o = \varepsilon_o A/d$ is the capacitance of the parallel-plate capacitor formed by the anode and cathode in the absence of space charge. This equation indicates that at low frequencies the diode acts as a conductance $g_o = A\partial J_o/\partial V_{ao}$ shunted by a capacitance equal to $\frac{3}{5}$ times the capacitance of the diode in the absence of space charge. A low-frequency equivalent network for the diode is shown in Figure 7.1-8.

# APPENDIX XI

## *LLEWELLYN-PETERSON COEFFICIENTS*

Tabulated below are the coefficients in the Llewellyn-Peterson equations discussed in Section 7.2.

$$A^* = \frac{1}{\epsilon_o}(u_{ao} + u_{bo})\frac{T^2}{2}\frac{1}{\beta}\left[1 - \frac{\zeta}{3}\left(1 - \frac{12S}{\beta^3}\right)\right]$$

$$B^* = \frac{1}{\epsilon_o}\frac{T^2}{\beta^3}\left[u_{ao}(P - \beta Q) - u_{bo}P + \zeta(u_{ao} + u_{bo})P\right]$$

$$C^* = -\frac{1}{\eta}2\zeta(u_{ao} + u_{bo})\frac{P}{\beta^2}$$

$$D^* = 2\zeta\left(\frac{u_{ao} + u_{bo}}{u_{bo}}\right)\frac{P_i}{\beta^2}$$

$$E^* = \frac{1}{u_{bo}}[u_{bo} - \zeta(u_{ao} + u_{bo})]\epsilon^{-\beta}$$

$$F^* = \frac{\epsilon_o}{\eta}\frac{2\zeta}{T^2}\left(\frac{u_{ao} + u_{bo}}{u_{bo}}\right)\beta\epsilon^{-\beta}$$

$$G^* = -\frac{\eta}{\epsilon_o}\frac{T^2}{\beta^3}\frac{1}{u_{bo}}[u_{bo}(P - \beta Q) - u_{ao}P + \zeta(u_{ao} + u_{bo})P]$$

$$H^* = -\frac{\eta}{\epsilon_o}\frac{T^2}{2}\left(\frac{u_{ao} + u_{bo}}{u_{bo}}\right)(1 - \zeta)\frac{\epsilon^{-\beta}}{\beta}$$

$$I^* = \frac{1}{u_{bo}}[u_{ao} - \zeta(u_{ao} + u_{bo})]\epsilon^{-\beta}$$

The following dc quantities and relationships pertain to these coefficients:

$u_{ao}$ and $u_{bo}$ are the dc electron velocities at planes $a$ and $b$.

$T$ is the electron transit time from plane $a$ to plane $b$ in the presence of the dc space charge in the interelectrode space but in the absence of applied ac signals.

$T_o$ is the electron transit time from plane $a$ to plane $b$ in the absence of space charge and in the absence of applied ac signals. (Note that $T$ and $T_o$ have meanings in this appendix different from those in Appendix X.)

$\zeta$ is a space charge factor related to $T$ and $T_o$ by $\zeta = 3(1 - T_o/T)$.

If $d$ is the distance from plane $a$ to plane $b$, then

$$d = (1 - \zeta/3)(u_{ao} + u_{bo})\frac{T}{2}$$

If $J_o$ is the dc current density passing through either of the planes,

$$J_o = \frac{\epsilon_o}{\eta}(u_{ao} + u_{bo})\frac{2\zeta}{T^2}$$

The following quantities contain the angular frequency $\omega$ of the ac signal.

$\beta = j\omega T$, where $j = \sqrt{-1}$, and $\omega$ is $2\pi$ times the frequency of the ac signal.

$$P = 1 - \epsilon^{-\beta} - \beta\epsilon^{-\beta} \approx \frac{\beta^2}{2} - \frac{\beta^3}{3} + \frac{\beta^4}{8} \cdots$$

$$Q = 1 - e^{\beta} \approx \beta - \frac{\beta^2}{2} + \frac{\beta^3}{6} - \frac{\beta^4}{24} \cdots$$

$$S = 2 - 2\epsilon^{-\beta} - \beta - \beta\epsilon^{-\beta} \approx -\frac{\beta^3}{6} + \frac{\beta^4}{12} - \frac{\beta^5}{40} + \frac{\beta^6}{180} \cdots$$

# APPENDIX XII

## *SOME USEFUL VECTOR RELATIONSHIPS*

Let **A** be an arbitrary vector and $\Phi$ an arbitrary scalar in the real or phasor notation. $\mathbf{i}_x$, $\mathbf{i}_y$, etc. are unit vectors in the coordinate directions indicated by the subscripts. The following relationships apply to **A** and $\Phi$. In rectangular coordinates:

$$\mathbf{A} = \mathbf{i}_x A_x + \mathbf{i}_y A_y + \mathbf{i}_z A_z \tag{1}$$

$$\nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{2}$$

$$\nabla \times \mathbf{A} = \mathbf{i}_x \left( \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) + \mathbf{i}_y \left( \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) + \mathbf{i}_z \left( \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \tag{3}$$

$$\nabla \Phi = \mathbf{i}_x \frac{\partial \Phi}{\partial x} + \mathbf{i}_y \frac{\partial \Phi}{\partial y} + \mathbf{i}_z \frac{\partial \Phi}{\partial z} \tag{4}$$

In cylindrical coordinates:

$$\mathbf{A} = \mathbf{i}_r A_r + \mathbf{i}_\theta A_\theta + \mathbf{i}_z A_z \tag{5}$$

$$\nabla \cdot \mathbf{A} = \frac{\partial A_r}{\partial r} + \frac{A_r}{r} + \frac{1}{r} \frac{\partial A_\theta}{\partial \theta} + \frac{\partial A_z}{\partial z} \tag{6}$$

$$\nabla \times \mathbf{A} = \mathbf{i}_r \left( \frac{1}{r} \frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z} \right) + \mathbf{i}_\theta \left( \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r} \right) + \mathbf{i}_z \left( \frac{\partial A_\theta}{\partial r} - \frac{1}{r} \frac{\partial A_r}{\partial \theta} + \frac{A_\theta}{r} \right) \tag{7}$$

$$\nabla \Phi = \mathbf{i}_r \frac{\partial \Phi}{\partial r} + \mathbf{i}_\theta \frac{1}{r} \frac{\partial \Phi}{\partial \theta} + \mathbf{i}_z \frac{\partial \Phi}{\partial z} \tag{8}$$

In any coordinate system, the following two theorems are true:
*Gauss's Theorem*

$$\int_{\text{closed surface}} \mathbf{A} \cdot \mathbf{n} dS = \int_{\text{volume}} \nabla \cdot \mathbf{A} dv \tag{9}$$

*Stoke's Theorem*

$$\oint_{\text{closed loop}} \mathbf{A} \cdot d\mathbf{l} = \int_{\text{surface}} (\nabla \times \mathbf{A}) \cdot \mathbf{n} dS \tag{10}$$

The following vector identities may be useful:

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \tag{11}$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) \tag{12}$$

$$\nabla^2 \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}) \tag{13}$$

Equation (13) defines the Laplacian of a vector. In *rectangular* coordinates it is given by

$$\nabla^2 \mathbf{A} = \mathbf{i}_x \nabla^2 A_x + \mathbf{i}_y \nabla^2 A_y + \mathbf{i}_z \nabla^2 A_z \tag{14}$$

where

$$\nabla^2 A_x = \frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_x}{\partial y^2} + \frac{\partial^2 A_x}{\partial z^2} \tag{15}$$

$$\nabla^2 A_y = \frac{\partial^2 A_y}{\partial x^2} + \frac{\partial^2 A_y}{\partial y^2} + \frac{\partial^2 A_y}{\partial z^2} \tag{16}$$

$$\nabla^2 A_z = \frac{\partial^2 A_z}{\partial x^2} \quad \frac{\partial^2 A_z}{\partial y^2} \quad \frac{\partial^2 A_z}{\partial z^2} \tag{17}$$

Detailed proofs of the above relationships are given in texts on vector analysis or advanced calculus. A good discussion is given in J. B. Hildebrand, *Advanced Calculus for Engineers*, Chapter 6, Prentice-Hall, Inc., New York, 1949.

## GROUP VELOCITY AND ENERGY FLOW

The physical significance of the group velocity can be made clearer by considering the propagation of a pulse modulated carrier wave down a lossless transmission line or waveguide of arbitrary length $L$.[1] This line is assumed to propagate the energy in a single mode, which can be characterized by one branch of an $\omega$-$\beta$ diagram of the type shown in Figure 8.5-2.

Let us suppose that some information in the form of a pulse of electromagnetic energy is impressed into the waveguide at the input end. We can ask ourselves the question: How much time will elapse before we can detect this pulse or information at the output end, a distance $L$ away?

Since all of the field components in a single mode are related by Maxwell's Equations, we can study the propagation of just one of them, and this will characterize the behavior of the other components as well. Assume that this
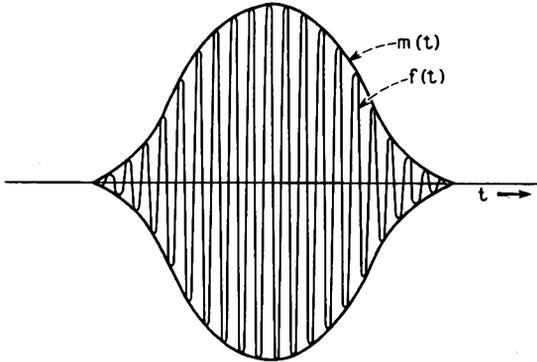


Fig. XIII-1 A high-frequency carrier with a modulation envelope $m(t)$.

component has a time variation $f(t)$ at the input as shown in Figure XIII-1. As is usual with modulation systems, the rate of change of the envelope $m(t)$ is assumed to be very slow compared with the frequency of the carrier. This means that a Fourier analysis of the resultant waveform $f(t)$ would yield frequency components clustered closely to the carrier frequency. Thus $f(t)$ may be written as

$$f(t) = \operatorname{Re} m(t)\epsilon^{i\omega_o t} \tag{1}$$

where $\omega_o$ is the carrier radian frequency.

The modulation envelope $m(t)$ may be written in terms of its Fourier transform

$$m(t) = \int_{-\infty}^{\infty} M(p)\epsilon^{ipt}dp \tag{2}$$

---

[1]Reference 8h, pp. 81–84.

Since any practical modulation system will have a finite bandwidth, this equation can be written

$$m(t) = \int_{-p_o}^{p_o} M(p)\epsilon^{ipt}dp \qquad (3)$$

This expression may be used in Equation (1), obtaining

$$f(t) = \mathrm{Re} \int_{-p_o}^{p_o} M(p)\epsilon^{i(\omega_o+p)t}dp \qquad (4)$$

The quantity $f(t)$ is the superposition of terms of slightly different frequencies, centered about the carrier frequency. Thus, the pulse of energy is a superposition of waves of different frequencies, each with its own propagation constant. Each component has a different phase shift from the input to the output. Since the system is linear and lossless, the output is a superposition of the input frequencies, each one shifted in phase by the proper amount. If $g(t)$ is the time variation of the field component at the output,

$$g(t) = \mathrm{Re} \int_{-p_o}^{p_o} M(p)\epsilon^{i(\omega_o+p)t}\epsilon^{-i\beta L}dp \qquad (5)$$

where $\beta$ is a function of $\omega_o + p$, the frequency. Since all of the frequency components are close to the carrier frequency, the variation of the propagation constant with frequency may be adequately represented by the first term of a Taylor series.

$$\beta = \beta_o + \frac{\partial \beta}{\partial \omega}p \qquad (6)$$

where $\beta_o$ is the propagation constant for the carrier frequency. We thus obtain

$$g(t) = \mathrm{Re}\ \epsilon^{i(\omega_o t-\beta_o L)} \int_{-p_o}^{p_o} M(p)\epsilon^{ip[t-(\partial\beta/\partial\omega)L]}dp \qquad (7)$$

This represents a carrier modulated by a modulation envelope $n(t)$, where

$$n(t) = \int_{-p_o}^{p_o} M(p)\epsilon^{ip[t-(\partial\beta/\partial\omega)L]}dp \qquad (8)$$

Comparing with Equation (3), we see that

$$m(t) = n\left(t + \frac{\partial\beta}{\partial\omega}L\right) \qquad (9)$$

that is, the modulation envelope at the output is exactly reproduced from the input, but at a time $(\partial\beta/\partial\omega)L$ later. This is just as if the pulse of energy had traveled with a velocity

$$v_g = \frac{\partial\omega}{\partial\beta} \qquad (10)$$

This is the physical significance of the group velocity.

The same relationship holds for all of the field components. Hence, if we could visualize the electromagnetic bundle of energy, we would see it move physically with the velocity $v_g$.

These results may be applied to continuous wave propagation at a single frequency. That is, we may think of a continuous wave as being a superposition of rectangular modulated pulses placed end-to-end. It is clear that the power flow in this case is given by

$$P = v_o W_l \tag{11}$$

where $W_l$ is the total energy stored per unit length.

## APPENDIX XIV

### *TIME AVERAGE STORED ENERGY*

The instantaneous magnetic stored energy is given by

$$\mathcal{W}(t) = \tfrac{1}{2}\mu_o \int_{\text{volume}} |\, \mathcal{H}(t)\,|^2 dv \tag{1}$$

The instantaneous magnetic field may be written in the phasor notation

$$\mathcal{H}(t) = \mathrm{Re}\mathbf{H}\epsilon^{j\omega t} \tag{2}$$

where

$$\mathbf{H} = \mathbf{H}_r + j\mathbf{H}_i$$

and $\mathbf{H}_r$ and $\mathbf{H}_i$ are real.  Thus

$$\mathcal{H}(t) = \mathbf{H}_r \cos \omega t - \mathbf{H}_i \sin \omega t \tag{3}$$

Equation (1) may be written

$$\mathcal{W}(t) = \tfrac{1}{2}\mu_o \int_{\text{volume}} (|\,\mathbf{H}_r\,|^2 \cos^2 \omega t + |\,\mathbf{H}_i\,|^2 \sin^2 \omega t - 2\mathbf{H}_r\cdot\mathbf{H}_i \sin \omega t \cos \omega t)dv \tag{4}$$

The time average of this quantity is given by

$$W_{\text{avg}} = \tfrac{1}{4}\mu_o \int_{\text{volume}} [|\,\mathbf{H}_r\,|^2 + |\,\mathbf{H}_i\,|^2]dv \tag{5}$$

But this is simply

$$W_{\text{avg}} = \tfrac{1}{4}\mu_o \int_{\text{volume}} |\,\mathbf{H}\,|^2 dv \tag{6}$$

Now in a periodic structure, since the time average magnetic stored energy equals the time average electric stored energy, the total time average stored energy per cell is given by

$$W_L = \tfrac{1}{2}\mu_o \int_{\text{unit cell}} |\,\mathbf{H}\,|^2 dv \tag{7}$$

## KLYSTRON INTERACTION FOR A HIGH DEGREE OF BUNCHING

In the theory of klystron interaction, large values of the bunching parameter ($X > 1$) result in crossings of the electron trajectories such that a given arrival time near the bunch center corresponds to three different departure times. This behavior is illustrated in Figure 9.1-2.

In Figure XV-1, the curve for $X = 1.5$ is replotted as $\tau_b$ vs. $\tau_a$, where

$$\tau_a = \omega t_o \tag{1}$$

and

$$\tau_b = \omega t - \theta \tag{2}$$

That is, $\tau_a$ and $\tau_b$ are normalized departure and arrival times, respectively. Certain
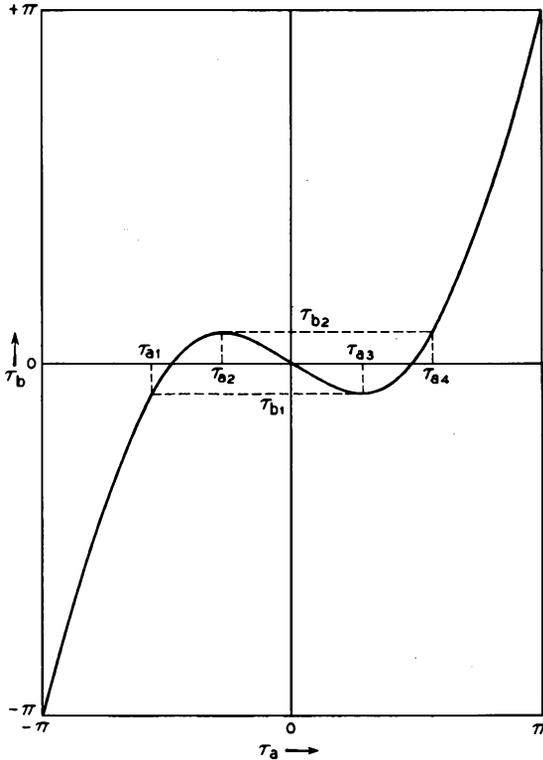


FIG. XV-1 Output gap arrival time in radians plotted vs. the input gap departure time in radians for $X = 1.5$.

points have been labeled on this graph which characterize the limits of the multi-valued branches.

Assuming that the subscript 2 in Equation (9.1-19) refers to the branch with negative slope from $\tau_{a2}$ to $\tau_{a3}$, Equation (9.1-19) may be written as

$$i(t) = -I_o\left[\frac{d\tau_a}{d\tau_b}\bigg|_1 - \frac{d\tau_a}{d\tau_b}\bigg|_2 + \frac{d\tau_a}{d\tau_b}\bigg|_3\right] \tag{3}$$

Equation (9.1-20) for the Fourier coefficient $a_n$ becomes

$$a_n = -\frac{I_o}{\pi}\int_{-\pi}^{\tau_{b1}}\frac{d\tau_a}{d\tau_b}\cos n\tau_b d\tau_b$$

$$-\frac{I_o}{\pi}\int_{\tau_{b1}}^{\tau_{b2}}\left[\frac{d\tau_a}{d\tau_b}\bigg|_1 - \frac{d\tau_a}{d\tau_b}\bigg|_2 + \frac{d\tau_a}{d\tau_b}\bigg|_3\right]\cos n\tau_b d\tau_b$$

$$-\frac{I_o}{\pi}\int_{\tau_{b2}}^{\pi}\frac{d\tau_a}{d\tau_b}\cos n\tau_b d\tau_b \tag{4}$$

Consider the second integral term. When the variable of integration is changed, one has:

$$\int_{\tau_{b1}}^{\tau_{b2}}\frac{d\tau_a}{d\tau_b}\bigg|_1 \cos n\tau_b d\tau_b = \int_{\tau_{a1}}^{\tau_{a2}}\cos n\tau_b d\tau_a \tag{5}$$

$$\int_{\tau_{b1}}^{\tau_{b2}} -\frac{d\tau_a}{d\tau_b}\bigg|_2 \cos n\tau_b d\tau_b = \int_{\tau_{a2}}^{\tau_{a3}} -\cos n\tau_b d\tau_a = \int_{\tau_{a3}}^{\tau_{a2}}\cos n\tau_b d\tau_a \tag{6}$$

and

$$\int_{\tau_{b1}}^{\tau_{b2}}\frac{d\tau_a}{d\tau_b}\bigg|_3 \cos n\tau_b d\tau_b = \int_{\tau_{a3}}^{\tau_{a4}}\cos n\tau_b d\tau_a \tag{7}$$

When Equations (5), (6), and (7) are added together, one obtains the simple result:

$$\int_{\tau_{a1}}^{\tau_{a4}}\cos n\tau_b d\tau_a \tag{8}$$

Equation (4) thus becomes simply:

$$a_n = -\frac{I_o}{\pi}\int_{-\pi}^{\pi}\cos n\tau_b d\tau_a \tag{9}$$

This equation is the same as the first of Equations (9.1-22) which was derived for the case of small bunching parameter $(X < 1)$, and hence it leads to the same Fourier coefficients. The $b_n$ coefficients may be shown to be identical in the same fashion.

# APPENDIX XVI

## A DERIVATION OF THE EXPRESSION FOR THE THERMAL NOISE GENERATED BY A RESISTANCE

Here we derive an expression for the thermal noise generated by a resistance, using as a basis for our derivation the random motions of the charge carriers in the
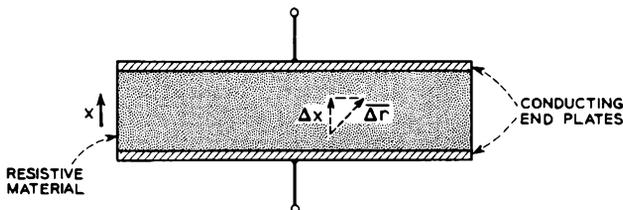


FIG. XVI-1 A cylinder of resistive material with two conducting end plates.

resistance.[1] Figure XVI-1 shows a view of the resistance. We shall assume that the resistance consists of a cylinder of resistive material between two conducting end plates. The end plates are joined by an external wire. We shall further assume that the charge carriers within the resistance have a charge $-e$, a density $n$ per unit volume, a mobility $\mu$, and a diffusion constant $D$. (See Section 14.2 for a discussion of diffusion and mobility.) If a voltage $v$ were applied across the resistance, there would be an electric field $v/L$ within the resistance, where $L$ is the distance from one end plate to the other, and a current

$$i = \frac{v}{L} e\mu n A = \frac{v}{R} \tag{1}$$

would flow through the resistance, where $A$ is the area of the end plates, and $R$ is the magnitude of the resistance. We assume that the linear dimensions of the end plates are sufficiently large in comparison with $L$ that the effects of fringing fields can be neglected. The resistance $R$ is then given by

$$R = \frac{L}{e\mu n A} \tag{2}$$

If there is no applied voltage across the resistance, the motions of the charge carriers is of a random-walk nature, sometimes called Brownian motion. (For a discussion of Brownian motion, see E. H. Kennard, *Kinetic Theory of Gases*, Sections 160–164, McGraw-Hill Book Co., Inc., New York, 1938. A second informative discussion is given by D. K. C. MacDonald, *Noise and Fluctuations: An Introduction*, Section 1.2, John Wiley and Sons, Inc., New York 1962.) On a microscopic scale the individual charge carriers travel along highly irregular paths which are char-

---

[1] This derivation follows that given by K. M. Van Vliet and J. Blok, *Physica* **22**, 231, 1956.

acterized by frequent abrupt changes in direction. For such motion, it can be shown that the mean squared displacement per carrier in one coordinate direction, say the $x$ direction, in time $t$ is given by

$$\overline{\Delta x^2} = 2Dt. \tag{3}$$

A derivation of this equation is given in *Kinetic Theory of Gases*, Section 163.

In Figure XVI-1 let the $x$ direction be parallel to the axis of the cylindrical resistance. If an individual carrier with charge $-e$ travels a distance $\Delta r$ within the resistance, where $\Delta r$ has a component $\Delta x$ in the $x$ direction, there is a net displacement of charge

$$\Delta q = e\frac{\Delta x}{L} \tag{4}$$

in the external circuit, where again it is assumed that the linear dimensions of the end plates are large compared with the distance $L$. This equation is similar to Equation (6.1-7). Squaring both sides of this equation and taking the mean value for a large number of carriers, we obtain

$$\overline{\Delta q^2} = e^2\frac{\overline{\Delta x^2}}{L^2} \tag{5}$$

Combining this with Equation (3), *the mean squared displacement of charge in the external circuit per carrier in time $\Delta t$ is*

$$\overline{\Delta q(\Delta t)^2} = \frac{e^2 2D\Delta t}{L^2} \tag{6}$$

Now the diffusion coefficient is related to the mobility by

$$D = \frac{kT}{e}\mu \tag{7}$$

as in Equation (14.2-12). Combining Equations (2), (6), and (7), we obtain

$$\overline{\Delta q(\Delta t)^2} = \frac{2kT\Delta t}{N_o R} \tag{8}$$

where $N_o = nAL$ is the total number of carriers within the resistance.

If we set $i_e(\Delta t) = \dfrac{\Delta q(\Delta t)}{\Delta t}$, where $i_e(\Delta t)$ is the mean current in time $\Delta t$ resulting from the charge displacement $\Delta q(\Delta t)$, we have

$$\overline{i_e(\Delta t)^2} = \frac{2kT}{N_o R\Delta t} \tag{9}$$

The subscript $e$ is used here to indicate that we are dealing with the motion of an individual carrier. This expression gives the mean of the square of the average current flowing in time $\Delta t$ in the external wire as a result of the motion of an individual carrier within the resistance. By Fourier analysis it can be shown[2] that the

---

[2] C. J. Bakker, *Physica* **5**, 581, July, 1938. See Equation (17).

frequency distribution associated with a rapidly fluctuating current with mean square average value over a time $\Delta t$ of $\overline{i_e(\Delta t)^2}$ is given by

$$\overline{i_e^2} = 2\overline{i_e(\Delta t)^2}\Delta t \Delta f \tag{10}$$

where $\Delta f$ is the bandwidth over which $\overline{i_e(t)^2}$ is measured. Substituting from Equation (9), we obtain

$$\overline{i_e^2} = \frac{4kT}{N_oR}\Delta f \tag{11}$$

Since the motions of the individual carriers are independent of one another, the mean-square current flowing in the external wire as a result of the motions of all $N_o$ carriers is $N_o$ times that given by Equation (11) and hence is given by

$$\overline{i^2} = \frac{4kT}{R}\Delta f \tag{12}$$

# INDEX