

# Expt. 4

## Vacuum Tube & Semi-Conductor Electronics - J. Millman

### CHAPTER 3

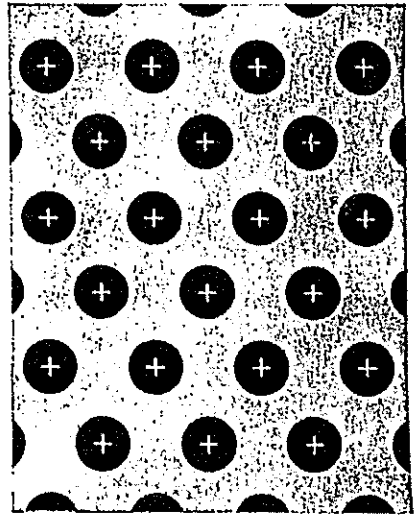
### METALS AND SEMICONDUCTORS

IN THIS chapter we present physical "pictures" of the inside of a metal and a semiconductor. The nature of conduction in a solid and the laws governing the emission of electrons from the surface of a metal are considered.

**3-1. Free Electrons in Metals.** X-ray and other studies reveal that most metals and semiconductors are crystalline in structure. A crystal

consists of a space array of atoms or molecules (strictly speaking, ions) built up by regular repetition in three dimensions of some fundamental structural unit. In a metal the outer electrons of the atom are as much associated with one ion as with another, so that the electron attachment to any individual atom is practically zero. Depending upon the metal, at least one and sometimes two or three electrons per atom are free to move throughout the interior of the metal under the action of applied forces.

Figure 3-1 shows the charge distribution within a metal, specifically, sodium. The plus signs represent the heavy positive sodium nuclei of the individual atoms. The heavily shaded regions represent the electrons in the sodium atom that are tightly bound to the nucleus. These



0 1 2 3 4 5  
A UNITS

FIG. 3-1. Arrangement of the sodium atoms in one plane of the metal. (W. Shockley, *J. Appl. Phys.*, 10, 543, 1939.)

are inappreciably disturbed as the atoms come together to form the metal. The light shading represents the outer, or valence, electrons in the atom; and it is these electrons that cannot be said to belong to any particular atom. Instead, they have completely lost their individuality and can

wander freely about from atom to atom in the metal. Thus a metal is visualized as a region containing a periodic three-dimensional array of heavy, tightly bound ions permeated with a swarm of electrons that may move about quite freely. This picture is known as the "electron-gas" description of a metal.

**3-2. Mobility and Conductivity.** According to the electron-gas theory of a metal, the electrons are in continuous motion, the direction of flight being changed at each collision with the heavy (almost stationary) ions. The average distance between collisions is called the *mean free path*. Since the motion is random, then, on an average, there will be as many electrons passing through unit area in the metal in any direction as in the opposite direction in a given time. Hence, the average current is zero.

Let us now see how the situation is changed if a constant electric field of magnitude  $\mathcal{E}$  volts per meter is applied to the metal. As a result of this electrostatic force the electrons would be accelerated and the velocity would increase indefinitely with time, were it not for the collisions with the ions. However, at each inelastic collision with an ion an electron loses energy, and a steady-state condition is reached where a finite value of *drift speed*  $v$  is attained. This drift velocity is in the direction opposite to that of the electric field, and its magnitude is proportional to  $\mathcal{E}$ . Thus,

$$v = \mu \mathcal{E} \quad (3-1)$$

where  $\mu$  square meters per volt-second is called the *mobility* of the electrons.

According to the above theory, a steady-state drift speed has been superimposed upon the random thermal motion of the electrons. Such a directed flow of electrons constitutes a current. If the concentration of free electrons is  $n$  electrons per cubic meter, then the current density  $J$  amperes per square meter is (Sec. 1-10)

$$J = nev = ne\mu\mathcal{E} = \sigma\mathcal{E} \quad (3-2)$$

where

$$\sigma = ne\mu \quad (\text{ohm-meter})^{-1} \quad (3-3)$$

is the *conductivity* of the metal. Equation (3-2) is recognized as Ohm's law: namely, the conduction current is proportional to the applied voltage. As already mentioned, the energy which the electrons acquire from the applied field is, as a result of collisions, given to the lattice ions. Hence, power is dissipated within the metal by the electrons, and the power density (Joule heat) is given by  $J\mathcal{E} = \sigma\mathcal{E}^2$  watts per cubic meter.

**3-3. The Energy Method of Analyzing the Motion of a Particle.** A method is considered in Chap. 1 by which the motion of charged particles may be analyzed. It consists of the solution of Newton's second law in which the forces of electric and magnetic origin are equated to the product

of the mass and the acceleration of the particle. Obviously, this method is not applicable when the forces are as complicated as they must be in a metal. Furthermore, it is neither possible nor desirable to consider what happens to each individual electron.

It is necessary, therefore, to consider an alternative approach. This method employs the law of the conservation of energy, use being made of the potential-energy curve corresponding to the field of force. The principles involved may best be understood by considering specific examples of the method.

*Example.* An idealized diode consists of plane-parallel electrodes, 5 cm apart. The anode *A* is maintained 10 volts negative with respect to the cathode *K*. An electron leaves the cathode with an initial energy of 2 ev. What is the maximum distance it can travel from the cathode?

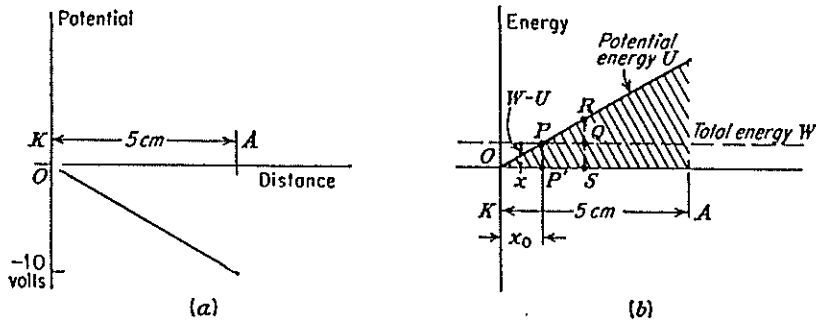


FIG. 3-2. Illustration of the potential-energy barrier encountered by an electron in a retarding field.

*Solution.* This problem will be analyzed by the energy method. Figure 3-2a is a linear plot of potential vs. distance, and in Fig. 3-2b is indicated the corresponding potential energy vs. distance. Since potential is the potential energy per unit charge (Sec. 1-4), curve *b* is obtained from curve *a* by multiplying each ordinate by the charge on the electron (a negative number). Since the total energy *W* of the electron remains constant, it is represented as a horizontal line. The kinetic energy at any distance *x* equals the difference between the total energy *W* and the potential energy *U* at this point. This difference is greatest at *O*, indicating that the kinetic energy is a maximum when the electron leaves the cathode. At the point *P* this difference is zero, which means that no kinetic energy exists, so that the particle is at rest at this point. This distance, *x*<sub>0</sub>, is the maximum that the electron can travel from the cathode. At point *P* it comes momentarily to rest and then reverses its motion and returns to the cathode. From geometry it is seen that  $x_0/5 = \frac{2}{10}$  or  $x_0 = 1$  cm.

Consider a point such as *S* which is at a greater distance than 1 cm from the cathode. Here the total energy *QS* is less than the potential energy *RS*, so that the difference, which represents the kinetic energy, is negative. This is an impossible physical condition, however, since negative kinetic energy ( $\frac{1}{2}mv^2 < 0$ ) implies an imaginary velocity. We must conclude that the particle can never advance a distance greater than *OP* from the cathode.

The foregoing analysis leads to the very important conclusion that the shaded portion of Fig. 3-2b can never be penetrated by the electron. Thus, at point *P* the particle

acts as if it had collided with a solid wall, hill, or barrier and the direction of its flight had been altered. Potential energy barriers of this sort will play important roles in the analyses to follow.

It must be emphasized that the words "collides with" or "rebounds from" a potential "hill" are convenient descriptive phrases and that an actual encounter between two material bodies is not implied.

As a second illustration, consider a mathematical pendulum of length *l*, consisting of a "point" bob of mass *m* that is free to swing in the earth's gravitational field. If the lowest point of the swing (point *O*, Fig. 3-3) is chosen as the origin, then the potential energy of the mass at any point *P* corresponding to any angle  $\theta$  of the swing is given by

$$U = mgy = mgl(1 - \cos \theta) \tag{3-4}$$

where *g* is the acceleration of gravity. This potential-energy function is illustrated graphically in Fig. 3-4.

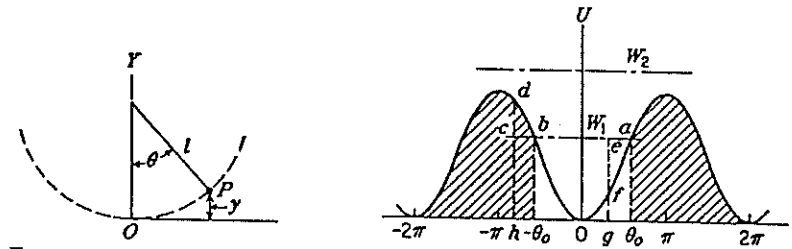


FIG. 3-3. Point *P* represents the mass *m* of a mathematical pendulum swinging in the earth's gravitational field.

FIG. 3-4. The potential energy of the bob in Fig. 3-3 plotted as a function of the angle of swing.

Consider the resultant motion of the bob if it is given a potential energy *U*<sub>1</sub> by raising it through an angle  $\theta_0$  and releasing it with zero initial velocity. If dissipation is neglected, the particle will swing back and forth through the angle  $2\theta_0$ , going from  $\theta_0$  on one side to  $\theta_0$  on the other side of the vertical axis. How might we analyze the motion of the physical system if only the potential-energy field of Fig. 3-4 were given without specifying the physical character of the system?

The procedure is the same as that followed in the simple diode problem considered above. A horizontal line *abc* is drawn at a height equal to the total energy *W*<sub>1</sub> of the particle. At any point, such as *e*, the total energy is represented by *eg* = *W*<sub>1</sub>, and the potential energy is represented by *fg*. The difference between these two, namely, *ef*, represents the kinetic energy of the particle when the angle of swing, given by the intercept of *eg* on the axis, corresponds to *Og*. In other words, the difference between the total-energy line and the potential-energy curve at any angle represents the kinetic energy of the particle under these conditions. This

difference is greatest at  $O$ , indicating that the kinetic energy is a maximum at the bottom of the swing, an almost evident result. At the points  $a$  and  $b$  this difference is zero. This condition means that no kinetic energy exists, or that the particle is at rest at these points. This result is evident, since corresponding to the points  $a$  ( $\theta = \theta_0$ ) and  $b$  ( $\theta = -\theta_0$ ), the particle is about to reverse its motion.

Consider a point in the shaded region outside the range  $-\theta_0$  to  $+\theta_0$ , such as  $h$ . Here the total energy  $ch$  is less than the potential energy  $dh$ . This impossible condition is interpreted by our previous reasoning to mean that the particle whose total energy is  $W_1$  can never swing to the angle  $Oh$ , so that the motion must be confined to the region  $ab$ . The shaded portions of Fig. 3-4 represent the potential barrier which can never be penetrated by the bob, if its total energy is no greater than  $W_1$ . This type of constrained motion about a point  $O$  is closely analogous to that of the so-called "bound" electrons in a metal, as will be seen later.

Now consider the case when the bob has a total energy equal to  $W_2$ , which is greater than the maximum of the potential-energy curve. Clearly from Fig. 3-4 the horizontal line corresponding to this energy cannot intersect the curve at any point. Consequently, the particle does not "collide" with the potential barrier, and its course is never altered, so that it moves through an ever-increasing angle. Of course, its kinetic energy varies over wide limits, being maximum for  $\theta = 0, 2\pi, 4\pi, \dots$  and minimum for  $\theta = \pi, 3\pi, 5\pi, \dots$ . Physically, this type of motion results when the bob has enough energy to set it spinning completely around in a circular path. This type of motion is somewhat analogous to that experienced by the so-called "free" electrons in a metal.

This simple but powerful energy method facilitates the discussion of the motion of a particle in a conservative field of force, such as that found in the body of a metal. It will also be applied to many other types of problem. For example, the method of analysis just considered is extremely useful in determining whether electrons will possess sufficient energy to pass through grids and reach the various electrodes in a vacuum tube, whether or not electrons or ions will be able to penetrate electron clouds in a vacuum tube or ion sheaths in a gaseous-discharge tube, and whether charge carriers can cross a semiconductor junction. This method will now be applied to the analysis of the motion of electrons in metals.

**3-4. The Potential-energy Field in a Metal.** It is desired to set up the potential-energy field for the three-dimensional array of atoms that exists in the interior of a metal and to discuss the motion of electrons in this field. The resultant potential energy at any point in the metal is simply the sum of the potential energies produced at this point by all the ions of the lattice. To determine the potential energy due to one ion, it is noted that

an atom of atomic number  $Z$  has a net positive charge  $Ze$  on its nucleus. Surrounding this nucleus is an approximately spherical cloud, or shell, of  $Z$  electrons. By Gauss's law the potential at a point at a distance  $r$  from the nucleus varies inversely as  $r$  and directly as the total charge enclosed within a sphere of radius  $r$ . Since the potential  $V$  equals the potential energy  $U$  per unit charge (Sec. 1-4), then  $U = -eV$ . The minus sign is introduced since  $e$  represents the magnitude of the (negative) electronic charge.

The potential of any point may be chosen as the zero reference of potential because it is only differences of potential that have any physical significance. For the present discussion it is convenient to choose zero potential at infinity, and then the potential energy at any point is negative. Enough has been said to make plausible the potential-energy

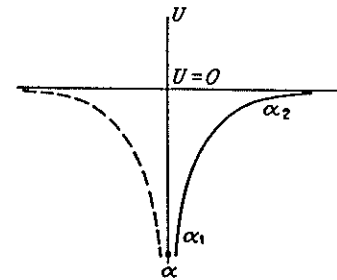


FIG. 3-5. The potential energy of an electron as a function of radial distance from an isolated nucleus.

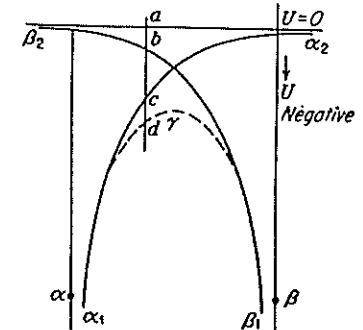


FIG. 3-6. The potential energy resulting from two nuclei  $\alpha$  and  $\beta$ .

curve illustrated in Fig. 3-5. Here  $\alpha$  represents a nucleus, the potential energy of which is given by the curve  $\alpha_1\alpha_2$ . The vertical scale represents  $U$ , and the horizontal scale gives the distance  $r$  from the nucleus. It must be emphasized that  $r$  represents a radial distance from the nucleus and hence can be taken in any direction. If the direction is horizontal but to the left of the nucleus, then the dashed curve represents the potential energy.

To represent the potential energy at every point in space requires a four-dimensional picture, three dimensions for the three space coordinates and a fourth for the potential-energy axis. This difficulty is avoided by plotting  $U$  along some chosen line through the crystal, say through a row of ions. From this graph and the method by which it is constructed it is easy to visualize what the potential energy at any other point might be. In order to build up this picture, consider first two adjacent ions, and neglect all others. The construction is shown in Fig. 3-6.  $\alpha_1\alpha_2$  is the  $U$  curve for nucleus  $\alpha$ , and  $\beta_1\beta_2$  is the corresponding  $U$  curve for the adjacent

nucleus  $\beta$ . If these were the only nuclei present in the metal, the resultant  $U$  curve in the region between  $\alpha$  and  $\beta$  would be the sum of these two curves, as shown by the dashed curve  $\alpha_1\gamma\beta_1$  (since  $ad = ab + ac$ ). It is seen that the resultant curve is very nearly the same as the original curves in the immediate vicinity of a nucleus, but it is lower and flatter than either individual curve in the region between the nuclei.

Let us now single out an entire row of nuclei  $\alpha, \beta, \gamma, \delta, \epsilon, \dots$  from the metallic lattice (Figs. 3-1 and 3-7) and sketch the potential energy as we proceed along this line from one nucleus to the other, until the surface of the metal is reached. Following the same type of construction as above, but considering the influence of other nearby nuclei, an energy distribution somewhat as illustrated in Fig. 3-7 is obtained.

According to classical electrostatics, which does not take the atomic structure into account, the interior of a metal is an equipotential region.

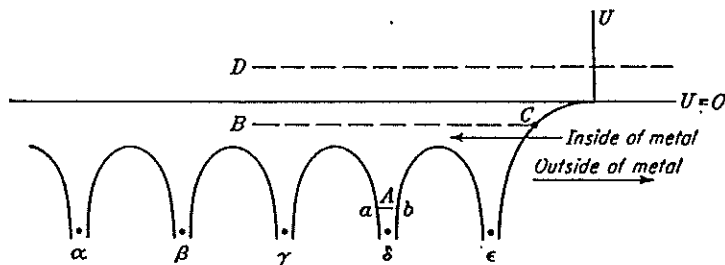


FIG. 3-7. The potential-energy distribution within and at the surface of a metal.

The present, more accurate, picture shows that the potential energy varies appreciably in the immediate neighborhoods of the nuclei and actually tends to  $-\infty$  in these regions. However, the potential is approximately constant for the greatest volume of the metal, as indicated by the slowly varying portions of the diagram in the regions between the ions.

Consider the conditions that exist near the surface of the metal. It is evident, according to the present point of view, that the exact position of the "surface" cannot be defined. It is located at a small distance from the last nucleus  $\epsilon$  in the row. It is to be noted that, since no nuclei exist to the right of  $\epsilon$ , there can be no lowering and flattening of the potential-energy curve such as prevails in the region between the nuclei. This leads to a most important conclusion, viz.: A potential-energy "hill," or "barrier," exists at the surface of the metal.

**3-5. Bound and Free Electrons.** The motion of an electron in the potential-energy field of Fig. 3-7 will now be discussed by the method given in Sec. 3-3. Consider an electron in the metal that possesses a total energy corresponding to the level  $A$  in Fig. 3-7. This electron collides with, and rebounds from, the potential walls at  $a$  and  $b$ . It cannot drift

very far from the nucleus but can move about only in the neighborhood  $ab$  of the nucleus. Obviously this electron is strongly bound to the nucleus and so is called a *bound electron*. It is evident that these bound electrons contribute very little to the conductivity of the metal since they cannot drift in the metal, even under the stimulus of an externally applied electric field. These electrons are responsible for the heavy shading in the neighborhood of the nuclei of Fig. 3-1.

Our present interest is in the *free*, or *conduction*, electrons in the metal rather than in the bound ones. A free electron is one having an energy corresponding to the level  $B$  of the figure. At no point within the metal is its total energy entirely converted into potential energy. Hence, at no point is its velocity zero, and the electron travels more or less freely throughout the body of the metal. However, when the electron reaches the surface of the metal, it collides with the potential-energy barrier there. At the point  $C$ , its kinetic energy is reduced to zero, and the electron is turned back into the body of the metal. An electron having an energy corresponding to the level  $D$  collides with no potential walls, not even the one at the surface, and so it is capable of leaving the metal.

In our subsequent discussions the bound electrons will be neglected completely since they in no way contribute to the phenomena to be studied. Attention will be focused on the free electrons. The region in which they find themselves is essentially a potential plateau, or equipotential region. It is only for distances close to an ion that there is any appreciable variation in potential. Since the regions of rapidly varying potential represent but a very small portion of the total volume of the metal, we shall henceforth assume that the field distribution within the metal is equipotential and the free electrons are subject to no forces whatsoever. The present viewpoint is therefore essentially that of classical electrostatics.

Figure 3-7 is redrawn in Fig. 3-8, all potential\* variations within the metal being omitted, with the exception of the potential barrier at the surface. For the present discussion, the zero of energy is chosen at the level of the plateau of this diagram. This choice of the zero-energy

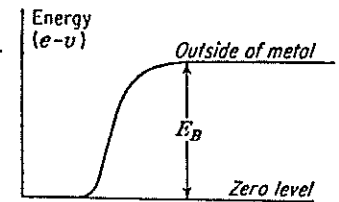


FIG. 3-8. For the free electrons, the interior of a metal may be considered an equipotential volume, but there is a potential barrier at the surface.

\* This figure really represents potential energy and not potential. However, the phrase "potential barrier" is much more common in the literature than the phrase "potential-energy barrier." Where no confusion is likely to arise, these two expressions will be used interchangeably. These barriers will be measured in electron volts, and hence the symbol  $E$  will replace the  $U$  of the preceding sections. It must be emphasized that one unit of  $E$  represents  $1.60 \times 10^{-19}$  joule of energy.

reference level is valid since, as has already been emphasized, only difference of potential has physical significance. The region outside the metal is now at a potential equal to  $E_B$ , the height of the potential-energy barrier in electron volts.

**3-6. Energy Distribution of Electrons.** In order to be able to escape, an electron inside the metal must possess an amount of energy at least as great as that represented by the surface barrier  $E_B$ . It is therefore important to know what energies are possessed by the electrons in a metal. This relationship is called the energy distribution function. We shall digress briefly in order to make clear what is meant by a distribution function.

Suppose that we were interested in the distribution in age of the people in the United States. A sensible way to indicate this relationship is

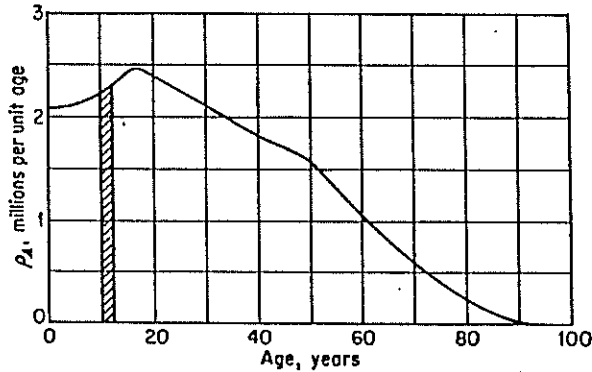


FIG. 3-9. The distribution function in age of the people in the United States.

shown in Fig. 3-9, where the abscissa is *age* and the ordinate is  $\rho_A$ , the *density* of the population in age. This density gives the number  $dN_A$  of people whose ages lie in the range between  $A$  and  $A + dA$ , or

$$dN_A = \rho_A dA \tag{3-5}$$

The data for such a plot are obtained from census information. We see, for example, that the number of persons having ages between 10 and 12 years is represented by  $dN_A$ , with  $\rho_A = 2.25$  million per year chosen as the mean ordinate between 10 and 12 years, and  $dA$  is taken as  $12 - 10 = 2$  years. Thus  $dN_A = \rho_A dA = 4.50$  million. Geometrically, this is the shaded area of Fig. 3-9. Evidently, the total population  $N$  is given by

$$N = \int dN_A = \int \rho_A dA \tag{3-6}$$

or simply the total area under the curve.

We shall now be concerned with the distribution in energy of the free electrons in a metal. By analogy with Eq. (3-5), we may write

$$dN_E = \rho_E dE \tag{3-7}$$

where  $dN_E$  represents the number of free electrons per cubic meter whose energies lie in the energy interval  $dE$  electron volts and where  $\rho_E$  gives the density of electrons in this interval. The question that immediately presents itself is: What is the mathematical expression for the density function  $\rho_E$ ? Fermi,<sup>1</sup> and independently Dirac,<sup>2</sup> taking into account the quantum nature of the electron, other physical facts, and the laws of probability, deduced this most probable distribution function<sup>3</sup> for electrons. The application of this statistics to the theory of metals is due primarily to Sommerfeld. The Fermi-Dirac-Sommerfeld energy density function may be expressed in the form

$$\rho_E = \frac{\gamma E^{\frac{1}{2}}}{1 + e^{(E-E_M)/E_T}} \quad (\text{electrons/m}^3)/\text{ev} \tag{3-8}$$

where  $\gamma$  is a constant defined by

$$\gamma = \frac{4\pi}{h^3} (2mc)^{\frac{1}{2}} \quad (\text{electrons/m}^3)/(\text{ev})^{\frac{1}{2}}$$

and where  $m$  is the mass of the electron in kilograms,  $h$  is a constant (its dimensions are joule-seconds) first introduced by Planck,  $e$  is the base of natural logarithms,  $E$  is the energy of the electron in electron volts,  $E_M$  is a parameter to be discussed later, and  $E_T$  is defined by the relationship

$$eE_T = kT \tag{3-9}$$

where  $k$  is the Boltzmann gas constant in joules per degree Kelvin,  $T$  is the temperature in degrees absolute or Kelvin, and  $e$  is the electronic charge in coulombs. The quantity  $E_T$  is called the electron-volt equivalent of temperature<sup>4</sup> and is a convenient abbreviation. The numerical values of the physical constants introduced here are contained in Appendix I. Equation (3-9) becomes, upon substituting numerical values for the constants contained in the equation,

$$E_T = \frac{T}{11,600} \tag{3-10}$$

This permits a rapid conversion from temperature to the electron-volt equivalent.

Several points must be emphasized before discussing Eq. (3-8). Since our interests are confined only to the free electrons, it will be assumed that there are no potential variations within the metal. Hence, there must be, a priori, the same number of electrons in each cubic meter of the metal. That is, the density in space (electrons per cubic meter) is a constant. However, within each unit volume of metal there will be electrons having all possible energies. It is this distribution in energy (per cubic meter of the metal) that is expressed by Eq. (3-8).

At a temperature of absolute zero, Eq. (3-8) attains a very striking form known as the *completely degenerate function*. When  $T = 0^\circ\text{K}$ , then  $E_T = 0$ , and two possible conditions exist: (1) If  $E > E_M$ , then the exponential term becomes infinite, whence  $\rho_E = 0$ . Consequently, there are no electrons with energies greater than  $E_M$  at absolute zero of temperature. That is,  $E_M$  is the *maximum energy that any electron may possess at absolute zero.* This important quantity  $E_M$  is often referred to as the "Fermi characteristic energy" or the "Fermi level." (2) If  $E < E_M$ , then the exponential in Eq. (3-8) becomes zero. Hence

$$\left. \begin{aligned} \rho_E &= \gamma E^{\frac{1}{2}} & \text{for } E < E_M \\ \rho_E &= 0 & \text{for } E > E_M \end{aligned} \right\} \text{ when } E_T = 0 \quad (3-11)$$

A plot of the distribution in energy given by Eqs. (3-8) and (3-11) for metallic tungsten at  $T = 0^\circ\text{K}$  and  $T = 2500^\circ\text{K}$  is shown in Fig. 3-10. The area under each curve is simply the total number of particles per cubic meter of the metal, whence the two areas must be equal. Also, the curves for all temperatures must pass through the same ordinate, namely,  $\rho_E = \gamma E_M^{\frac{1}{2}}/2$ , at the point  $E = E_M$ , as is evident from Eq. (3-8).

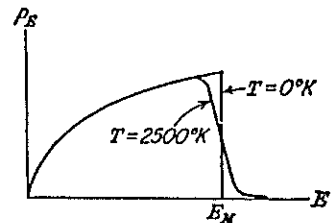


FIG. 3-10. Energy distribution in metallic tungsten at  $0^\circ$  and  $2500^\circ\text{K}$ .

A most important characteristic is to be noted, *viz.*, the distribution function changes only very slightly with temperature, even though the temperature change is as great as  $2500^\circ\text{K}$ . The effect of the high temperature is merely to give those electrons having the high energies at absolute zero (those in the neighborhood of  $E_M$ ) still higher energies, whereas those having lower energies have been left practically undisturbed. Since the curve for  $T = 2500^\circ\text{K}$  approaches the energy axis asymptotically, a few electrons will have large values of energy.

An expression for  $E_M$  may be obtained on the basis of the completely degenerate function. The area under the curve of Fig. 3-10 represents the total number of free electrons (as always, per cubic meter of the metal). Thus

$$N = \int_0^{E_M} \gamma E^{\frac{1}{2}} dE = \frac{2}{3} \gamma E_M^{\frac{3}{2}}$$

or

$$E_M = \left( \frac{3N}{2\gamma} \right)^{\frac{2}{3}} \quad \text{ev} \quad (3-12)$$

Inserting the numerical value ( $6.82 \times 10^{27}$ ) of the constant  $\gamma$  in this expression, there results

$$E_M = 3.64 \times 10^{-19} N^{\frac{2}{3}} \quad \text{ev} \quad (3-13)$$

Since the density  $N$  (electrons per cubic meter) varies from metal to metal, then  $E_M$  will also vary among metals. Knowing the specific gravity, the atomic weight, and the number of free electrons per atom, it is a simple matter to calculate  $N$ , and so  $E_M$ . For most metals the numerical value of  $E_M$  is less than 10 ev.

*Example.* The specific gravity of tungsten is 18.8, and its atomic weight is 184.0.\* Assume that there are two free electrons per atom. Calculate the numerical values of  $N$  and  $E_M$ .

*Solution.* A quantity of any substance equal to its molecular weight in grams is a mole of that substance. Further, 1 mole of any substance contains the same number of molecules as 1 mole of any other substance. This number is *Avogadro's number* and equals  $6.02 \times 10^{23}$  molecules per mole. Thus

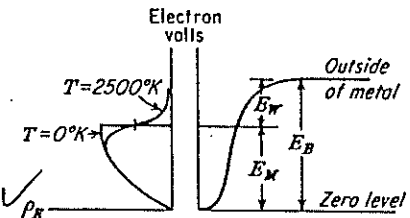
$$N = 6.02 \times 10^{23} \frac{\text{molecules}}{\text{mole}} \times \frac{1 \text{ mole}}{184 \text{ g}} \times 18.8 \frac{\text{g}}{\text{cm}^3} \times \frac{2 \text{ electrons}}{\text{atom}} \times \frac{1 \text{ atom}}{\text{molecule}}$$

$$= 12.3 \times 10^{22} \frac{\text{electrons}}{\text{cm}^3} = 1.23 \times 10^{23} \frac{\text{electrons}}{\text{m}^3}$$

since for tungsten the atomic and the molecular weights are the same. Therefore, for tungsten,

$$E_M = 3.64 \times 10^{-19} (123 \times 10^{22})^{\frac{2}{3}} = 8.95 \text{ ev}$$

**3-7. Work Function.** In Fig. 3-11, Fig. 3-10 has been rotated 90 deg counterclockwise and combined with Fig. 3-8 so that the vertical axis represents energy for both sets of curves.<sup>6</sup> At  $0^\circ\text{K}$  it is impossible for an electron to escape from the metal because this requires an amount of energy equal to  $E_B$  and the maximum energy possessed by any electron is only  $E_M$ . It is necessary to supply an additional amount of energy equal to the difference between  $E_B$  and  $E_M$  in order to make this escape possible. This difference, written  $E_W$ , is known as the work function of the metal.



$$E_W = E_B - E_M \quad (3-14)$$

Thus the work function of a metal represents the minimum amount of energy that must be given to the fastest-moving electron at the absolute zero of temperature in order for this electron to be able to escape from the metal.

FIG. 3-11. Energy diagram used to define the work function.

The experiments of Davisson and Germer<sup>6</sup> and of Rupp<sup>7</sup> on the diffraction of electrons in passing through matter have verified the existence of the potential-energy barrier at the surface of the metal. In fact, based on the results of these experiments together with experimentally determined values of  $E_W$ , it is possible to calculate the values of  $E_M$  for the

\* The atomic weights of the elements are given in the periodic table (Appendix III).

metals used. These data show fair agreement between the experimental and theoretical values.

A second physical meaning of the term "work function" may be obtained by considering what happens to an electron as it escapes from a metal, without particular regard to the conditions within the interior of the metal. A negative electron will induce a positive charge on a metal from which it escapes. There will then be a force of attraction between the induced charge and the electron. Unless the electron possesses sufficient energy to carry it out of the region of influence of this image force of attraction, it will be returned to the metal. The energy required for the electron to escape from the metal is the work function  $E_W$  (based upon this classical electrostatic model).

**3-8. Thermionic Emission.** The curves of Fig. 3-11 show that the electrons in a metal at absolute zero are distributed among energies which range in value from zero to the maximum energy  $E_M$ . Since an electron must possess an amount of energy at least as great as  $E_B$  in order to be able to escape, no electrons can leave the metal. Suppose now that the metal, in the form of a filament, is heated by sending a current through it. Thermal energy is then supplied to the electrons from the lattice of the heated metal crystal. The distribution of the electrons changes, owing to the increased temperature, as indicated in Fig. 3-11. As the temperature is raised, some of the electrons represented by the tail of the curve of Fig. 3-11 will have energies greater than  $E_B$  and so may be able to escape from the metal.

Using the analytical expression from the distribution function, it is possible to calculate the number of electrons which strike the surface of the metal per second with sufficient energy to be able to surmount the surface barrier and hence escape. Based upon such a calculation,<sup>8</sup> the thermionic current is given by

$$I_{th} = SA_0 T^2 e^{-E_W/E_T} \quad \text{or} \quad I_{th} = SA_0 T^2 e^{-b_0/T} \quad \text{amp} \quad (3-15)$$

where  $S$  = area of filament,  $m^2$

$A_0$  = constant whose dimensions are amp/( $m^2$ )( $^{\circ}K^2$ )

$T$  = temperature,  $^{\circ}K$

$E_T \equiv T/11,600$  is defined in Eq. (3-9)

$E_W$  = work function, eV

$b_0 \equiv 11,600E_W$ ,  $^{\circ}K$

Equations (3-15) are two forms of the equation of thermionic emission. They are sometimes referred to as the "Dushman equations" and sometimes as the "Richardson equations," since both workers developed equations of this form theoretically. The constant  $E_W$ , which has been termed the "work function," is known also as the "latent heat of evaporation

of electrons" from the metal, from the analogy of electron emission with the evaporation of molecules from a liquid.

The thermionic-emission equation has received considerable experimental verification.<sup>9</sup> The graphical representation between the thermionic-emission current and the temperature is generally obtained by taking the logarithm of Eq. (3-15), viz.,

$$\log_{10} I_{th} - 2 \log_{10} T = \log_{10} SA_0 - 0.434b_0 \frac{1}{T} \quad (3-16)$$

where the factor 0.434 represents  $\log_{10} e$ . Hence, if we plot  $\log_{10} I_{th} - 2 \log_{10} T$  vs.  $1/T$ , the result should be a straight line having a slope equal to  $-0.434b_0$ . The verification of this equation requires a knowledge of the cathode temperature. In those cases where the cathode is sufficiently exposed, the temperature can most accurately be determined by means of an optical pyrometer. Often, however, it is difficult or entirely impossible to see the cathode. Under these conditions a method that is based upon the energy radiated by the cathode is usually employed.

If a certain amount of power is supplied to a cathode, it will become heated and the temperature will increase until temperature equilibrium occurs. Equilibrium exists when the rate of heat removal by all causes equals the rate of heat produced as a result of the electrical input. Since the cathode is generally a thin filament in a vacuum, no convection of heat can occur. A small amount of heat will be conducted away by the leads, but most of the heat loss is due to the radiated energy. The rate at which energy is radiated from the heated surface is expressed explicitly as a function of the temperature of the body by the Stefan-Boltzmann relation

$$P = 5.67 \times 10^{-8} e_T T^4 \quad \text{watts}/m^2 \quad (3-17)$$

where  $P$  is the power radiated, in watts per square meter, by the surface whose emissivity is  $e_T$ ; the factor  $5.67 \times 10^{-8}$  watt/( $m^2$ )( $^{\circ}K^4$ ) is known as the Stefan-Boltzmann constant; and  $T$  is the temperature in degrees Kelvin. The value of  $e_T$  is always less than unity for all practical cases. It varies slightly with temperature and must be determined experimentally. Forsythe and Worthing<sup>10</sup> and Jones and Langmuir<sup>11</sup> have determined the temperature of tungsten as a function of the input power per square centimeter, over wide ranges of temperature. These data determine  $e_T$ . Hence, by measuring  $P$ , the temperature  $T$  is found.<sup>12</sup>

An early form of an emission equation suggested on the basis of the classical kinetic theory by O. W. Richardson<sup>13</sup> is

$$I_{th} = A' T^4 e^{-b'/T} \quad (3-18)$$

where  $A'$  is a quantity that depends upon the material and  $b'$  is a quantity related to, but not equal to,  $b_0$ . Experimentally, it is impossible to dis-

tinguish between Eqs. (3-15) and (3-18). This difficulty arises from the fact that both equations predict the same *exponential* dependence upon the temperature; and since this factor is such a rapidly varying one, it overshadows the dependence upon the  $T^1$  or the  $T^2$  term. For example, it follows from the second of Eqs. (3-15), by taking the derivative of the natural logarithm of this equation, that

$$\frac{dI_{th}}{I_{th}} = \left(2 + \frac{b_0}{T}\right) \frac{dT}{T} \quad (3-19)$$

For tungsten,  $b_0 = 52,400$ , so that at a normal operating temperature of  $2400^\circ\text{K}$  the fractional change in current  $dI_{th}/I_{th}$  is  $2 + 22$  times the fractional change in the temperature. It is to be noted that the term  $22 (= b_0/T)$  arises from the exponential term in the Dushman equation, and the term  $2$  arises from the  $T^2$  term. Because of this slight dependence upon the power of the  $T$  term, it is impossible to use the experimental results as a criterion to favor one or the other equation. We observe in passing that the thermionic current is a very sensitive function of the temperature, since a 1 per cent change in  $T$  results in a 24 per cent change in  $I_{th}$ .

It must be emphasized that Eqs. (3-15) give the electron emission from a metal at a given temperature provided that there are no external fields present. If there are either accelerating or retarding fields at the surface, then the actual current collected will be greater or less than the emission current, respectively. The effect of such surface fields is discussed later in this chapter.

**3-9. Contact Potential.** Consider two metals in contact with each other, as at the junction  $C$  in Fig. 3-12. The contact difference of potential between these two metals is defined as the potential difference  $E_{AB}$ , between a point  $A$  just outside metal 1 and a point  $B$  just outside metal 2. The reason for the existence of the difference of potential is easily understood.

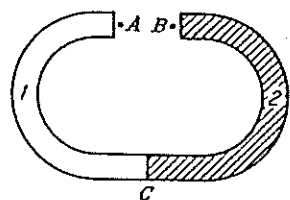


FIG. 3-12. Two metals in contact at the junction  $C$ .

When the two metals are joined at the boundary  $C$ , electrons will flow from the lower-work-function metal, say 1, to the other metal, 2. This flow will continue until metal 2 has acquired so

much negative charge that a retarding field has built up which repels any further electrons. A detailed analysis<sup>14</sup> of the requirement that the number of electrons traveling from metal 1 across junction  $C$  into metal 2 is the same as that in the reverse direction across  $C$  leads to the conclusion that this equilibrium condition is attained when the Fermi energies  $E_M$  of the two metals are located at the same height on the energy-level diagram. To satisfy this condition the potential-energy diagram for the

two metals must be drawn as in Fig. 3-13. The barriers at the two surfaces  $A$  and  $B$  are indicated as vertical lines instead of curves as in Fig. 3-8 because the distance between the surfaces  $A$  and  $B$  is very large in comparison with atomic dimensions.

The diagram should be clear if it is recalled that  $E_W = E_B - E_M$ . From this figure it is seen that

$$E_{AB} = E_{W2} - E_{W1} \quad (3-20)$$

which means that the contact difference of potential between two metals equals the difference between their work functions. This result has been verified experimentally by numerous investigators.

If metals 1 and 2 are similar, the contact potential between them is evidently zero. If they are dissimilar metals, the metal having the lower

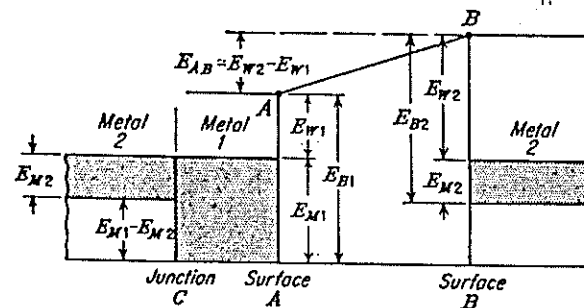


FIG. 3-13. The potential-energy system of two metals in contact.

work function becomes charged positively and the higher-work-function metal becomes charged negatively. In a vacuum tube the cathode is usually the lowest-work-function metal. If it is connected to any other electrode externally by means of a wire, then the effective voltage between the two electrodes is not zero but equals the difference in the work functions. This potential difference is in such a direction as to repel the electrons being emitted from the cathode. If a battery is connected between the two electrodes, then the effective potential is the algebraic sum of the applied voltage and the contact potential.

**3-10. Energies of Emitted Electrons.** Since the electrons inside a metal have a distribution of energies, then those which escape from the metal will also have an energy distribution. It is easy to demonstrate this experimentally. Thus consider a plane emitter and a plane-parallel collector. The current is measured as a function of the retarding voltage  $E_r$  (the emitter positive with respect to the collector). If all the electrons left the cathode with the same energy, then the current would remain constant until a definite voltage was reached and then it would fall abruptly to zero. For example, if they all had 2 eV energy, then when



the retarding voltage was greater than 2 volts the electrons could not surmount the potential barrier between cathode and anode and no particles would be collected. Experimentally no such sudden falling off of current is found, but instead there is an exponential decrease of current  $I_b$  with voltage according to the equation

$$I_b = I_{th} e^{-E_r/E_T} \quad (3-21)$$

This result may be obtained theoretically as follows: Since  $I_{th}$  is the current for zero retarding voltage, then the current obtained when the barrier height is increased by  $E_r$  is determined from the right-hand side of the first of Eqs. (3-15) by changing  $E_W$  to  $E_W + E_r$ . Hence,

$$I_b = S A_0 T^2 e^{-(E_W + E_r)/E_T} = I_{th} e^{-E_r/E_T}$$

If  $E_b$  is the applied (accelerating) anode potential and if  $E'$  is the (retarding) contact potential, then  $E_r = E' - E_b$ , and Eq. (3-21) becomes

$$I_b = I e^{+E_b/E_T} \quad (3-22)$$

where

$$I \equiv I_{th} e^{-E'/E_T} \quad (3-23)$$

represents the current which is collected at zero applied voltage. Since  $E' > E_T$ , this current  $I$  is a small fraction of  $I_{th}$ . If  $E_b$  is increased from zero, the current  $I_b$  increases exponentially until the magnitude of the applied voltage  $E_b$  equals the contact potential  $E'$ . At this voltage

$E_r = 0$ , and the thermionic current is collected. If  $E_b > E'$ , then the field acting on the emitted electrons is in the accelerating direction and the current remains at the value  $I_{th}$ . A plot of the term  $\log_{10} I$  vs.  $E_b$  should be of the form shown in Fig. 3-14. The nonzero slope of this broken-line curve is  $(11,600 \log_{10} e)/T = 5,030/T$ .

From the foregoing considerations, the potential represented by the distance from  $O$  to  $O'$  is the contact potential  $E'$ . Because most commercial diodes do not even approxi-

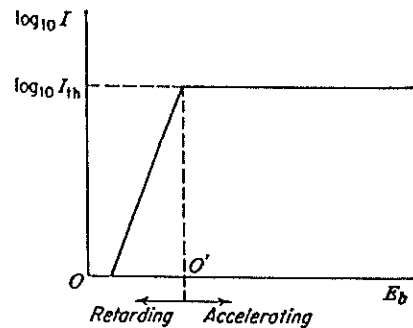


FIG. 3-14. To verify the retarding-potential equation,  $\log_{10} I$  is plotted vs.  $E_b$ .

mate a plane cathode with a plane-parallel anode, the volt-ampere characteristic indicated in Fig. 3-14 is only approached in practice. Furthermore, since the effect of space charge (Chap. 4) has been completely neglected, Eq. (3-21) is valid only for low values (microamperes) of current. For larger values of  $I_b$ , the current varies as the three-halves power of the plate potential (Sec. 4-4).

*Example.* What percentage of the electrons leaving a tungsten filament at 2700°K can surmount a barrier whose height is 1 ev?

*Solution.* Using Eq. (3-21), with  $E_r = 1$ , and remembering that  $E_T = T/11,600$  yields

$$\frac{I_b}{I_{th}} = e^{-(11,600 \times 1)/2,700} = e^{-4.3} \cong 0.014$$

Hence, only about 1.4 per cent of the electrons have energies in excess of 1 ev.

If the emitter is an oxide-coated cathode operating at 1000°K, then a calculation similar to the above gives the result that only about 0.001 per cent of the electrons have a surface-directed energy in excess of 1 ev!

A statistical analysis<sup>16</sup> shows that the average energy of the escaping electrons is given by the expression

$$\bar{E} = 2E_T \quad \text{ev} \quad (3-24)$$

For operating temperatures of 2700° and 1000°K the average energies of the emitted electrons are 0.47 and 0.17 ev, respectively.

These calculations demonstrate the validity of the assumption made in Chap. 1 in the discussion of the motion of electrons in electric and magnetic fields, *viz.*, that the electrons begin their motions with very small initial velocities. In most applications the initial velocities are of no consequence, but they are of significance in tubes which are operated at low electrode voltages.

**3-11. Accelerating Fields.** Under normal operating conditions, the field applied between the cathode and the collecting anode is accelerating rather than retarding, and so the field aids the electrons in overcoming the image force at the surface of the metal. This accelerating field tends, therefore, to lower the work function of the metal and so results in an increased thermionic emission from the metal. It can be shown<sup>16</sup> that the current  $I$  under the condition of an accelerating field of 8 volts per meter at the surface of the emitter is

$$I = I_{th} e^{+0.440e^{1/2} T} \quad (3-25)$$

where  $I_{th}$  is the zero-field thermionic current and  $T$  is the cathode temperature in degrees Kelvin. The fact that the measured thermionic currents continue to increase as the applied potential between the cathode and the anode is increased is often referred to as the *Schottky effect*, after the man who first predicted this effect. Some idea of the order of magnitude of this increase can be obtained from the following illustration.

*Example.* Consider a cylindrical cathode of radius 0.01 cm and a coaxial cylindrical anode of radius 1.0 cm. The temperature of the cathode is 2500°K. If an accelerating potential of 500 volts is applied between the cathode and the anode, calculate the percentage increase in the zero-external-field thermionic-emission current because of the Schottky effect.

*Solution.* The electric-field intensity at any point  $r$  (meters) in the region between the electrodes of a cylindrical capacitor, according to classical electrostatics, is given by the formula

$$\epsilon = \frac{E_b}{\ln(r_a/r_k)} \frac{1}{r} \quad \text{volts/m} \quad (3-20)$$

where  $\ln$  denotes the logarithm to the natural base  $e$ ,  $E_b$  is the plate voltage,  $r_a$  is the anode radius, and  $r_k$  is the cathode radius. Thus the electric-field intensity at the surface of the cathode is

$$\epsilon = \frac{500}{2.303 \log_{10} 100} \frac{1}{10^{-4}} = 1.085 \times 10^8 \text{ volts/m}$$

It follows from Eq. (3-25) that

$$\log_{10} \frac{I}{I_A} = \frac{(0.434)(0.44)(1.085 \times 10^8)^{\frac{1}{2}}}{2,500} = 0.0705$$

Hence,  $I/I_A = 1.20$ , which shows that the Schottky theory predicts a 20 per cent increase over the zero-field emission current.

**3-12. High-field Emission.** Suppose that the accelerating field at the surface of a "cold" cathode (one for which the thermionic-emission current is negligible) is very intense. Under these circumstances the variation of the emission-current density with the strength of the electric-field intensity at the surface of the metal has been calculated by several investigators.<sup>17</sup> The result obtained by Fowler and Nordheim is

$$J = C\epsilon^2 e^{-D/\epsilon} \quad \text{amp/m}^2 \quad (3-27)$$

where

$$\left. \begin{aligned} C &= \frac{6.2 \times 10^{-8}}{E_B} \left( \frac{E_M}{E_W} \right)^{\frac{1}{2}} \quad \text{amp/volt}^2 \\ D &= 6.8 \times 10^8 E_W^{\frac{1}{2}} \quad \text{volts/m} \end{aligned} \right\} \quad (3-28)$$

This equation has received direct experimental verification.<sup>18</sup> This effect is called *high-field, cold-cathode* or *autoelectronic emission*. The electric-field intensity at an electrode whose geometry includes a sharp point or edge may be very high even if the applied voltage is moderate. Hence, if high-field emission is to be avoided, it is very important to shape the electrodes in a tube properly so that a concentration of electrostatic lines of flux does not take place on any metallic surface. On the other hand, the cold-cathode effect has been used to provide several thousand amperes in an X-ray tube used for high-speed radiography.<sup>19</sup>

**3-13. Secondary Emission.**<sup>20</sup> The number of secondary electrons that are emitted from a material, either a metal or a dielectric, when subjected to electron bombardment has been found experimentally to depend upon a number of factors. Among these are the number of primary electrons, the energy of the primary electrons, the angle of incidence of the electrons on the material, the type of material, and the physical condition of the surface. The secondary-emission ratio, defined as the ratio of the number

of secondary electrons per primary electron, is small for pure metals, the maximum value being between 1.5 and 2. It is increased markedly by the presence of a contaminating layer of gas or by the presence of an electropositive or alkali metal on the surface. For such composite surfaces, secondary-emission ratios as high as 10 or 15 have been detected. This ratio as a function of the energy of the impinging primary electrons on a cesium-antimony and a silver-magnesium surface is shown in Fig. 3-15.

The maximum in the secondary-emission ratio curve can be explained qualitatively. For low-energy primaries, the number of secondaries that are able to overcome the surface attraction is small. As the energy of the impinging electrons increases, more energetic secondaries are produced and the secondary-emission ratio increases. Since, however, the depth of penetration increases with the energy of the incident electron, the secondaries must travel a greater distance in the metal before they reach the surface. This increases the probability of collision in the metal, with a consequent loss of energy of these secondaries. Thus, if the primary energy is increased too much,

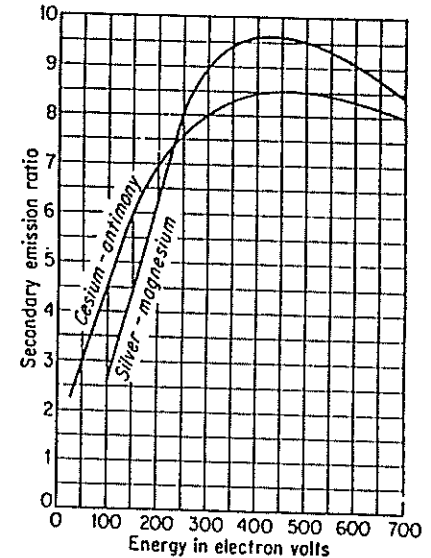


FIG. 3-15. Variation of secondary-emission ratio with primary voltage. (Courtesy of A. B. Du Mont Laboratories, Inc.)

the secondary-emission ratio must pass through a maximum.

Most secondary electrons are emitted with small energies. There is evidence<sup>21</sup> that more than 85 per cent of the secondary electrons emitted from a surface have energies of less than 3 eV. This condition is to be expected since a rapidly moving inner electron should be able to induce the same type of phenomenon as a fast-moving primary electron. The small percentage of high-energy electrons that is present is attributed to those primary electrons which have been reflected from the surface, rather than to true secondary electrons.

It is possible to induce electron emission by bombarding a surface with positive ions instead of with electrons.<sup>22</sup> This process is much less efficient than electron bombardment. As a result, the energies of the impinging ions must be much greater than those of electrons in order to yield a comparable secondary-emission ratio. Nevertheless, this process plays a fundamental role in some types of discharge to be discussed later.