Chapter 14

Electronic Conduction through Thin Insulating Films

JOHN G. SIMMONS

Department of Electrical Engineering, University of Toronto, Toronto, Canada

1. INTRODUCTION

In this chapter we will discuss conduction mechanisms in thin insulating films *sandwiched* between metal electrodes; we are thus concerned with conduction through, rather than along, the plane of the film. Furthermore, since we are concerned with thin films, generally less than 1 or 2 μ thick, it will also be apparent that we will be concerned primarily with the high-field electrical properties of the films, since applied biases as low as even a few volts will induce fields of the order 10^4 to 10^5 V cm^{-1} in the films. Thin films do in fact manifest much more interesting electrical properties when subjected to high fields: in contrast to the low-field properties which are usually ohmic in nature, that is, current I is linear with voltage V, the high-field I-V characteristics are often rich in structure. Quite often, too, the high-field electrical properties cannot be adequately described by a single conduction process; usually the various field-strength ranges manifest different electrical phenomena.

We will not be concerned with ionic conductivity per se in this chapter. Nor will we be discussing any of the unusual effects that one observes if the electrodes are permitted to go superconducting; this subject is discussed in Chap. 22.

The mks system of units has been used throughout in expressing the equations; this avoids the confusion that often arises when working with the mixed emu and esu cgs system of units. Some of the mks units are, however, unwieldy when applied to thin film calculations. For this reason, then, where appropriate, the equations have also been expressed in the more conventional units used by workers in the field, namely, energies in eV; field in V cm^{-1}; distance, area, and volume, respectively, in cm, cm^2, and cm^3.

2. PRELIMINARY COMMENTS

a. Conductivity of Thin Films

An insulator is a material which contains very few volume-generated carriers, in many instances considerably less than one per cm^3, and thus has virtually no conductivity. It will be clear later to the reader that the conductivity of the thin film materials we will discuss does not necessarily fall into this category; that is, although we will be concerned with materials which have energy gaps greater than about 2 eV or so, the electrical properties may bear no resemblance to what is *intrinsically* expected of such a material. This is because it is becoming increasingly clear that the electrical properties of thin film insulators are determined not by the intrinsic properties of the insulator but by other properties, such as the nature of the electrode-insulator contact. A suitable (ohmic) contact (Sec. 3*a*) is capable of injecting additional carriers into the insulator, far in excess of the bulk-generated carriers.[1–3] Also the application of a few volts bias is capable of causing inordinately high fields to be generated in a thin film insulator at the cathode-insulator interface. For fields in excess of 10^6 V cm^{-1}, field-emission injection of relatively large currents from the cathode into the conduction band of the insulator is possible [Sec. 5*c*(1)].

There are also several reasons for believing that the observed conductivity in thin film insulators is due often to extrinsically rather than intrinsically bulk-generated carriers. Consider first the intrinsic current density carried by an insulator.

$$I = e\mu N_c F \exp\left(-\frac{E_g}{2kT}\right) \tag{1}$$

where e is the electronic charge, μ is the mobility, F is the field in the insulator, N_c is the effective density of states in the insulator, E_g is the insulator energy gap, k is Boltzmann's constant, and T is the absolute temperature. At room temperature $N_c = 2.5 \times 10^{19}$ cm^{-3}, and assuming $E_g = 3$ eV and $\mu = 100$ cm^2 V^{-1} s^{-1}, then even for $F = 10^6$ V cm^{-1} the current density is only of the order 10^{-18} A cm^{-2}. This is many orders of magnitude smaller than the current densities we will be discussing for materials which have energy gaps greater than 3 eV. A second point is that the observed thermal-activation energy associated with the conductivity of the films is much smaller than would be expected ($\simeq E_g/2$) if the conductivity were intrinsic in nature.

The source of the extrinsic conductivity is thought to be the inherent defect nature of evaporated chemical compound films. Stoichiometric films of compound insulators, with which we are primarily concerned, are notoriously difficult to prepare by evaporation, because of decomposition and preferential evaporation of the lower-vapor-pressure constituent atom. For example, using the compound as starting material, elemental Cd tends to evaporate more rapidly from CdS; as a result CdS films contain donor centers of free cadmium.[5] SiO yields a film containing a mixture of compounds varying from SiO to SiO_2, as well as free Si;[6–9] the free Si per se may act as donor centers in these films[10] or alternatively vacancies existing in the insulator may be the source. A further problem that arises is the contamination of the films

by deposits arising from sublimation of the crucible, and by residual gases. Thus it requires dissociation of only one molecule per million, or that the crucible sublimes at one-millionth the rate of the evaporant, to yield an impurity level of the order 10^{17} cm^{-3} within the film. A good example of the existence of donors in evaporated MoO_3 films is obtained by examination of capacity measurements obtained from these films.[4] At room temperature the capacity is *independent* of the film thickness, but at liquid-nitrogen temperature it corresponds to the geometrical capacitance (see Fig. 1). The room-temperature results are explained by assuming that they correspond to the capacitance of Schottky barriers existing at the two electrode-insulator interfaces, which are independent of the insulator thickness [see Secs. 3*c* and 3*d*(2) and Fig. 3*d*], the rest of the bulk being of high conductivity and hence not contributing to the capacity. In order to account for the Schottky barriers and the high bulk conductivity, it is necessary to postulate that the insulator contains a high donor density. The liquid-nitrogen temperature results follow naturally from this model. At the low temperature very few donors are ionized, that is, there are very few carriers in the insulator conduction band. Thus the bulk has a high resistivity and now, together with the Schottky barriers, that is, the total insulator thickness, contributes to the capacity.

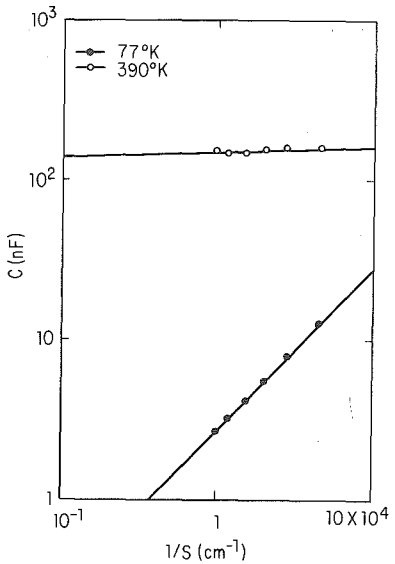

Fig. 1 The capacity c of MoO_3 as a function of inverse sample thickness $1/S$. Dielectric constant obtained from 77°K results corresponds to bulk dielectric constant.

Another important fact to be considered in thin film insulators is traps. Insulating films deposited onto amorphous (e.g., glass) substrates are usually, at best, polycrystalline, and in many cases are amorphous. For crystallite sizes of 100 Å, trapping levels as high as 10^{18} cm^{-3} are possible because of grain-boundary defects alone; in vacuum-deposited CdS, trapping densities as high as 10^{21} cm^{-3} have been reported.[11] Furthermore, vacuum-deposited films contain large stresses which induce further trapping centers.

It follows then that thin film vacuum-deposited insulators can contain a large density of *both* impurity and trapping centers. A judicious study of electrical conduction in vacuum-deposited thin films cannot be accomplished without consideration of these possibilities.

Throughout most of this chapter, since we are concerned primarily with insulators having large energy gaps ($E_g \gtrsim 3$ eV), we will be concerned mainly with the electron rather than the hole carrier, although the results in general can be applied equally well to either type of carrier. The reason for this is twofold. First, the hole mobility is usually much lower than the electron mobility, and thus the hole contribution to the conductivity can usually, but not always, be neglected. Second, in practical insulating films, where, as previously mentioned, the trapping density is high, the tendency is for a free hole to be trapped quickly and thus become immobilized.

b. Band Structure

It will be noted in the energy diagrams that the forbidden gap separating the conduction and valence bands has been shown with well-defined boundaries. Strictly speaking, a well-defined energy gap is a property of a crystalline solid, and in general we are not dealing with such materials; rather we are concerned with polycrystalline,

or even amorphous, insulators. However, it can be shown that the essential features of the band structure of a solid are determined by the *short-range* order within the solid; thus the general properties of the band structure of the crystalline state are carried over to the polycrystalline state. The lack of the long-range order in a noncrystalline solid causes smearing of the conduction and valence-band edges, among other things, so that the energy gap is no longer well defined. What is normally done in the study of thin insulating films is to assume that this smeared-out energy gap, to a first-order approximation, can be represented by a well-defined energy gap which is representative of perhaps an average value of the actual nondiscrete energy gap. With this model we can go a long way toward determining (within the limitation imposed by such a model) features of the band structure of thin film insulators. Indeed, it is found that parameters such as the barrier height at metal-insulator interfaces obtained by independent means (electrical and photoemissive techniques) agree remarkably well, even in what might be considered to be an extreme case of departure from the crystalline state, namely, extremely thin (~30 Å) layers of amorphous Al_2O_3.

3. METAL-INSULATOR CONTACTS

In order to measure the conductivity of an insulator, it is, of course, necessary to connect electrodes to its surfaces in order to facilitate injection of electrons into, and their withdrawal from, the bulk of the insulator. Clearly, the conductivity of the insulator per se will determine the conductivity of the system, since it is much lower than that of the electrodes. In terms of the energy-band picture, the action of the insulator is to erect between the electrodes a potential barrier, extending from the electrode Fermi level to the bottom of the insulator conduction band (see Figs. 2 and 3). This barrier impedes the flow of electrons from one electrode to the other, which would normally flow virtually unimpeded if the insulator were not there (i.e., metal-metal contact). Clearly, then, the height of the potential barrier is an important parameter in conductivity studies in metal-insulator systems. Furthermore, the height of this barrier is determined by the relative alignment of the electrode and insulator energy bands. How then are these bands aligned? This problem is resolved by making use of the rule that in thermal equilibrium the vacuum and Fermi levels must be continuous throughout the system. (The vacuum level represents the energy of an electron at rest just outside the surface of the material, and the energy difference between the vacuum and Fermi levels is called the work function of the material.) It would appear that the equilibrium conditions can be satisfied only when the work function of the metal ψ_m and that of the insulator ψ_i are equal (see Fig. 2*c* and *d*). We shall see a little later, however, that the equilibrium condition is satisfied when $\psi_m \neq \psi_i$, because of charge transfer from the electrode to the insulator or vice versa.

The shape of the potential barrier just within the surface of the insulator depends on whether or not the insulator is intrinsic or extrinsic, and on the relative magnitudes of the work functions of the metal and insulator, among other things.

At reasonable applied fields there will normally be a sufficient supply of carriers available to enter the insulator from the cathode (negatively biased electrode) to replenish the carriers drawn out of the bulk of the insulator. Under these conditions the current-voltage (I-V) characteristics of the sample will be determined by the bulk properties of the insulator; we thus refer to this conduction process as being *bulk-limited*. At high fields, or if the contact is blocking, the current capable of being supplied by the cathode to the insulator will be less than that capable of being carried in the bulk of the insulator. Under these conditions the I-V characteristics of the sample will be controlled primarily by conditions existing at the cathode-insulator interface; this conduction process is referred to as being *emission-limited* or *contact-limited*.

The types of contact that can exist at a metal-insulator interface fall into three categories: (1) ohmic contact, (2) neutral contact, and (3) blocking contact. Each of the contacts will now be discussed in some detail.

a. Ohmic Contact—Mott-Gurney Contact

In order to achieve an ohmic contact at a metal-insulator interface,[1] it is necessary that the electrode work function ψ_m be smaller than the insulator work function ψ_i, as shown in Fig. 2*a*. Here, the term ohmic contact is used to mean that the electrode can readily supply electrons to the insulator as needed. Under these conditions, in order to satisfy thermal-equilibrium requirements, electrons are injected from the

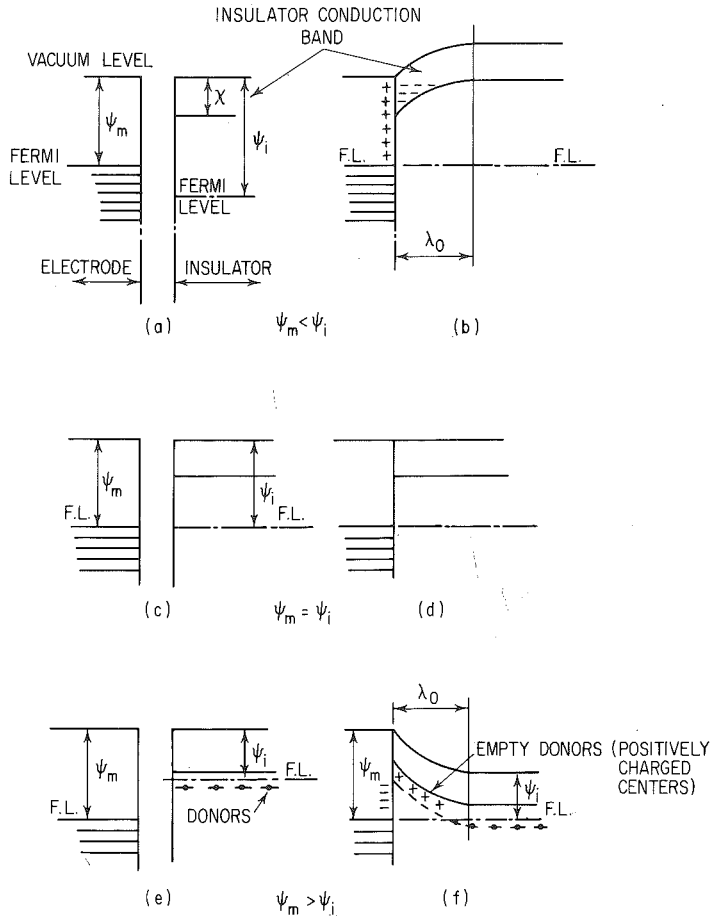

Fig. 2 Energy diagrams showing the requirements and type of contact of: (*a*, *b*) ohmic contact; (*c*, *d*) neutral contact; (*e*, *f*) blocking contact.

electrode into the conduction band of the insulator, thus giving rise to a space-charge region in the insulator. This space-charge region is shown in Fig. 2*b* to extend a distance λ_0 into the insulator, and is termed the accumulation region. In order to satisfy charge-neutrality requirements an equal amount of positive charge, say, Q_0, accumulates on the electrode surface. The electrostatic interaction between the positive and negative charge induces a local electric field within the surface of the insulator, the strength of which falls off with distance from the interface and is zero at the edge of the space-charge region (i.e., at λ_0 from the interface). This field causes the bottom of the conduction band to rise with distance of penetration into the insulator until it reaches the equilibrium value $\psi_i - \chi$, where χ is the insulator affinity.

The field F within the accumulation region is related to the space-charge density $\rho(x)$ within the accumulation region by Poisson's equation:

$$\frac{dF}{dx} = \frac{\rho(x)}{K\epsilon_0} \tag{2}$$

where K and ϵ_0 are, respectively, the dielectric constant and permittivity of free space. To obtain information regarding conditions in the space-charge region, we substitute $\rho(x) = eN_c \exp(-\psi/kT)$ in (2), where ψ is the potential energy of the insulator conduction-band edge with respect to the electrode Fermi level:

$$\frac{dF}{dx} = \frac{1}{e}\frac{d^2\psi}{dx^2} = -\frac{eN_c}{K\epsilon_0}\exp\left(-\frac{\psi}{kT}\right) \tag{3}$$

On integration, using $d\psi/dx = 0$ at $\psi = \psi_i - \chi$, (3) yields

$$\frac{d\psi}{dx} = \left(\frac{2e^2kTN_c}{K\epsilon_0}\right)^{\frac{1}{2}}\left[\exp\left(-\frac{\psi}{kT}\right) - \exp\left(-\frac{\psi_i - \chi}{kT}\right)\right]^{\frac{1}{2}} \tag{4}$$

Integrating (4) using the boundary conditions $\psi = \psi_m - \chi$ and $(\psi_i - \chi)$ at $x = 0$ and λ_0, respectively, we obtain

$$\lambda_0 = \left(\frac{2kTK\epsilon_0}{e^2N_c}\right)^{\frac{1}{2}}\left\{\frac{\pi}{2} - \sin^{-1}\left[\exp\left(-\frac{\psi_i - \psi_m}{2kT}\right)\right]\right\}\exp\left(\frac{\psi_i - \chi}{2kT}\right) \tag{5}$$

For $\psi_i - \psi_m > 4kT$, which will normally be the case, (5) reduces to

$$\lambda_0 \simeq \frac{\pi}{2}\left(\frac{2kTK\epsilon_0}{e^2N_c}\right)^{\frac{1}{2}}\exp\left(\frac{\psi_i - \chi}{2kT}\right) \tag{6}$$

In Table 1 we have calculated λ_0 for various values of $\psi_i - \chi$, from which it is seen that, at room temperature, a good ohmic contact in the thin insulators we are considering requires that the maximum height of the bottom of the conduction band above the Fermi level be less than about 0.3 eV. This statement is tantamount to the more important condition that, for an effective ohmic contact at room temperature, the interfacial barrier $\psi_m - \chi$ must be less than 0.3 eV.

TABLE 1 Depth of Accumulation Region for Several Values of $\psi_m - \chi$

$\psi_m - \chi$, eV	0.1	0.2	0.3	0.4
λ, cm	1.6×10^{-6}	1.2×10^{-5}	7.2×10^{-5}	7.2×10^{-4}

The total amount of charge injected into the insulator to form the ohmic contact is determined as follows: From (2), using the boundary conditions that $F = F_0$ and 0 at $x = 0$ and λ_0, respectively, we have

$$\int_{F_0}^{0} dF = -\frac{1}{K\epsilon_0}\int_0^{\lambda}\rho(x)\,dx$$

or

$$F_0 = \frac{Q_0}{K\epsilon_0} \tag{6a}$$

where $Q_0 = \int_0^{\lambda}\rho(x)\,dx$ is the total space charge per unit area in the accumulation region. Since $\psi = \psi_m - \chi$ at $x = 0$, we have from (4), assuming $\psi_m < \psi_i - 2kT$, which will normally be the case,

$$F_0 = \left(\frac{2kTN_c}{K\epsilon_0}\right)^{\frac{1}{2}}\exp\left(-\frac{\psi_m - \chi}{kT}\right) \tag{6b}$$

Substituting (6*b*) into (6*a*) yields

$$Q_0 = (2K\epsilon_0 kTN_c)^{\frac{1}{2}} \exp\left(-\frac{\psi_m - \chi}{kT}\right) \tag{6c}$$

b. Neutral Contact

When there is no reservoir of charge at the contact ($Q_0 = 0$), this type of contact is known as the neutral contact. The condition $Q_0 = 0$ implies that $\psi_m = \psi_i$, which means that the conduction band is flat right up to the interface; that is, no band bending is present, as shown in Fig. 2*d*.

For initial voltage bias the cathode is capable of supplying sufficient current to balance that flowing in the insulator, so that the conduction process is ohmic. There is no theoretical limit to the maximum current an insulator per se may carry, provided a high enough voltage supply is available. (There are, in fact, practical limitations such as Joule heating and dielectric breakdown, which determine the maximum voltage that an insulator can withstand before catastrophic breakdown occurs.) There is, however, a limit to the current that the cathode can supply, and this is the saturated thermionic (Richardson) current over the barrier. When this limit is reached, the conduction process is no longer ohmic in nature. The maximum field that may be applied to the insulator before the current supplied by the cathode saturates is obtained by equating the saturated thermionic current $n_0 ev/4$ to the current flowing in the insulator $n_0 e\mu F$, where n_0 is the density of electrons in the cathode with energies greater than the interfacial potential barrier and v is the thermal velocity of the carriers:

$$\frac{n_0 ev}{4} = n_0 e\mu F$$

or

$$F = \frac{v}{4\mu}$$

In actual fact we will see later that, because of attendant image forces interacting with the electric field at the cathode-insulator interface, the saturation current increases with applied field [Richardson-Schottky effect; see Sec. 5*a*(2)].

c. Blocking Contact—Schottky Barrier

A blocking contact occurs when $\psi_m > \psi_i$, and in this case electrons flow from the insulator into the metal to establish thermal-equilibrium conditions. A space-charge region of positive charge, the *depletion* region, is thus created in the insulator and an equal negative charge resides on the metal electrode. As a result of the electrostatic interaction between the oppositely charged regions, a local field exists within the surface of the insulator. This causes the bottom of the conduction band to bend downward until the Fermi level within the bulk of the insulator lies ψ_i below the vacuum level. An intrinsic insulator, however, contains such a low density of electrons that it would have to be inordinately thick to provide the required positive space-charge region to satisfy the above condition. Thus the conduction band of an intrinsic insulator at a blocking contact does in fact slope only imperceptibly downward; that is, we have what is essentially a neutral contact.

Let us assume now that the insulator contains a large density N_d of donors, as shown in Fig. 2*f*. The method of determining quantitative information about this depletion region parallels that applied to the doped semiconductor-metal contact, and was first studied by Schottky.[12,13] We assume that the donors are fully ionized and uniformly distributed (N_d cm^{-3}) in the region extending from the interface to a depth λ_0, the depletion region, into the insulator. Poisson's equation (2) for this system is therefore given by

$$\frac{d^2\psi}{dx^2} = \frac{e^2 N_d}{K\epsilon_0}$$

For convenience we measure ψ from the Fermi level of the electrode to the bottom of the conduction band of the insulator. Integrating this expression, and using the boundary condition $d\psi/dx = 0$ at $x = \lambda_0$, we have

$$\frac{d\psi}{dx} = \frac{e^2 N_d}{K\epsilon_0}(x - \lambda_0) \tag{7}$$

Integrating again and using the boundary condition $\psi = \psi_m - \chi$ at $x = 0$, we have

$$\psi = \psi_m - \chi + \frac{e^2 N_d}{K\epsilon_0}\left(\frac{x^2}{2} - \lambda_0 x\right) \tag{8}$$

To determine the depth of the depletion region λ_0, we note that $\psi = \psi_i - \chi$ at $x = \lambda_0$, so that

$$\lambda_0 = \left[\frac{2(\psi_m - \psi_i)K\epsilon_0}{e^2 N_d}\right]^{\frac{1}{2}} \tag{9}$$

or in conventional units

$$\lambda_0 = 1.05 \times 10^3 \left[\frac{(\psi_m - \psi_i)K}{N_d}\right]^{\frac{1}{2}}$$

In Table 2 we calculate a few values for λ_0 using $\psi_m - \psi_s = 2$ eV and $K = 5$. It will be apparent that the donor density is required to be of the order 10^{16} cm^{-3} or higher if the insulator is to be able to accommodate the depletion region adequately.

TABLE 2 Depth of Depletion Region for Several Values of N_d

N_d, cm^{-3}	10^{15}	10^{17}	10^{19}	10^{21}
λ, cm	10^{-4}	10^{-5}	10^{-6}	10^{-7}

To determine the effect of voltage bias on the depletion region, we note that the depletion region has a much lower electron density than the bulk of the insulator; hence its conductivity is much lower. Thus any voltage applied to the system can, to a good approximation, be assumed to be absorbed entirely across the depletion region. Under these conditions, when the metal is negatively biased, the boundary condition leading to (9) is modified to read $\psi = \psi_i - \chi - eV$ at $x = \lambda$. The depth of the depletion region λ now becomes

$$\lambda = \left[\frac{2(\psi_m - \psi_i + eV)K\epsilon_0}{N_d e^2}\right]^{\frac{1}{2}} \tag{10}$$

which increases with increasing voltage bias. The field at the interface $F_0 = e^{-1}\, d\psi/dx$, which we will be needing later, is from (7) and (10)

$$F_0 = -\left[\frac{2N_d(\psi_m - \psi_i + eV)}{K\epsilon_0}\right]^{\frac{1}{2}} \tag{11}$$

d. Metal-insulator-metal Systems

(1) Two Ohmic Contacts Figure 3*a* and *b* illustrates two ohmic contacts on an insulator when the electrodes are at the same potential. In the case of Fig. 3*a* it is seen that the accumulation regions extend right into the bulk of the insulator. As a result, the bottom of the conduction band is curved throughout its length (that is, an electric field exists at all points within the insulator), the highest point of which is less than the equilibrium position $\psi_i - \chi$ above the Fermi level. This condition is due to either the insulator's being too thin or the interfacial potential barriers' being too high, with the result that insufficient charge is contained within

the insulator to screen the interior effectively from the surface. Figure 3*b*, on the other hand, illustrates the case of good ohmic contacts (small interfacial potential barriers) on an insulator. In this case the accumulation regions effectively screen the interior of the insulator from conditions at the surfaces. The bottom of the conduction band is thus flat within the interior of the insulator (field-free interior) and the height of the conduction band within the interior, reaching its equilibrium level $\psi_i - \chi$ above the Fermi level.

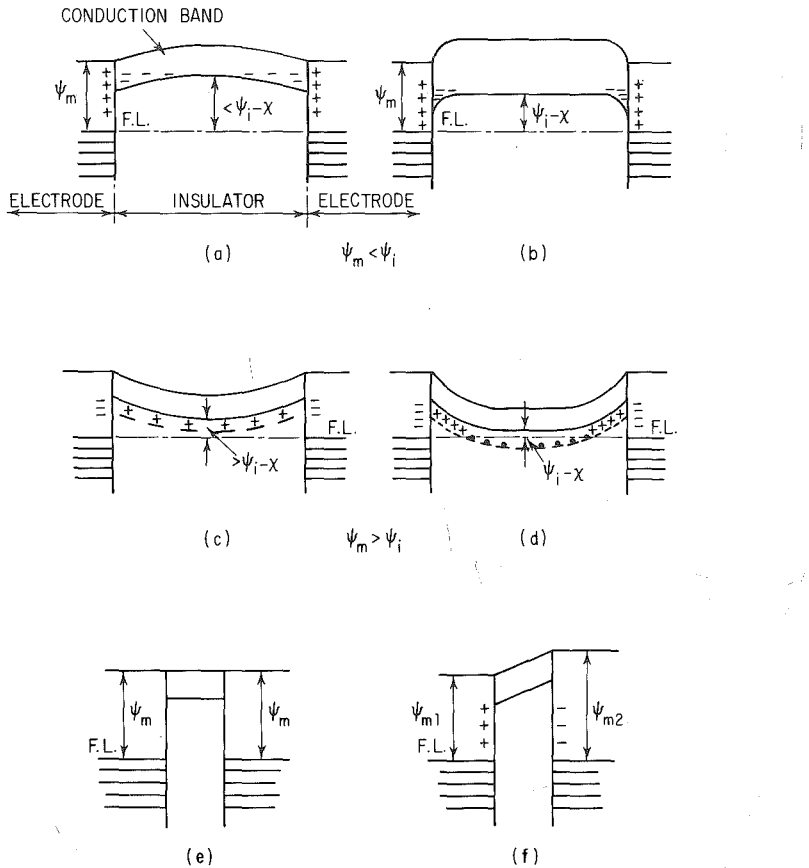

Fig. 3 Energy diagrams of two metal contacts on an insulator. (*a*) and (*b*) represent imperfect and good ohmic contacts, respectively; (*c*) and (*d*) imperfect and good blocking contacts; (*e*) and (*f*) similar and dissimilar neutral contacts.

(2) Two Blocking Contacts Figure 3*c* and *d* illustrates two blocking contacts on an insulator. In Fig. 3*c* the depletion regions extend right into the interior of the insulator, as a result of the insulator's being too thin, or of the doping concentration's being too low. As in the case of Fig. 3*a*, the bottom of the conduction band is curved throughout its length, but in this case concave upward, indicating that an electric field exists at all points throughout the length of the insulator. Because the depletion regions do not effectively screen the interior from the surfaces, the lowest point of the bottom of the conduction band is greater than $\psi_i - \chi$ above the Fermi level. In contrast to Fig. 3*c* the depletion regions in Fig. 3*d* effectively screen the interior

from the surfaces, which means that the insulator is thick or the doping density is high. Under these conditions, as in the case of good ohmic contacts, the interior of the insulator is field-free, and the bottom of the conduction band attains its equilibrium position $\psi_i - \chi$ above the Fermi level, within the interior.

(3) Other Contacts Figure 3*e* and *f* illustrates the case of blocking electrodes on an intrinsic or a *very* thin doped insulator. In this case the insulator is incapable of transferring sufficient charge from its interior to the electrodes to give rise to any effective band bending. In the case of similar electrodes (Fig. 3*e*), the result is that the bottom of the insulator conduction band is flat throughout its entire length.

When the insulator has dissimilar electrodes connected to its surfaces, it is clear that the interfacial potential barriers differ in energy by an amount

$$(\psi_{m2} - \chi) - (\psi_{m1} - \chi) = \psi_{m2} - \psi_{m1}$$

as shown in Fig. 3*f*. This means that the conduction band must slope upward from the lower barrier with a gradient $(\psi_{m2} - \psi_{m1})/s$, where s is the insulator thickness; this in turn means that a uniform intrinsic field F_{in} of strength $(\psi_{m2} - \psi_{m1})/es$ exists within the insulator. The origin of this zero-bias intrinsic field is a consequence of charge transfer between the *electrodes*. The electrode of lower work function, electrode 1, transfers electrons to electrode 2, so that a positive surface charge appears on electrode 1 and a negative surface charge on electrode 2. The amount of charge Q transferred between the electrodes (the surface charge on the electrodes) depends on the contact potential difference $(\psi_{m2} - \psi_{m1})/e$ existing between the electrodes, and the capacity C of the system as follows:

$$Q = \frac{C(\psi_{m2} - \psi_{m1})}{e} = \frac{(\psi_{m2} - \psi_{m1})AK\epsilon_0}{es} \tag{12}$$

where A is the electrode area.

If the insulator is very thin, the intrinsic field within the insulator can be very large; for example, suppose $s = 20$ Å, as one finds in metal-insulator-metal tunnel junction (see Sec. 4), and $(\psi_{m2} - \psi_{m1})/e = 1$ V, $F_{in} = 5 \times 10^6$ V cm^{-1}. When electrode 1 is positively biased, the intrinsic field augments the applied field; that is, the field in the insulator F is given by $F = F_{in} + V/s$. When it is negatively biased, the intrinsic field acts to reduce the effect of the applied field; that is, $F = V/s - F_{in}$. In the latter case the initial effect of the applied voltage $V < (\psi_{m2} - \psi_{m1})/e$ is to *reduce* the field in the insulator, which is zero when $V = (\psi_{m2} - \psi_{m1})/e$. Further increasing voltage application, $V > (\psi_{m2} - \psi_{m1})/e$, then results in a *reversal* of the field in the insulator which increases with increasing voltage bias. It will be apparent, then, when the electrode of lower work function is negatively biased, the insulator will withstand greater voltage application before it breaks down. By investigating the breakdown voltage of very thin metal-insulator-metal junctions as a function of voltage bias, the effect of the intrinsic field can be detected.[14]

e. Effect of Surface States

(1) Free Surface of an Insulator In the above discussions we have taken the interfacial barrier height to be $\psi_m - \chi$. In actual fact experiment shows, in many instances, that this is an oversimplification of conditions existing at real surfaces; in particular the potential barrier is often found to be independent of the work function of the electrode material. In order to account for similar observations on metal-semiconductor contacts, Bardeen[15] suggested surface states as the cause. These states owe their existence in part to the sudden departure from periodicity of the potential at the surface, and in part to the defect chemical nature of the surface, i.e., absorbed gas, etc.

The effect of surface states on the surface potential of the insulator can be seen by reference to Fig. 4. The surface states, of density N_s (per unit area per unit energy), are assumed to be uniformly distributed throughout the energy gap of the insulator, as shown in Fig. 4*a*. When filled up to an energy E_0 below the bottom of the conduction band, the surface is assumed to be electrically neutral. However, since the

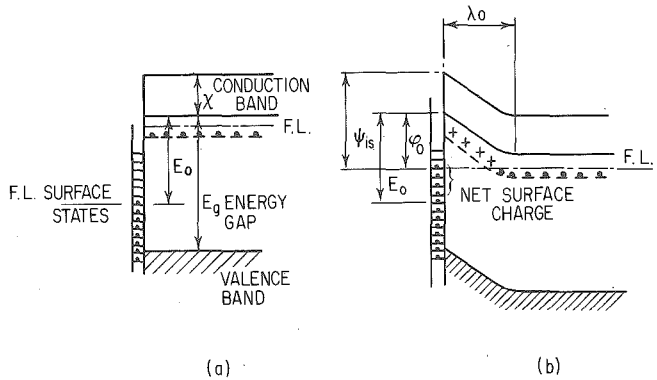

Fig. 4 Energy diagram of surface states on a semiconductor. (*a*) Before surface and bulk are in equilibrium. (*b*) In equilibrium.

Fermi level associated with the surface states should coincide with that of the bulk, electrons from the conduction band fill up additional surface states. The result is that a sheet of negative charge resides on the surface and a positive space charge (depletion region) of depth λ_0 exists just inside the surface of the insulator, as shown in Fig. 4*b*. This double-charge layer causes the conduction band to bend upward until the highest filled surface state coincides with the Fermi level of the bulk (Fig. 4*b*). This band bending is possible only in a *doped* insulator; in an intrinsic insulator too few electrons are available in the bulk to effect this process.

Electrical neutrality of the crystal requires that the number of ionized donors in the depletion region be equal to the excess number of electrons in surface states; thus (from Fig. 4*b*)

$$N_s(E_0 - \phi_0) = N_d\lambda_0$$

or

$$E_0 - \phi_0 = \frac{N_d\lambda_0}{N_s} \tag{13}$$

From (13), for an infinite density of surface states we have $\phi_0 = E_0$; *thus the potential barrier at the surface ϕ_0 is determined entirely by surface states.* To see the effect of practical values of surface-charge densities ($N_s \simeq 10^{14}$ cm^{-2} eV^{-1}) let us assume that the bands at the surface are bent upward by an amount 1 eV; therefore, from Table 2 we see that for $N_d = 10^{17}$ cm^{-3}, $\lambda_0 = 10^{-5}$ cm. Using these values and $N_s = 10^{14}$ cm^{-2} eV^{-1} in (13), we have $E_0 - \phi_0 = 10^{-2}$ eV or $\phi_0 \simeq E_0$. Thus for practical values of surface states, ϕ_0 ($\simeq E_0$) is essentially controlled by conditions existing at the surface rather than in the bulk.

(2) Surface States at a Metal-insulator Contact Let us now consider what happens when an electrode is attached to the surface of an insulator on which surface states are prominent. If the work function of the electrode ψ_m is the same as the work function of the insulator surface ψ_{is} the potential barrier ϕ_0 remains undisturbed when contact is made, because the two surfaces are in equilibrium before and after contact.

If $\psi_m > \psi_{is}$, electrons are transferred out of the surface states into the surface of the electrode and provided the density of the surface states is sufficiently high, the difference in potential $\psi_m - \psi_{is}$ of the two surfaces can be virtually all dropped in the double-charged surface layers without appreciably altering the potential barrier at the insulator surface. For example, the number of electrons per unit area transferred between the surfaces is $Q/e = (\psi_m - \psi_{is})K\epsilon_0/e^2a$ [cf. Eq. (12)], where a is the distance between the two surfaces, that is, a distance of the order of the interatomic spacing. Assuming $\psi_m - \psi_{is} = 1$ eV and $a = 5 \times 10^{10}$ m, we have $Q/e \simeq 10^{17}$ m^{-2} or 10^{13} cm^{-2}. Assuming $N_s = 10^{14}$ cm^{-2} eV^{-1}, only $\frac{1}{10}$ eV spread of energies at the top of the filled surface states are depleted in order to supply the necessary charge to the electrode to accommodate the difference in work function (see Fig. 5). This

means that the Fermi energy of the surface of the insulator is lowered by only $\frac{1}{10}$ eV in releasing the required charge, or in other words, the potential barrier at the surface is increased by only $\frac{1}{10}$ eV as compared with 1 eV if the surface states had not been present. *Thus the interfacial potential barrier is virtually independent of the electrode work function.* Using similar arguments, it can be shown that if ψ_m is 1 eV *less* than ψ_{is}, the surface potential of the insulator *decreases* by only $\frac{1}{10}$ eV.

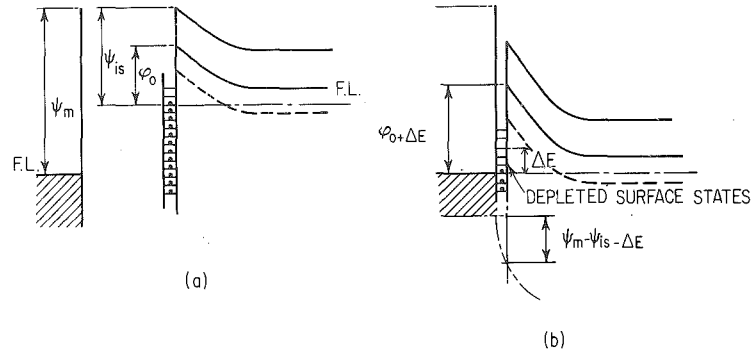

Fig. 5 Energy diagram of a metal surface and insulator surface which contains surface states. (*a*) Before contact. (*b*) After contact (the chain-dotted line below the energy diagram shows the change in potential in passing through the interface).

The above arguments have been simplified somewhat in that we have not taken into account changes that occur in the depletion region of the insulator when changes occur at its surface, but this is a matter of detail rather than principle. The important point is that these two examples serve to show why surface states tend to make the surface potential barrier almost independent of the work function of the electrode, and why it is no longer appropriate invariably to assign the value $\psi_m - \chi$ to the height of the interfacial barrier. Henceforth we will define the height of the potential barrier by the parameter ϕ_0, its value being dependent on whether or not surface states play a significant part in determining the barrier height.

4. TUNNEL EFFECT—VERY THIN INSULATORS

a. Tunnel Effect

If the energy of an electron is less than the interfacial potential barrier in a metal-insulator-metal junction upon which it is incident, classical physics predicts certain reflection of the electron at the interface; that is, the electron cannot penetrate the barrier and hence its passage from one electrode to the other is precluded. Quantum theory, however, contradicts this thesis. The quantum-mechanical wave function $\psi(x)$ of the electron has finite values within the barrier (see Fig. 6), and since $\psi(x)\psi^*(x)\cdot dx$ is the probability of finding the electron within the incremental range x to $x + dx$, this means that the electron can penetrate the forbidden region of the barrier. The wave function decays rapidly with depth of penetration of the barrier from the electrode-insulator interface and, for barriers of macroscopic thickness, is essentially zero (Fig. 6*a*) at the opposite interface, indicating zero probability of finding the electron there. However, if the barrier is very thin (<50 Å), the wave function has a nonzero value at the opposite interface. For this case, then, there is a finite probability that the electron can pass from one electrode to the other by penetrating the barrier, as shown in Fig. 6*b*. When the electron passes from one electrode to the other by this process, one speaks of the electron as having tunneled through the barrier.

If two metal surfaces are placed within 50 Å of each other, it follows from the above discussion that the electrons will pass from one electrode to the other by means of the

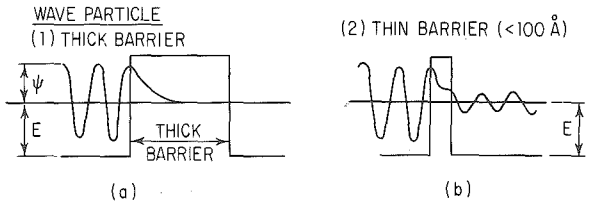

Fig. 6 Schematic representation of the tunnel effect. (*a*) At a thick barrier. (*b*) At a very thin barrier.

tunnel effect. The close separation of the electrode is achieved in practice by means of a thin insulating film, usually the thermally or anodically grown surface oxide of one of the electrodes, and it is this system we will now consider.

b. Theory of Isothermal Tunneling

The tunnel effect between metal electrodes was first studied, in elementary fashion, by Frenkel.[16] Sommerfeld and Bethe[17] made the first comprehensive studies, in which they included image-force effects but confined their calculations to very low ($V \ll \phi_0/e$) and very high ($V \gg \phi_0/e$) voltages. Holm[18] made the next notable investigations and extended the calculation to intermediate voltages, although approximations he used have been found questionable. Stratton[19] and Simmons[20–22] further extended the theory, and the results of these studies are those currently most commonly used in the analysis of experimental data. All the above studies were concerned with a barrier which in its simplest forms is shown in Fig. 3*e* and *f*. These models appear to be quite suitable for predicting the salient features of the tunneling *I-V* characteristics. There have been several other studies of a more detailed nature. These have considered the effect of space charge,[23–25] traps[26] and ions[27] in the insulator, the effect of the shape of forbidden band,[28,29] representation of the insulator by a series of potential wells,[30,31] electric-field penetration of the electrodes,[32] and diffuse reflection.[33] We will be concerned here with tunneling through a barrier of the type shown in Fig. 3*e* and *f*.

We will discuss the theory of the tunnel effect using the notation and type of approximation developed by the author.[20,21] This formulation is readily applicable to potential barriers of arbitrary shape and to all practical voltage ranges.

The generalized formula gives the relationship connecting the tunnel current density with the applied voltage for a barrier of arbitrary shape (see Fig. 7) as

$$I = I_0\{\bar{\phi}\exp(-A\bar{\phi}^{\frac{1}{2}}) - (\bar{\phi}+eV)\exp[-A(\bar{\phi}+eV)^{\frac{1}{2}}]\} \qquad (14)$$

where $$I_0 = \frac{e}{2\pi h(\beta\,\Delta s)^2} \quad \text{and} \quad A = \frac{4\pi\beta\,\Delta s}{h}(2m)^{\frac{1}{2}}$$

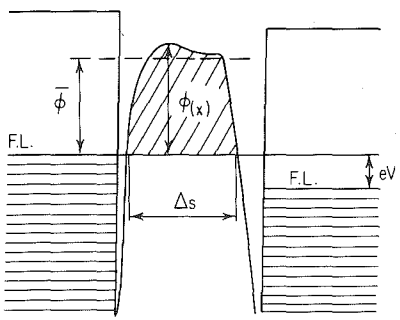

Fig. 7 Energy-diagram arbitrary barrier [see (14)] illustrating the parameters $\bar{\phi}$ and Δs.

Δs = width of the barrier at the Fermi level of the negatively biased electrode
$\bar{\phi}$ = mean barrier height above the Fermi level of the negatively biased electrode
h = Planck's constant
m = mass of the electrons
e = unit of electronic charge
β = a function of barrier shape and is usually approximately equal to unity, a condition we will assume throughout

Expressed in conventional units, except for Δs, which is expressed in angstroms, (14) becomes

$$I = \frac{6.2\times10^{10}}{(\Delta s)^2}\{\bar{\phi}\exp(-1.025\Delta s\,\bar{\phi}^{\frac{1}{2}}) - (\bar{\phi}+V)\exp[-1.025\Delta s(\bar{\phi}+V)^{\frac{1}{2}}]\} \qquad (15)$$

This equation was derived using the *WKBJ* approximation and assuming an electronic-energy distribution corresponding to absolute zero of temperature. Since the current density I is only very slightly temperature-dependent, this equation is also suitable for use at higher temperatures. We will, however, discuss the detailed temperature dependence of I later.

We will now use (15) to derive the *I-V* characteristics of an asymmetric junction (see Fig. 3*f*), as this is the type of junction most often met with in practice. Since the *I-V* characteristic will be shown to be polarity-dependent, in the subsequent discussion the following convention will be adopted: any electrical characteristic is described as the *forward* characteristic when the electrode of lower work function is positively biased, and as the *reverse* characteristic when the electrode of lower work function is negatively biased. J_1 will denote the reverse current density and J_2 the forward current density.

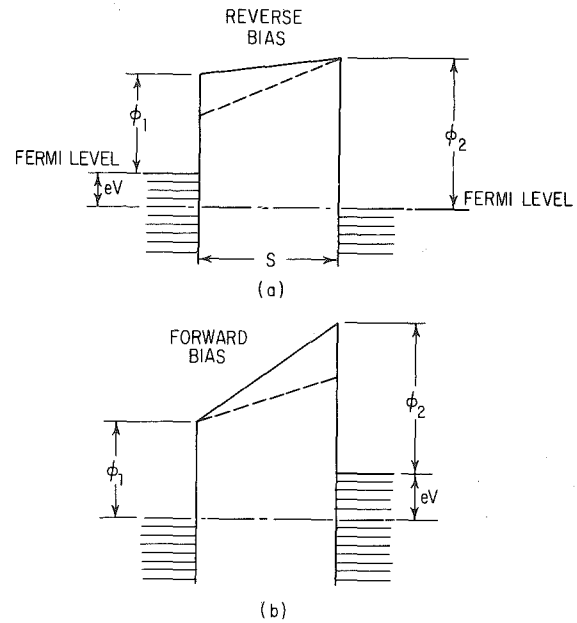

Fig. 8 Energy diagrams of asymmetric potential barrier for $0 \le V \le \phi_1/e$. (*a*) is reverse-bias condition, and (*b*) is forward-bias condition. The dotted lines illustrate the shape of the barrier (with respect to the Fermi level of the *positively* biased electrode in both cases) under zero-bias condition.

(1) Voltage Range $0 \le V \le \phi/e$ Figure 8*a* and *b* illustrates the energy diagram for the reverse and forward directions of polarity, respectively. It follows for both these cases that

$$\bar{\phi} = \frac{\phi_1+\phi_2-eV}{2} \qquad \Delta s = s \qquad (16)$$

which on substitution in (15) yields

$$J_1 = J_2 = \frac{3.1\times10^{10}}{s^2}\{(\phi_1+\phi_2-V)\exp[-0.725s(\phi_1+\phi_2-V)^{\frac{1}{2}}] - (\phi_1+\phi_2+V)\exp[-0.725s(\phi_1+\phi_2+V)^{\frac{1}{2}}]\} \qquad (17)$$

Since $J_1 = J_2$, the I-V characteristic is symmetric with polarity of bias for the voltage range $0 \leq V \leq \phi_1/e$.

(2) Voltage Range $V > \phi_2/e$ From Fig. 9*a* we have for the reverse-biased condition, using $\Delta\phi = \phi_2 - \phi_1$,

$$\bar{\phi} = \frac{\phi_1}{2} \qquad \Delta s = \frac{s\phi_1}{eV - \Delta\phi} \tag{18}$$

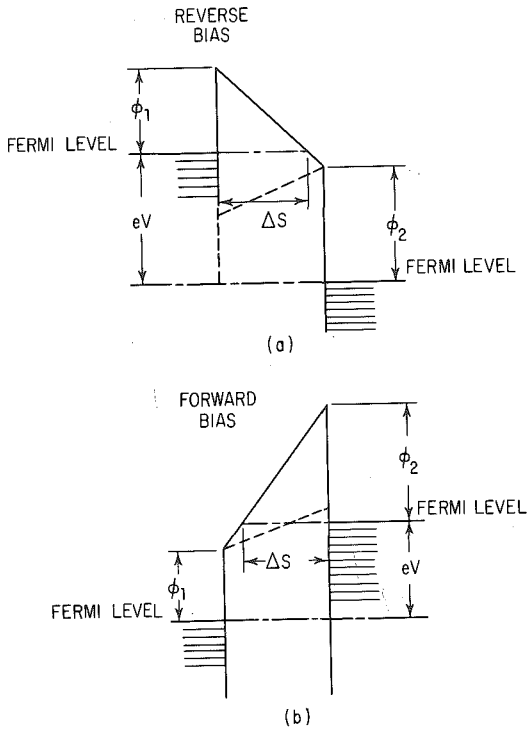

Fig. 9 Energy diagrams of potential barriers for $V > \phi_2/e$. (See Fig. 8 for explanation of dotted lines.)

which on substitution in (15) yields

$$J_1 = \frac{3.38 \times 10^{10}(V - \Delta\phi)^2}{\phi_1 s^2}\left\{\exp\left(-\frac{0.69s\phi_1^{\frac{3}{2}}}{V - \Delta\phi}\right) - \left(1 + \frac{2V}{\phi_1}\right)\exp\left[-\frac{0.69s\phi_1^{\frac{3}{2}}(1 + 2V/\phi_1)^{\frac{1}{2}}}{V - \Delta\phi}\right]\right\} \tag{19}$$

From Fig. 9*b* we have for the forward-biased condition

$$\bar{\phi} = \frac{\phi_2}{2} \qquad \Delta s = \frac{s\phi_2}{eV + \Delta\phi} \tag{20}$$

which on substitution in (15) yields

$$J_2 = \frac{3.38 \times 10^{10}(V + \Delta\phi)^2}{\phi_2 s^2}\left\{\exp\left(-\frac{0.69s\phi_2^{\frac{3}{2}}}{V + \Delta\phi}\right) - \left(1 + \frac{2V}{\phi_2}\right)\exp\left[-\frac{0.69s\phi_2^{\frac{3}{2}}(1 + 2V/\phi_2)^{\frac{1}{2}}}{V + \Delta\phi}\right]\right\} \tag{21}$$

In this case, Eqs. (19) and (21) are not equivalent. It follows, then, that the J-V characteristic is asymmetric in this range. In actual fact, not only is the J-V characteristic asymmetric with polarity of bias, but also the direction of easy conductance reverses at some particular voltage, as shown in Fig. 10.

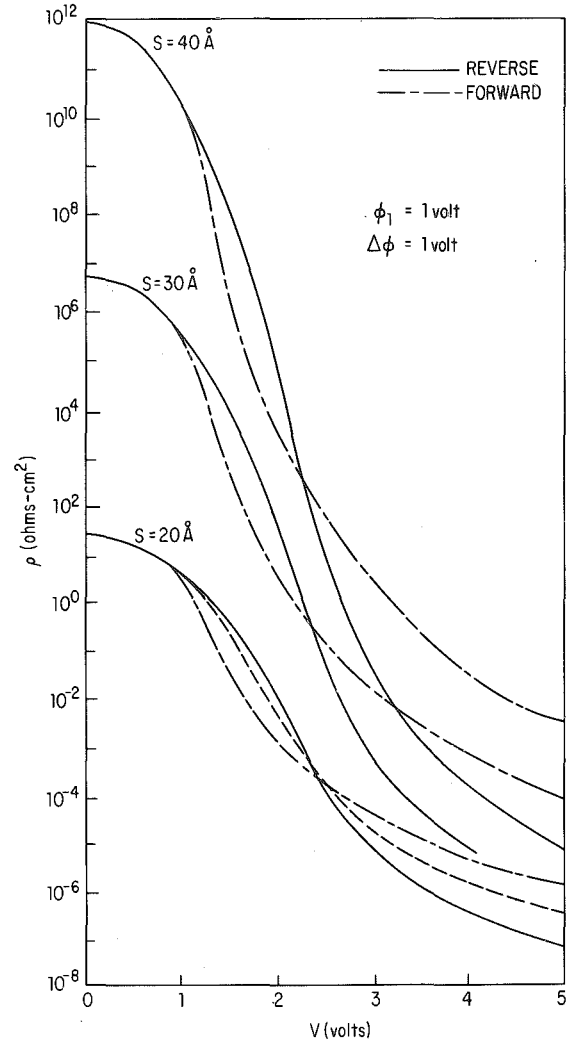

Fig. 10 $\rho(= J/V) - V$ tunnel characteristics for $\phi_1 = 1$ eV, $\phi_2 = 2$ eV, and $s = 30$, 30, and 40 Å.

At high voltages, i.e., $V \gg \phi_2/e$, both (20) and (21) reduce to the familiar Fowler-Nordheim[34] form:

$$J = \frac{3.38 \times 10^{10}F^2}{\phi}\exp\left(-\frac{0.69\phi^{\frac{3}{2}}}{F}\right) \tag{22}$$

where F is the field in the insulator.

(3) Numerical Evaluation Figure 10 illustrates the tunnel resistance $\rho(V/J)$ as a function of V for $s = 20$, 30, and 40 Å, $\phi_1 = 1$ eV and $\phi_2 = 2$ eV ($\Delta\phi = 1$ eV). The plot of ρ vs. V, rather than I vs. V, is chosen for illustration, because it is a more efficacious way of illustrating the effect of the junction parameters on the electrical characteristics. The reverse and forward directions are depicted, respectively, by the full and dotted curves. Several features of the curves will be immediately apparent. At very low voltages the curves are horizontal, which means that the junction resistance is ohmic. The junction resistance falls off rapidly with increasing voltage bias, and the effect is more pronounced the thicker the film. The junction characteristics are symmetric for $V < \phi_1/e$, as noted above; thereafter they are asymmetric, with the direction of asymmetry *reversing* at a voltage ($V \simeq 2.5$ V) which is practically independent of insulator thickness.

The dotted curve accompanying the characteristic marked $s = 20$ Å illustrates the ρ-V characteristic for a *symmetric* junction having $\phi = (\phi_1 + \phi_2)/2 = 1.5$ V. This characteristic is, of course, symmetric over the entire voltage range. As would be intuitively expected, this characteristic is approximately equal to the logarithmic mean value of the forward and reverse characteristics.

The above results apply to an ideal barrier inasmuch as image forces have been neglected (see Sec. 5*a*). The effect of image forces is to reduce the size of the barrier and to lower the impedance of the junction without affecting the general functional form of the curves and the observations noted above. For further information the reader is referred to the literature.[20,21,24]

c. Temperature Dependence of the Tunnel Characteristic

Stratton[19] first derived the I-V characteristic for a tunnel junction for $T \neq 0$°K. The thermal I-V characteristic has also been expressed in terms of the generalized theory, which we will now present.[22]

The generalized formulation connecting the tunnel current $J(V,T)$ at T°K to the tunnel current $J(V,0)$ at 0°K [Eq. (14)] is given by

$$\frac{J(V,T)}{J(V,0)} = \frac{\pi BkT}{\sin \pi BkT}$$

where

$$B = \frac{A}{2\bar{\phi}^{\frac{1}{2}}}$$

In conventional units, but with Δs in angstroms,

$$\frac{J(V,T)}{J(V,0)} \simeq 1 + \frac{3 \times 10^{-9}(\Delta s T)^2}{\bar{\phi}} \tag{23}$$

This result, which states that the thermal component of the tunnel current for a given applied bias V should be proportional to T^2, has been observed experimentally.

A more interesting and important result is apparent, however, when we rewrite (23) in the following form:

$$\hat{J} = \frac{100[I(V,T) - I(V,0)]}{I(V,0)} = \frac{3 \times 10^{-7}(\Delta s T^2)}{\bar{\phi}} \tag{24}$$

$\hat{J}$ will be recognized as the percentage change in current (for a fixed voltage bias) as the temperature is raised from 0 to T°K. From (16) and (18) we have, for reverse bias,

$$\hat{J}_1 = \begin{cases} 6 \times 10^{-7} \dfrac{(sT)^2}{\phi_1 + \phi_2 - V} & V \leq \phi_2 \\ 6 \times 10^{-7}\phi_1 \left(\dfrac{sT}{V - \Delta\phi}\right)^2 & V \geq \phi_2 \end{cases} \tag{25}$$

From (16) and (20), for forward bias,

$$\hat{J}_2 = \begin{cases} 6 \times 10^{-7} \dfrac{(sT)^2}{\phi_1 + \phi_2 - V} & V \leq \phi_1 \\ 6 \times 10^{-7}\phi_2 \left(\dfrac{sT}{V + \Delta\phi}\right)^2 & V \geq \phi_1 \end{cases} \tag{26}$$

These results are exemplified by the curves in Fig. 11. Both $\hat{J}_1$ and $\hat{J}_2$ increase initially until they reach a maximum value, after which they decrease. However, the maximum of $\hat{J}_1$ and the voltage at which it occurs ($V = \phi_2$) exceed the corresponding maximum and voltage ($V = \phi_1$) for $\hat{J}_2$. The importance of (25) and (26) is now apparent: the $\hat{J}$ maxima occur at voltages equal to the electrode-insulator interfacial barrier heights ϕ_1 and ϕ_2 (or ϕ_0, in the case of a symmetric barrier) and, because of the prominence of the peaks, provide a good method of experimentally determining barrier heights.

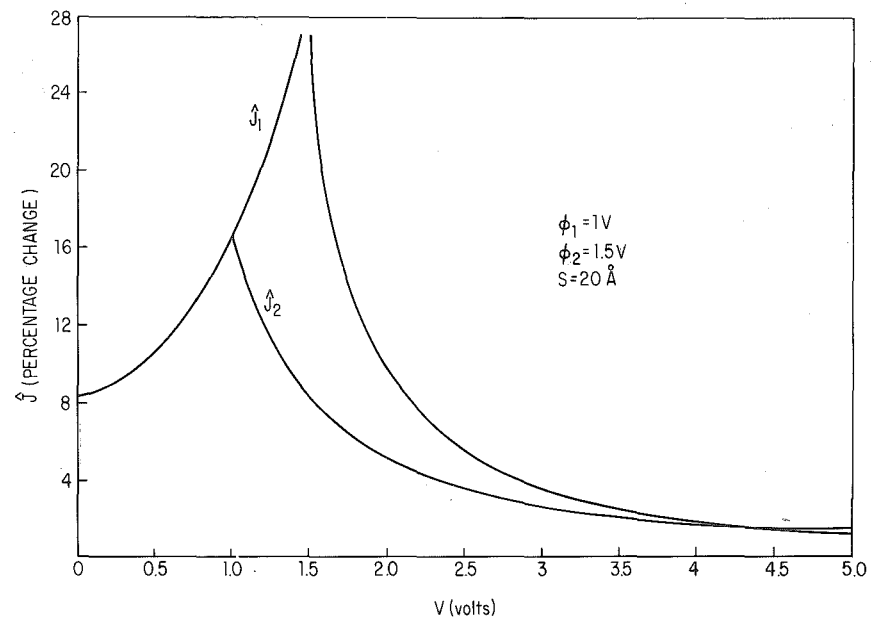

Fig. 11 Graph of $\hat{J}$ vs. V. In this case $\hat{J}$ represents the percentage change in current at a voltage bias V as the temperature of the tunnel junction is raised from 0 to 300°K. Curve *abc* is the forward-biased ($\hat{J}_2$-V) characteristic and *ade* the reverse-bias ($\hat{J}_1$-V) characteristic.

d. Experimental

(1) Thermally Grown Oxides Since Fisher and Giaever[35] made the first successful thin film tunnel junctions, there have been many reported observations of tunneling through thin films.[35–48] Here, however, we will confine our attention to the work of Pollack and Morris,[46,47] which offers good experimental examples of such observations. The results of Pollack and Morris[46] for a Al-Al_2O_3-Al junction are reproduced in Fig. 12. The oxide layer was thermally grown on the surface of the first-deposited Al layer. Their data are compared with the theory, and it is seen that the room-temperature data correlate with theory over *nine* decades of current. Further support for the theory that the conduction phenomena are due to the tunnel effect is seen by the fact that the J_2-V characteristic is only very slightly temperature-

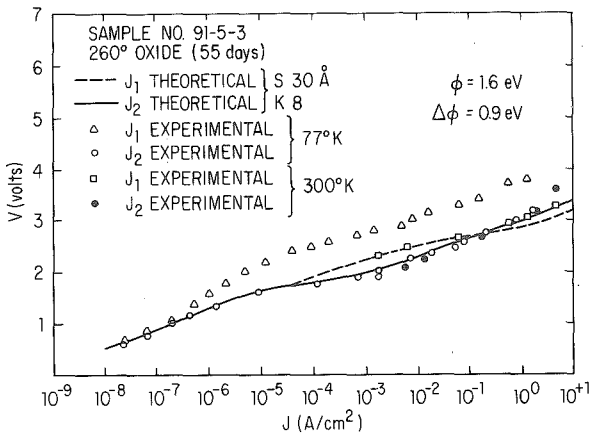

Fig. 12 Tunnel data from Al-Al_2O_3-Al junction.

sensitive. The fact that the barrier heights ($\phi_1 = 1.6$ eV and $\phi_2 = 2.5$ eV) are not the same even though the junction is apparently symmetric, and that the J_1-V curve is temperature-sensitive, is attributed to a semiconducting transition region existing between the oxide layer and its parent electrode.

(2) Gaseous Anodized Layers Pollack and Morris[47] also fabricated Al-Al_2O_3-Al tunnel junctions in which the oxide layer was grown on the Al electrode in an oxygen glow discharge.[49] Again correlation between theory and experiment was observed over many decades of current density, and the barrier heights for the Al-Al_2O_3-Al system were determined to be $\phi_1 = 1.5$ eV and $\phi_2 = 1.85$ eV. These barrier heights were later confirmed remarkably closely by Braunstein et al.,[50] who determined the barriers to be $\phi = 1.49$ eV and $\phi = 1.92$ eV using photoemissive techniques.

The thermal-tunnel ($\hat{J}$-V characteristic, Sec. 4c) results[47] of Pollack and Morris for Al-Al_2O_3-Al junctions are transcribed in Fig. 13. These curves are very similar in structure to the theoretical curves of Fig. 11, but the percentage increase in the current is greater than theoretically expected. The barrier heights determined from these data, that is, the voltage at which the current peaks occur, are $\phi_1 = 2.0$ eV and $\phi_2 = 2.2$ eV, which are somewhat higher than the isothermal results; these

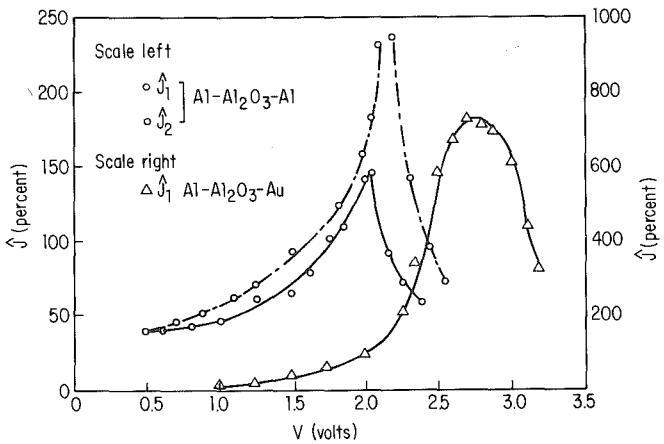

Fig. 13 $\hat{I}$-V tunnel characteristic for Al-Al_2O_3-Al and Al-Al_2O_3-Au junction.

deviations have been attributed to the electric-field penetration of the electrodes.[32,51] The $\hat{J}$-V characteristic for an Al-Al_2O_3-Au junction is also shown in Fig. 13, but the peak in this characteristic is not as sharp as for the Al-Al_2O_3-Al junction. The displacement between the peak of this characteristic and that of the corresponding Al electrode of the Al-Al_2O_3-Al junction is 0.6 V, which is in reasonable agreement with the difference in work junction (0.75 eV) of Al and Au, as required by theory.

e. Other Tunneling Investigations

(1) Al-Al_2O_3-SnTe Junctions Esaki and Stiles[52,53] have fabricated Al-Al_2O_3-SnTe junctions by evaporating SnTe onto an oxidized strip of Al. The SnTe layer was highly P-type with about 8×10^{20} carriers cm^{-3}. The I-V characteristics of these junctions (Al electrode negatively biased) taken at $T = 4.2$, 77, and 300°K, are shown in Fig. 14. The interesting feature of these curves is the negative resistance region in the range 0.55 to 0.85 V.

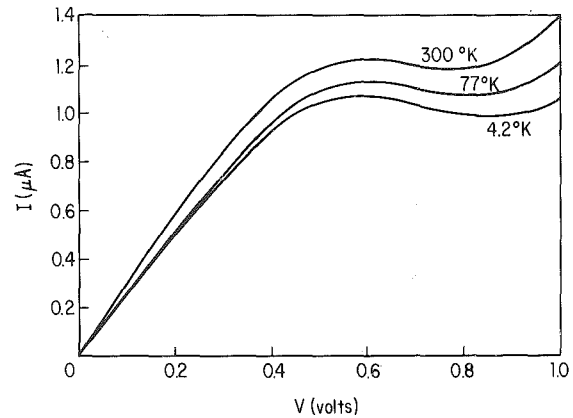

Fig. 14 J-V characteristic for Al-Al_2O_3-SnTe tunnel junction.

The I-V characteristic can be explained along the following lines, using Fig. 15, which is an idealized energy diagram for the system for various voltage biases (the Al electrode is negatively biased in all cases). For voltages $eV < E_F$ (Fig. 15b) the electrons tunnel from the metal electrode into the empty states lying between the electrode Fermi level and the top of the valence band of the SnTe. The tunnel probability [$\simeq \exp(-\Delta s\, \bar{\phi}^{\frac{1}{2}})$, where Δs and $\bar{\phi}$ are defined in Sec. 4b] for electrons at the Fermi level of the metal increases with increasing bias according to

$$\exp\left[-1.025s\left(E_c + E_g + E_F - \frac{eV}{2}\right)^{\frac{1}{2}}\right] \tag{27}$$

which means that the tunnel current into the SnTe valence band I_v increases with increasing bias.

When $E_F < eV < E_F + E_g$ (Fig. 15c), electrons in energy levels between AB in the electrode only may tunnel into the SnTe. Electrons in levels in the electrode between A and the electrode Fermi level face the energy gap of the SnTe and hence cannot tunnel into the SnTe, because there are no available empty states there. The tunnel probability of electrons entering the SnTe just below the top of the valence band is given by

$$\exp\left[-1.025s\left(E_g + E_c + \frac{eV}{2}\right)^{\frac{1}{2}}\right] \tag{28}$$

The tunnel probability now *decreases* with increasing voltage bias, which means

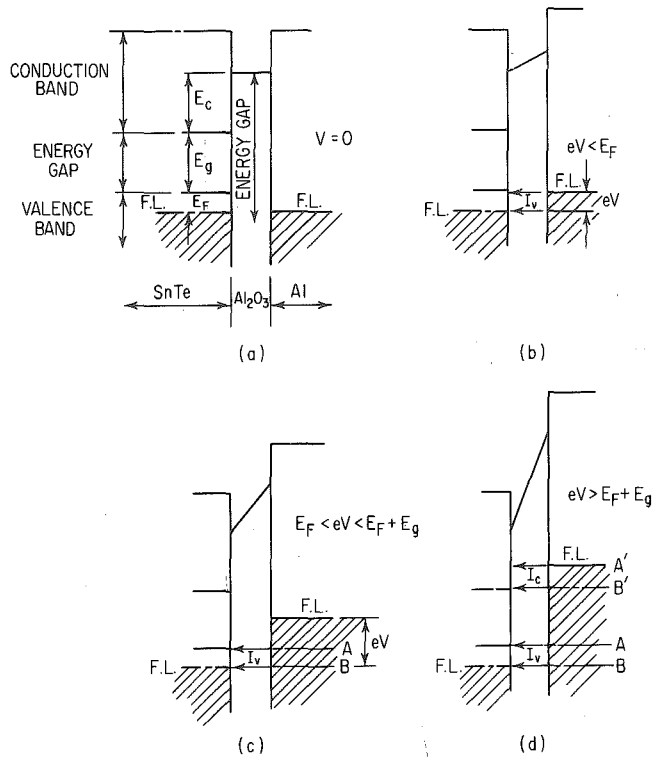

Fig. 15 Energy diagram for Al-Al_2O_3-SnTe tunnel junction for various voltage biases.

I_v *decreases* for $E_F < eV < E_F + E_g$, resulting in the negative-resistance region shown in Fig. 14.

For $eV > E_F + E_g$, the tunnel probability for the electrons tunneling into the SnTe valence band is still given by (28). However, we now have an additional component of current I_c which is due to electrons in the electrode in levels between the Fermi level of the electrode and the bottom of the SnTe conduction band (i.e., between A' and B') tunneling into the *conduction* band of the electrode. The tunnel probability of an electron at the Fermi level is

$$\exp\left[-1.025s\left(E_c + E_g + E_F - \frac{eV}{2}\right)^{\frac{1}{2}}\right] \tag{29}$$

which *increases* with increasing voltage bias; thus I_c *increases* with voltage bias. Since the total tunnel current is given by $I_c + I_v$, and since I_c increases much more rapidly than I_v decreases, the tunnel current increases when $eV \gtrsim E_F + E_g$.

In view of the above considerations and by inspection of Fig. 14, it can be seen that $E_F \simeq 0.06$ eV and the energy gap of the SnTe is $E_g \simeq 0.30$ eV.

(2) Si-SiO_2-(Au + Cr) Dahlke[54] has reported a somewhat different type of tunneling in metal-insulator-semiconductor structures. In this case the semiconductor used was *bulk* degenerate P^{++}-type silicon, <111> orientation, 0.0006 ohm-cm resistivity. The oxide was the thermally grown oxide (SiO_2) of the semiconductor; the metal electrodes were evaporated films of Au + Cr.

Figure 16 illustrates the dc tunnel conductance, obtained by differentiation of the measured I-V characteristics, of three samples in which the oxides were grown under different conditions; the polarity of the voltage axis corresponds to that of the metal

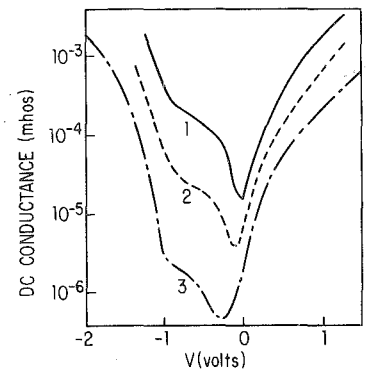

Fig. 16 The dc tunnel conductance vs. voltage of a Si-SiO-(Au + Cr) tunnel junction.

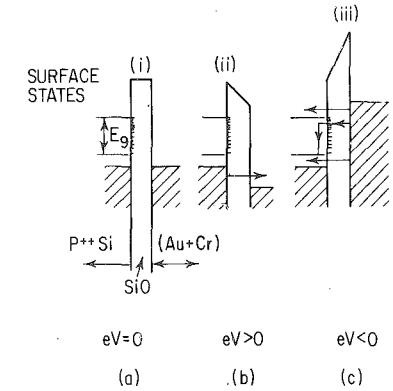

Fig. 17 Energy diagram for Si-SiO(Au + Cr) tunnel junctions for various voltage biases.

electrode. Samples 1 and 2 were grown in dry oxygen and steam, respectively, by brief rf heating under bias at 700°C; sample 3 was treated as sample 2 followed by a 30-min annealing in H_2 at 350°C. The right-hand half of Fig. 16 is the conductance characteristic for electrons tunneling from the valence band into the electrode, as shown in Fig. 17*b*, and increases monotonically with voltage, as is to be expected. The reason that negative resistance, as observed by Esaki and Stiles in essentially the same type of system, is not observed in these samples when the metal electrode is negatively biased ($V < 0$) is apparently the existence of surface states in the forbidden gap of the semiconductor. Thus electrons in levels in the metal electrode positioned opposite the forbidden band of the semiconductor can, unlike in the case of Esaki and Stiles [Sec. 4*e*(1)] where the effect of the surface states was apparently negligible, tunnel into the semiconductor by way of the surface states. They then recombine with holes in the valence band, as shown in Fig. 17*c*. Thus the structure in the characteristics ($V < 0$) is due to the combined effects of the band gap and the distribution of the surface states therein. From these results Dahlke estimates an increase of one to two orders of magnitude in surface-state density in changing from annealed-steam-grown to steam-grown to dry-oxygen-grown oxide layers, which is in qualitative agreement with other workers' results.

(3) Determination of E-K Relationship in Aluminum Nitride Lewicki and Mead[55] have reported a series of measurements carried out on Al-AlN-Mg tunnel junctions of various thicknesses. The junctions were fabricated by exposing freshly deposited Al films in a N_2 glow discharge at a pressure of 200 millitorr for 3 min and subsequently evaporating into the AlN a Mg counterelectrode. From the dependence of the tunneling current upon the insulator thickness they have determined the relationship between the imaginary component of the propagation vector in the forbidden band of the AlN as a function of energy.

The experimental data were analyzed in terms of theory developed by Stratton et al.,[28] which is summarized as follows: For applied voltages higher than the average spread of the tunneling electrons, but less than the barrier energy of the positively biased electrode, the current is related to the average wave vector $\bar{k}(eV)$, corresponding to an incident electron with energy equal to the metal Fermi energy, by

$$I = \frac{B}{s^2}\exp\left[-2\bar{k}(eV)s\right]$$

which is simply the first part of (14) when $\bar{k}(eV) = (2m/h^2)\bar{\phi}^{\frac{1}{2}}$. Thus the ratio of the currents I_1 and I_2 from two samples of thickness s_1 and s_2 for a given voltage

bias is related to $\bar{k}(eV)$ by

$$2\bar{k}(eV) = (s_1 - s_2)^{-1} \ln\left(\frac{I_2 s_2^2}{I_1 s_1^2}\right)$$

Since $\bar{k}$ is an integral of $k(E)$, $k(E)$ is related to $\bar{k}$ as follows:

$$k(\phi_2 - eV) = \frac{1}{e}\frac{\partial}{\partial V}(\phi_1 - \phi_2 + eV)\bar{k}(eV) \tag{30}$$

Lewicki and Mead applied (30) to their results and determined the $k(E)$ vs. E dependence for their junction insulator, which was found to be in agreement with Franz's[55a] empirical relationship for a material with the 4.2-eV energy band gap of AlN, as shown in Fig. 18.

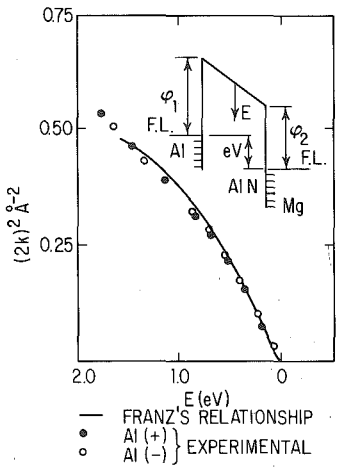

Fig. 18 E-k relationship for AlN. Inset is an energy-band diagram of junction with Al electrode negatively biased.

(4) Magnetointernal Field Emission Esaki et al.[56] have prepared tunnel junctions using thin films of Eu chalcogenides, such as EuSe and EuS. We will confine our remarks to observations made on EuSe, which has a magnetic-transition temperature at 4.7°K. The junctions were prepared by the successive evaporation of Al, EuSe, and Au, onto a heated sapphire substrate without breaking vacuum. The thickness of the EuSe ranged from 200 to 600 Å.

The I-V characteristics on a log (I/V^2) vs. V^{-1} plot at 4.2°K are shown in Fig. 19. It is noted that the curve shifts toward lower voltage in a magnetic field, and also that the I-V characteristic obeys a relationship of the form

$$I = \frac{aV^2}{s^2}\exp\left(-\frac{bs}{V}\right) \tag{31}$$

which is typical of Fowler-Nordheim tunneling [see (22)]. In (31) a and b are constants of the material and s is the EuSe thickness. It was

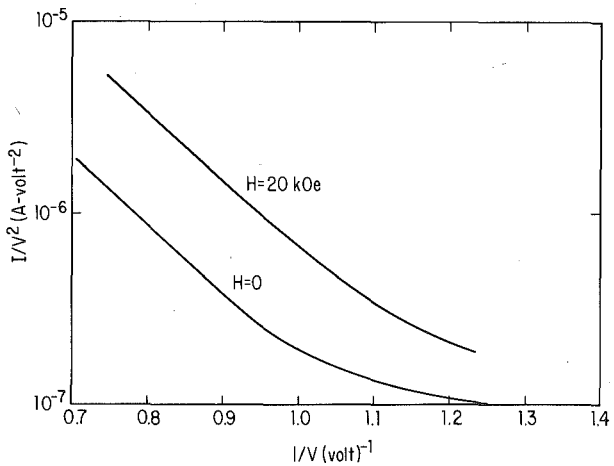

Fig. 19 J/V^2 vs. I/V characteristic for Al-EuSe-Au junction at $H = 0$ and 20 kOe.

verified that application of magnetic fields affects only the constant b, reducing it with increasing field.

The temperature dependence of the EuSe junction is shown in Fig. 20 in terms of the voltage (normalized to its value at 4.2°K) required to maintain a constant current through the junction. The full line illustrates the results with zero magnetic field and the dotted line with a field of 20 kOe. The interesting features of these curves are: (1) a rapid change near the magnetic-transition temperature of EuSe, and (2) no temperature dependence above and below the transition temperature, except for a gradual decrease above 20°K.

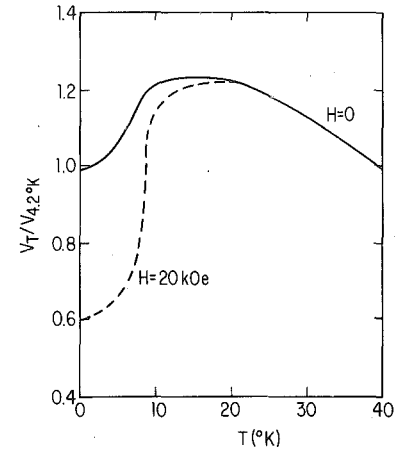

Fig. 20 $V_T/V_{4.2°K}$ vs. temperature for Al-EuSe-Au junction at $H = 0$.

Esaki et al.[56] explain the above observations as follows: They assume that the barrier ϕ of the junction extends from the Fermi level of the electrodes up to the bottom of the conduction band provided by the $5d$ band of the EuSe, and that the conduction process is due to Fowler-Nordheim tunneling through this barrier [see Fig. 9 and Eq. (22)]. They estimate

$$\phi = 0.5\text{ eV}$$

at the Al-EuSe interface from the constant b, which is related to ϕ by [see (22)]

$$b = \frac{4\sqrt{2}m^{\frac{1}{2}}\phi^{\frac{3}{2}}}{3e\hbar} \tag{32}$$

Because b is related to only one of the material constants, namely, ϕ, it is this parameter only that is affected by the magnetic transition at 4.2°K and the applied magnetic field. Esaki et al. conclude, therefore, that the magnetic perturbation of the barrier height is due to spin ordering of the $4f$ electrons of the EuSe, which causes a shift of the $5d$ band with respect to the electrode Fermi level.

5. HIGH-FIELD EFFECTS

a. Electrode-limited Processes

(1) Image Forces The abrupt changes in potential at the metal-insulator interface we have been showing in our energy diagrams are physically unrealistic, since abrupt changes in potential imply infinite electric fields. In actual fact the potential step changes smoothly as a result of the image force.[57] This arises as a result of the metal surface's becoming polarized (positively charged) by an escaping electron, which in turn exerts an attractive force $e^2/16\pi\epsilon_0 K^* x^2$ on the electron. The potential energy of the electron due to the image force is thus

$$\phi_{im} = -\frac{e^2}{16\pi\epsilon_0 K^* x} \tag{33}$$

where x is the distance of the electron from the electrode surface. The dielectric constant K^* used in (33) is the *high-frequency* constant, since in the course of emission from the cathode the escaping electrons spend only an extremely short time in the immediate vicinity of the surface.

Image-force effects play an important role in the conduction process when the current is electrode-limited, as we will see.

(2) Neutral Contact—Theory The potential step (with respect to the Fermi level) at a neutral barrier with attendant image potential as a function of the distance x from the interface is given by

$$\phi(x) = \phi_0 + \phi_{im} = \phi_0 - \frac{e^2}{16\pi\epsilon_0 K^* x} \tag{34}$$

The barrier $\phi(x)$ in the presence of image forces is illustrated by the line AB in Fig. 21. Clearly (34) is not valid at the electrode surface, since $\phi = -\infty$ there. Schottky circumvents this anomaly by assuming that the image force holds only for x greater than some critical distance x. For $x < x_0$, he assumes a constant image force; that is, the potential energy is a linear function of x, and such that it matches the bottom of the electrode conduction band at the surface (see Fig. 21).

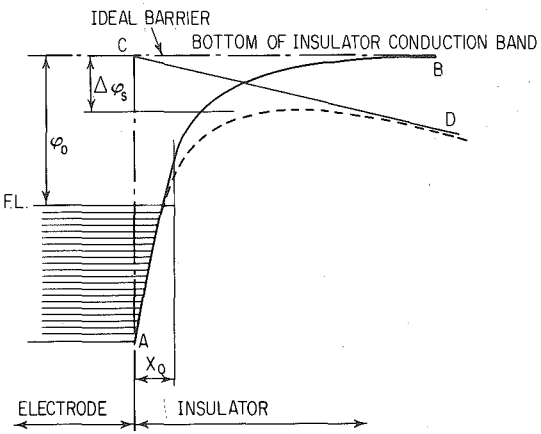

Fig. 21 Schottky effect at a neutral contact.

When an electric field exists at a metal-insulator interface, it interacts with the image force and lowers the potential barrier. This can be seen by reference to Fig. 21. The line CD represents the potential due to a uniform field, which when added to the barrier potential $\phi(x)$ produces the potential step shown by the dotted line, which is seen to be $\Delta\phi_s$ lower than without the electric field. The potential energy of the barrier under the influence of the field with respect to the Fermi level of the electrode is now given by

$$\phi(x) = \phi_0 - \frac{e^2}{16\pi K^*\epsilon_0 x} - eFx$$

This equation has a maximum at $x_m = (e/16\pi K^*\epsilon_0 F)^{\frac{1}{2}}$. The change $\Delta\phi\,[= \phi_0 - \phi(x_m)]$ in the barrier height due to the interaction of the applied field with the image potential is thus given by

$$\Delta\phi_s = \left(\frac{e^3}{4\pi K^*\epsilon_0}\right)^{\frac{1}{2}} F^{\frac{1}{2}} \equiv \beta F^{\frac{1}{2}} \tag{35}$$

In conventional units $\Delta\phi_s = (3.8 \times 10^{-4}/K^{*\frac{1}{2}})F^{\frac{1}{2}}$.

Because of image-force lowering of the barrier, the electrode-limited current does not saturate according to the Richardson law $J = AT^2 \exp(-\phi_0/kT)$ but rather obeys the Richardson-Schottky[57] law:

$$J = AT^2 \exp\left(-\frac{\phi_0 - \Delta\phi}{kT}\right) \tag{36}$$

$$J = AT^2 \exp\left(-\frac{\phi_0}{kT}\right) \exp\frac{\beta_s F^{\frac{1}{2}}}{kT} \tag{37}$$

where $A = 4\pi em(kT)^2/h^3$. When I is expressed in A cm^{-2}, A takes the value 120. Equation (36) was first applied successfully to metal-vacuum interfaces. Its use in insulators has to be qualified: Simmons[58] has pointed out that it holds only if the

electron mean free path is of the order of the insulator thickness; otherwise the pre-exponential factor becomes $2e(2\pi mkT/h^2)^{\frac{1}{2}}\mu V/S$; and Crowell[59] has shown that the effect of the electron's having differing masses in the metal and insulator is to change m to m^*, the effective electron mass, in the constant A.

When high fields exist at the interface, tunneling through the potential barrier can dominate the conduction process, and the relevant I-V characteristic is given by (22). We have seen in Sec. 4 that this mechanism requires that the thickness of the barrier at the electrode Fermi energy be less than about 50 Å. Thus the onset of field emission at a neutral contact occurs at a field given by $F \simeq \phi_0 \times 10^8/50 = 2 \times 10^5\phi_0$ V cm^{-1}.

A comparison of Richardson-Schottky current and tunnel currents in thin insulators has also been studied; the reader is referred to the literature[60,61] for details.

(3) Neutral Contact—Experimental The Richardson-Schottky effect in insulators appears to have been first observed by Emptage and Tantraporn,[62] who reported a log I vs. $F^{\frac{1}{2}}$ relationship in their samples; since then there have been many other reported similar observations. Mead,[63] however, has pointed out that a log J vs. $F^{\frac{1}{2}}$ relationship does not necessarily imply the Richardson-Schottky effect, but has reported convincing evidence of the effect. His data for Zn-ZnO-Au samples are shown in Fig. 22. The slope of the straight line (β_s/kT) corresponds to a thickness of 61 Å, whereas the capacitively determined thickness was 57 Å. Furthermore, the barrier height determined from thermal-activation measurements gave $\phi = 0.75$ eV, while the intercept of the activation curves yielded an intercept of $\phi \simeq 0.8$ eV; photoemissive measurements also apparently yielded similar barrier heights. These results are good evidence for Schottky emission.

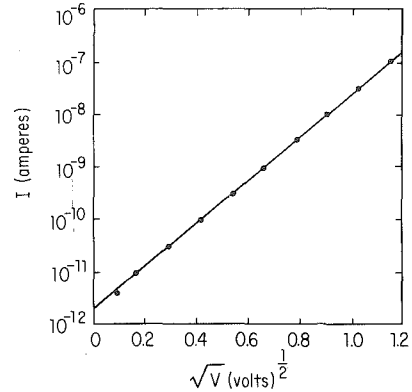

Fig. 22 I-V characteristic of a Zn-ZnO-Au structure.

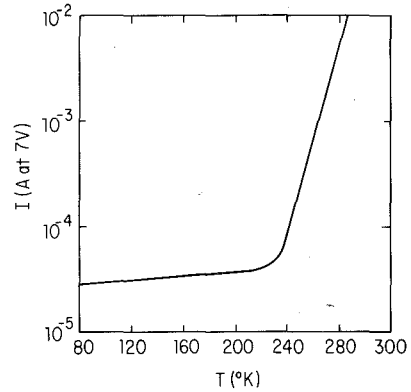

Fig. 23 Temperature variation of current through a Pb-Al_2O_3-Pb sample. Thickness of Al_2O_3 = 340 Å and V = 7 V.

Because thermal emission falls off rapidly with decreasing temperature, and since field emission is virtually temperature-insensitive (Sec. 4c), it means that low temperatures are amenable to field emission dominating the conduction process. Pollack[64,65] observed the electrode-limited conduction-process change from the Richardson-Schottky to the field-emission mechanism in both relatively thick and very thin insulators as the temperature of the sample was lowered from room to liquid-nitrogen temperatures (see Fig. 23).

(4) Blocking Contact The potential energy of the barrier (Schottky) of a blocking contact (Sec. 3c) with attendant image potential is, from (8) and (33),

$$\phi(x) = \phi_0 - \frac{N_d e^2}{K\epsilon_0}\lambda x - \frac{x^2}{2} - \frac{e^2}{16\pi K^*\epsilon_0 x}$$

This barrier has a maximum at $x_m \simeq 1/[4(\pi N_d \lambda)^{\frac{1}{2}}]$. In this case the image force lowers the potential barrier by an amount

$$\Delta\phi = \phi_0 - \phi(x_m) = \left[\frac{(V + \psi_m - \psi_i)e^7 N_d}{2(8\pi)^2(\epsilon_0 K^*)^3}\right]^{\frac{1}{4}} \tag{38}$$

Substituting (38) into (36) yields, in conventional units,

$$J = 120T^2 \exp\left\{-\frac{\phi_0 - 8.26 \times 10^{-6}[N_d(V + \psi_m - \psi_i)/K^{*3}]^{\frac{1}{4}}}{kT}\right\} \tag{39}$$

As at the neutral contact, tunneling will occur within the barrier when the field at the interface is sufficiently high to reduce the width of the barrier measured at the electrode Fermi level to about 50 Å or so. For $V > \phi_0$, the barrier at the interface can be considered approximately triangular; so we may use the Fowler-Nordheim expression (22) to estimate the current tunneling through the contact. Hence, substituting (11) into (22) yields, using conventional units,[66]

$$J = \frac{1.22 \times 10^{-11} N_d(V + \psi_m - \psi_i)}{\phi K^{*3}} \exp\left\{-\frac{3.65 \times 10^{10}(K^*\phi^3)^{\frac{1}{2}}}{[N_d(V + \psi_m - \psi_i)]^{\frac{1}{2}}}\right\} \tag{40}$$

We are not aware of any incontrovertible experimental evidence for these processes in thin film insulators, but the tunneling process is probably the dominant mechanism in the electrode-limited range described in Sec. 5*c*(3).

b. Poole-Frenkel Effect

(1) Theory The Poole-Frenkel effect[67,68] (field-assisted thermal ionization) is lowering of a Coulombic potential barrier when it interacts with an electric field, as shown in Fig. 24. This process is the bulk analog of the Schottky effect at an interfacial

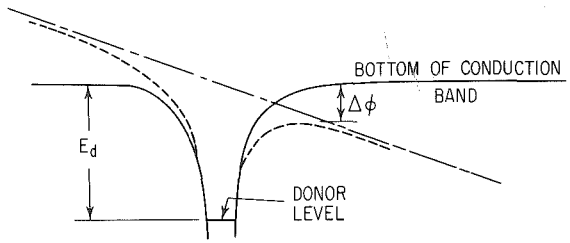

Fig. 24 Poole-Frenkel effect at a donor center.

barrier. Since the potential energy of an electron in a Coulombic field $-e^2/4\pi\epsilon_0 Kx$ is four times that due to image-force effects, the Poole-Frenkel attenuation of a Coulombic barrier $\Delta\phi_{PF}$ in a uniform electric field is twice that due to the Schottky effect at a *neutral* barrier:

$$\Delta\phi_{PF} = \left(\frac{e^3}{\pi\epsilon_0 K^*}\right)^{\frac{1}{2}} F^{\frac{1}{2}} \equiv \beta_{PF} F^{\frac{1}{2}} \tag{41}$$

This result was first applied by Frenkel[67,68] to the host atoms in *bulk* semiconductors and insulators. He argued that the ionization potential E_g of the atoms in a solid are lowered an amount given by (41) in the presence of a uniform field. Thus the conductivity is obtained by substituting $E_g - \Delta\phi_{PF}$ for E_g in (1), yielding a field-dependent conductivity of the form

$$\sigma = \sigma_0 \exp\left(\frac{\beta_{PF} F^{\frac{1}{2}}}{2kT}\right) \tag{42}$$

where σ_0 [$= e\mu N_c \exp(-E_g/2kT)$] is the low-field conductivity. Equation (42) may be written in the form

$$J = J_0 \exp\left(\frac{\beta_{PF} F^{\frac{1}{2}}}{2kT}\right) \tag{43}$$

where J_0 ($= \sigma_0 F$) is the low-field current density.

It is interesting to note that although $\Delta\phi_{PF} = 2\Delta\phi_s$, the coefficient of $F^{\frac{1}{2}}$ in the exponential is the same for both the Richardson-Schottky [at a neutral contact, see Eq. (37)] and Poole-Frenkel *J-F* characteristics (i.e., $\beta_{PF}/2 = \beta_s$). Mead[69] has suggested, however, that since traps abound in an insulator and that a trap having a Coulombic-type barrier would experience the Poole-Frenkel effect at high fields, thereby increasing the probability of escape of an electron immobilized therein, the current density in thin film insulators containing *shallow* traps is given by

$$J = J_0 \exp\left(\frac{\beta_{PF} F^{\frac{1}{2}}}{kT}\right) \tag{44}$$

Note in this case that the coefficient of $F^{\frac{1}{2}}$ is *twice* that in (43). Mead[63] also first reported field-dependent conductivity apparently of the form given by (44), and for this reason (44) is usually the form of Poole-Frenkel equation associated with *thin film* insulators rather than that given by (43).

(2) Experimental From what has been said it follows that it should be possible to differentiate between the Schottky and Poole-Frenkel effects in *thin-film* insulators from their different rates of change of conductivity with field strength;[67] viz., a plot of ln J vs. $F^{\frac{1}{2}}/kT$ results in a straight line of slope β_s or β_{PF}, depending on whether the conduction process is Richardson-Schottky or Poole-Frenkel. These experimentally determined slopes can be compared with the theoretical β_s and β_{PF}, which can be calculated quite accurately provided the high-frequency dielectric constant K^* for the material is known. Alternatively, the high-frequency dielectric constant may be determined from the slopes assuming the controlling mechanism is known. The resulting K^* should satisfy the equation $K^* = n^2$, where n is the refractive index for the material.

Several investigators have observed a field-dependent conductivity of the form given by (44) in what is apparently *bulk-limited* conduction in Ta_2O_5 and SiO films.[70–72] Johansen[71] has noted, however, that the coefficient of $F^{\frac{1}{2}}/kT$ is compatible with the Schottky effect rather than the Poole-Frenkel effect. On closer examination of the results of Mead[69] and of Hirose and Wada,[70] although they claim Poole-Frenkel emission, it is apparent that their experimental β is actually compatible with Schottky emission.[73] These authors based their conclusions on dielectric-constant values which are typically four times too high for the high-frequency dielectric constants of the materials used. Hartman et al.[72] conclude that neither the Richardson-Schottky equation (37), because conductivity is bulk-limited, nor the Poole-Frenkel equation (44), because of the incompatibility of the experimental and theoretical β, can adequately explain their results.

In actual fact in films of Ta_2O_5 and SiO, because of their wide energy gaps, in order to have any detectable current, even in the absence of traps, it is necessary to have donor or acceptor centers within the insulator to supply the necessary carriers (see Ref. 73 and Sec. 2). Furthermore, if it is assumed that the insulator contains donor centers which lie below the Fermi level—this assumption is supported by the fact that the conductivity of the films continues to increase with increasing temperature above room temperature—and shallow *neutral* traps (see Fig. 25), the bulk *J-V* characteristic of the film is given by[73]

$$J = J_0 \exp\left(\frac{\beta_{PF} F^{\frac{1}{2}}}{2kT}\right) \tag{45}$$

where

$$J_0 = e\mu N_c \left(\frac{N_d}{N_t}\right)^{\frac{1}{2}} F \exp\left(-\frac{E_d + E_t}{2kT}\right)$$

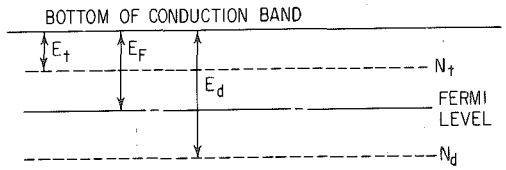

Fig. 25 Energy diagram of insulator film used in deriving (45).

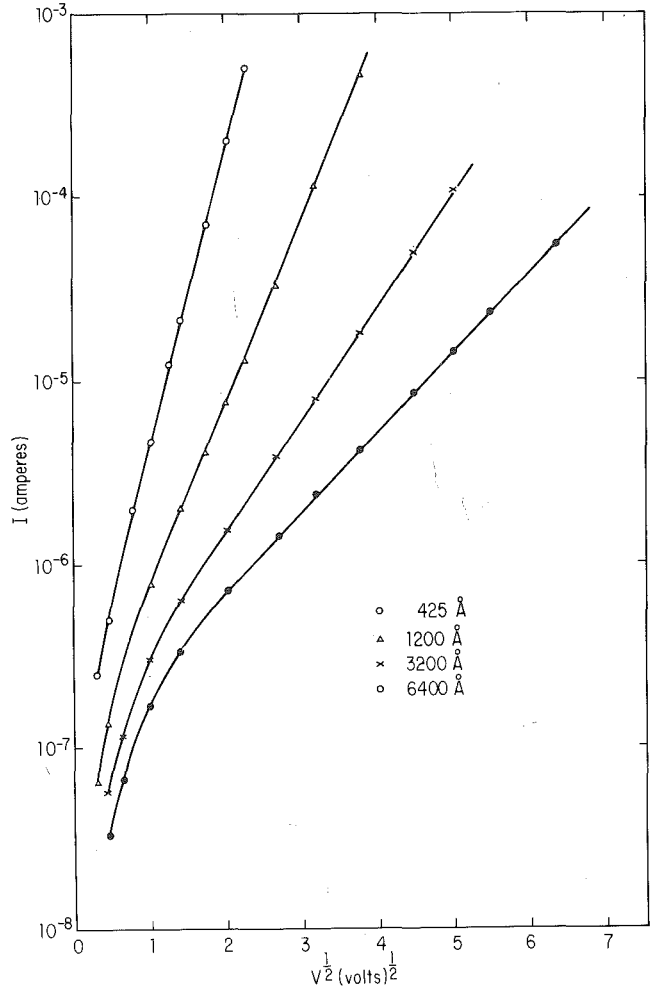

Fig. 26 I-$V^{\frac{1}{2}}$ characteristic of Al-SiO-Al samples illustrating the Poole-Frenkel effect in the higher-voltage range.

Thus in this case the coefficient of $F^{\frac{1}{2}}/kT$ is $\beta_{PF}/2$ ($= \beta_s$) *even though the conductivity is not electrode-limited,* which explains the anomalous experimental results.

Stuart[74,75] has recently shown that his bulk-limited I-V characteristics can be explained in terms of (45). Figure 26 shows a plot of his I-V characteristic for four film thicknesses,[74] which are observed to be linear on a $\ln I$ vs. $V^{\frac{1}{2}}$ plot. In Table 3 we have transcribed his calculation of the ratios of the experimentally determined $\beta(\beta_{exp})$ and the theoretical $\beta(\beta_{theor})$ assuming (45) as the relevant J-V characteristic.[74,75] The agreement between β_{exp} and β_{theor} is seen to be excellent over the entire thickness range, which covers almost four decades.

TABLE 3 Stuart's[74,75] Results for $\beta_{exp}/\beta_{theor}$ Obtained from SiO Films

s, μ	0.043	0.12	0.32	0.64	1.2	3.3	7.0	13.7
$\beta_{exp}/\beta_{theor}$	1.06	1.07	1.09	1.07	1.04	1.05	1.03	1.00

c. Electrode-limited to Bulk-limited Process

(1) Physical Concepts[66] Suppose we have a heavily doped ($N_d \gtrsim 10^{18}$ cm^{-3}) insulator with the band structure shown in Fig. 3*d*, so that the depletion region is very thin (see Table 2). Although the donor density is high, the conductivity of the bulk will be reasonably low, particularly if traps are present, since the donors lie below the Fermi level, which means a few only will be ionized. Under these conditions it is possible to observe an electrode- to bulk-limited transition in the conduction process, as illustrated by the following qualitative arguments.

For initial voltage bias the conduction process will clearly be electrode-limited, because of the high cathode-insulator barrier. At very low voltages the electrode-limited conduction process is by the thermal excitation of electrons from the cathode over the interfacial barrier. The current is only a slow function of the applied voltage for this process [see (39)] and for barriers in excess of about 0.5 eV will be very low.

At higher voltages one of two processes can occur. If the barrier at the Fermi level becomes thin enough, field emission of electrons from the cathode into the conduction band of the insulator can occur [see Eq. (40)], or when the voltage bias exceeds $3E_g/2$, impact ionization can occur in the depletion region of the insulator. Both these processes are characterized by a rapid increase of current with applied voltage; i.e., the contact resistance decreases extremely rapidly with increasing voltage bias. Since the contact resistance under these conditions at low voltages is much higher than that of the bulk, it follows that while either or both of these processes predominate the I-V characteristic will be virtually *thickness-independent* (because essentially all the applied voltage appears across the contact and very little across the bulk), and very steep (see chain-dotted line in Fig. 27).

The electrode-limited process cannot continue indefinitely, however, because the bulk resistance [see (45)] decreases much more slowly with increasing voltage than does the contact (compare dotted and chain-dotted lines in Fig. 27). Thus, at some voltage V_T, the transition voltage, the contact resistance falls to a value equal to that of the bulk, and when this occurs the applied voltage V_T is shared equally between the contact and the bulk. Thereafter, practically all the voltage in excess of V_T will fall across the bulk and the remaining fraction across the barrier, just sufficient to ensure current continuity throughout the system. Hence, for voltages in excess of V_T, the current will cease to rise as rapidly as for $V < V_T$, since it is controlled by bulk processes, *and will be thickness-dependent.*

(2) Theoretical Results In Fig. 27 the theoretical J-V characteristics for three thicknesses of dielectric (S = 1, 3, and 10 μ) have been plotted, using (40) and (45) and the parameters $N_d = 10^{18}$ cm^{-3}, $N_t = 10^{19}$ cm^{-3}, $K^* = 3$, $\mu = 10$ cm^2, (V sec)$^{-1}$, $(E_t + E_d)/2 = 0.55$ eV, and $\phi = 1.5$ eV. For these parameters the current given by (39) is too small to be shown on the graph. The dotted and chain-dotted lines

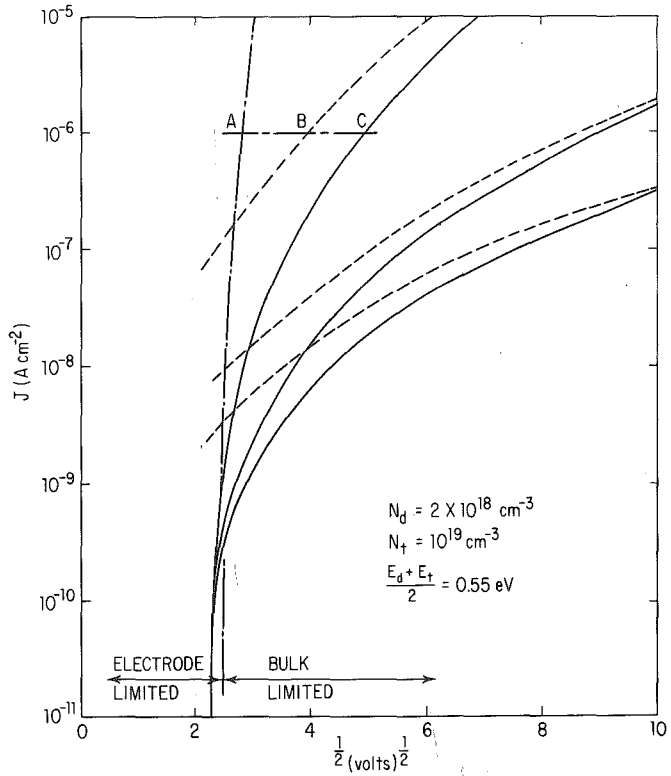

Fig. 27 Theoretical curves showing electrode- to bulk-limited characteristic.

represent the bulk and contact characteristics given by (45) and (40), respectively. The full lines represent the combined effect of the contact and bulk, that is, the actual characteristic of the junction. Since the contact and bulk impedances act in series, the actual curves are arrived at by adding together the contact and bulk voltages corresponding to a given current on the individual curves. Hence, to obtain the point C on the upper curve of Fig. 27 we add the voltages corresponding to the points A and B lying, respectively, on the individual contact and bulk curves.

The curves clearly delineate contact and bulk effects, the transitional stage occurring about 12 V. Below 12 V we have the contact-limited portion, which is characterized by the thickness-independent and rapidly increasing current-voltage tunnel curve. Above 12 V we have the bulk-limited region, which manifests the effect of varying insulator thickness. It will be noted that these curves are displaced to lower current densities and lower slopes than their corresponding individual curves, thus reflecting to a small extent contact effects on the bulk-limited characteristics.

(3) Experimental Results Stuart's[75] results, obtained from thin film Al-SiO-Al samples, are shown in Fig. 28. The thickness of the SiO films ranges from 1.2 to 13.7 μ. The interesting point about these results is their similarity to the theoretical curves in Fig. 27. Note in particular below $V \simeq 70$ V that the I-V characteristics are essentially independent of the film thickness, even though there is over a decade difference between the two extreme thicknesses. Above $V \simeq 70$ V the J-V characteristics are thickness-dependent and obey a relationship of the form given by (45).

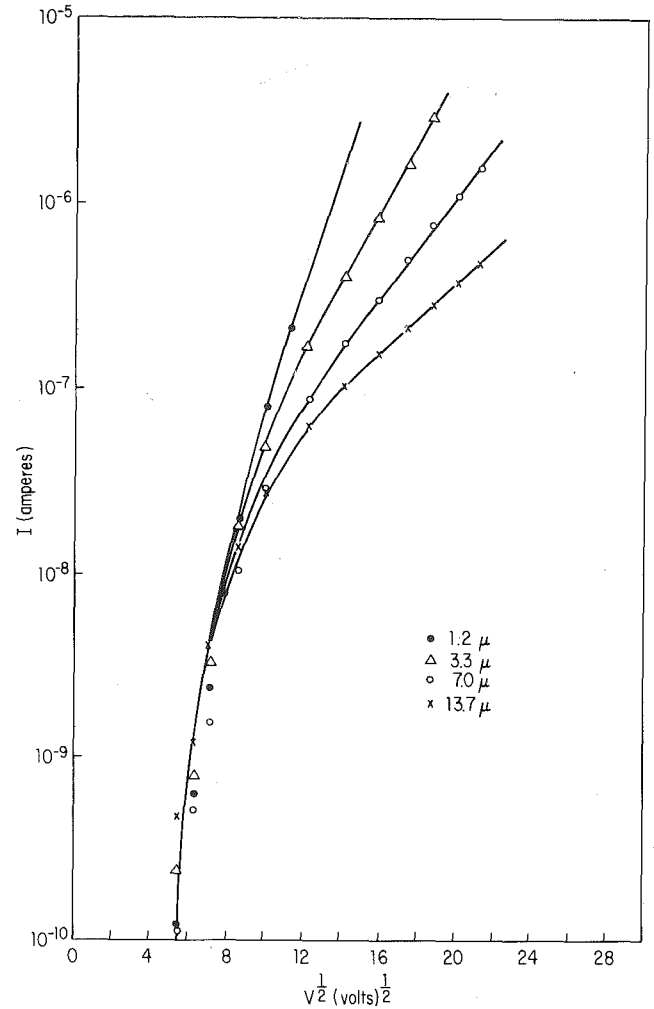

Fig. 28 Electrode-limited to bulk-limited conduction in Al-SiO-Al films. Electrode area 0.1 cm^2.

6. SPACE-CHARGE-LIMITED (SCL) CURRENTS IN INSULATORS

a. Physical Concepts

We have seen that an insulator which does not contain donors and which is sufficiently thick to inhibit tunneling will not normally conduct significant current. However, if an ohmic contact (Sec. 3a) is made to the insulator, the space charge injected into the conduction band of the insulator is capable of carrying current; this process is termed SCL conduction. In order to gain physical insight into this process, let us consider what happens when a bias is applied to the system shown in Fig. 29, that is, an insulator having two ohmic contacts on its surfaces.

The result of the applied bias is to add *positive* charge to the anode and *negative* charge to the cathode, as would be the case with any such capacitive system. Thus, as the voltage bias increases, the *net* positive charge on the anode *increases* and that on the cathode decreases. Calling the charge on the cathode Q_1, that on the anode Q_2, and the negative space-charge density $\rho(x)$, the condition of electrical neutrality demands that[76]

$$\int_0^s \rho(x)\,dx = Q_1 + Q_2 \qquad (45a)$$

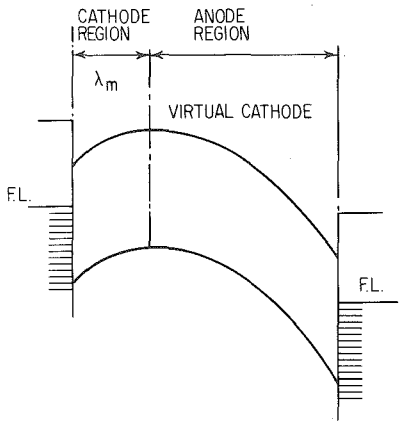

Fig. 29 Energy diagram illustrating virtual cathode, cathode region, and anode region under space-charge-limited condition.

Equation (45*a*) may be rewritten

$$\int_0^{\lambda_m} \rho(x)\,dx + \int_{\lambda_m}^{L} \rho(x)\,dx = Q_1 + Q_2$$

where λ_m is chosen such that

$$\int_0^{\lambda_m} \rho(x)\,dx = Q_1 \qquad (46)$$

and

$$\int_{\lambda_m}^{s} \rho(x)\,dx = Q_2 \qquad (47)$$

The insulator has thus been divided into two portions (see Fig. 29) with λ_m as the boundary separating the two.

The significance of (46) and (47) is that the *positive* charge on either contact is neutralized by an equal amount of *negative* charge contained between the contact and the plane at $x = \lambda_m$. Thus the field in the insulator due to Q_1 is zero for $x \geq \lambda_m$; that is, the negative charge between $x = 0$ and λ_m screens the insulator beyond $x = \lambda_m$ from conditions at the cathode-insulator interface. Similarly the field due to Q_2 is zero for $x \leq \lambda_m$. Since the field due to Q_1 and Q_2 is zero at $x = \lambda_m$, the net field there must be zero, as shown in Fig. 29, and for this reason the plane at $x = \lambda_m$ is termed the virtual cathode. The region $0 \leq x \leq \lambda_m$ is designated the cathode region, and the region $\lambda_m \geq x \geq s$ the anode region.

From a consideration of (46) and (47) and the fact that Q_1 *decreases* and Q_2 *increases* with increasing voltage, it will be clear that the virtual cathode moves closer to the cathode as the applied voltage increases; that is, the cathode region decreases and the anode region increases. Eventually, when $Q_1 = 0$, the virtual cathode coincides with the physical cathode-insulator interface. Under this condition, then, the anode region extends throughout the whole of the insulator, and an ohmic contact no longer exists at the cathode-insulator interface. Thus, for further increasing voltage bias, the conduction process is no longer space-charge-limited, but rather it is emission-limited.

b. Assumptions and Boundary Conditions

If the cathode makes a good ohmic contact to the insulator (e.g., $\phi \simeq 0$) the cathode region [i.e., the accumulation region at zero bias (see Table 1)] will normally be small compared with thickness of the insulator. We may therefore assume that under voltage bias the anode region occupies the whole of the insulator thickness; or alternatively stated, the virtual cathode coincides with the physical cathode (cathode-insulator interface). This assumption leads us to the following boundary conditions: (1) the field at $x = 0$ is zero, and (2) $V = 0$ and V_a, respectively, at $x = 0$ and s, where V_a is the applied voltage.

c. Trap-free Insulator

Since we are assuming the anode region extends throughout the insulator, we may neglect diffusion effect;[1] thus

$$J = \rho(x)\mu F \qquad (48)$$

and from Poisson's equation (2),

$$\frac{dF}{dx} = -\frac{\rho(x)}{K\epsilon_0} \qquad (49)$$

Substituting (49) into (48) yields

$$J = -\frac{\mu K\epsilon_0 F\,dF}{dx} = -\frac{(\mu K\epsilon_0/2)\,dF^2}{dx} \qquad (50)$$

which on integration, subject to the boundary condition $F = 0$ at $x = 0$, yields

$$F = -\frac{dV}{dx} = \left(\frac{2Jx}{\mu K\epsilon_0}\right)^{\frac{1}{2}} \qquad (51)$$

Integrating (51), and using $V = 0$ and V_a at $x = 0$ and $x = s$, we arrive at

$$J = \frac{9\mu K\epsilon_0}{8s^3} V^2 \qquad (52)$$

Equation (52) was originally derived by Mott and Gurney.[1] The interesting features of (52) are that it predicts that SCL current is proportional to V^2 and inversely proportional to s^3; both these predictions have since been confirmed by experiment. However, (52) predicts much higher currents than are observed in practice, and also that SCL currents are temperature-insensitive, which is also contrary to observation. These deviations from the simple trap-free theory are readily accounted for when a more realistic insulator, that is, one which contains traps, is considered.

d. Defect Insulator

The theory of SCL currents in defect insulators is due initially to Rose.[2] If the insulator contains traps, a large fraction of the injection space charge will condense therein, which means that the free-carrier density will be much lower than in a perfect insulator. Furthermore, since the occupancy of traps is a function of temperature, the SCL current is temperature-dependent.

Our starting point, as in the trap-free insulator, is (48) and (49), but now it is necessary to separate the space charge into a trapped ρ_t and a free ρ_f component; thus

$$J = \rho_f \mu F \qquad (53)$$

and

$$\frac{dF}{dx} = -\frac{\rho_f + \rho_t}{K\epsilon_0} \qquad (54)$$

(1) Shallow Traps If the insulator contains N_t shallow traps, positioned an energy E_t below the conduction band (see Fig. 30), then $\rho_f = eN_c \exp(-E_F/kT)$ and $\rho_t = eN_t \exp[-(E_F - E_t)/kT]$; thus

$$\theta \equiv \frac{\rho_f}{\rho_t} = \frac{N_c}{N_t} \exp\left(-\frac{E_t}{kT}\right) \qquad (55)$$

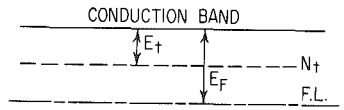

Fig. 30 Energy diagram showing shallow traps in an insulator.

An inspection of (55) shows that ρ_f can normally be neglected in comparison with ρ_t; for example, assuming $N_t = 10^{19}$ cm^{-3} and $E_t = 0.25$ eV, and taking $N_c = 10^{19}$ cm^{-3} at room temperature, we have $\theta < 10^{-5}$. Thus substituting (54) and (55) in (53) yields $J = e\mu K\epsilon_0 \theta F\,dF/dx$. By comparison with (50) and (52), the latter expression integrates to

$$J = \frac{9\mu K\epsilon_0 \theta}{8s^3} V^2 \qquad (56)$$

It will be clear, then, that the effect of shallow trapping enters into the theory by way of the factor θ, and since it is independent of V, $J\alpha V^2$ as in the trap-free case. We have seen, however, that θ is a very small quantity and temperature-dependent; thus the inclusion of shallow traps in the insulator brings the theory into line with experimental observation (see Sec. 6c).

Lampert[77,78] has noted that only when the injected free-carrier density n_i exceeds the volume-generated free-carrier density n_0 will space-charge effects be observed; when $n_0 > n_i$ the volume (ohmic) conductivity will predominate. Lampert calculates the voltage V_x at which the transition from ohmic to SCL conduction occurs to be

$$V_x = \frac{en_0s^2}{\theta K\epsilon}$$

Lampert[77,78] has also pointed out that if sufficient charge is injected into the insulator, the traps will become filled (trap-filled limit TFL). Further injected charge then exists as free charge in the conduction band and contributes in toto to the current. Beyond the TFL, then, the J-V characteristic will be given by (52) rather than (56), thus as V just exceeds V_{TFL}, the current rises rapidly by an amount θ^{-1}. The voltage V_{TFL} at which the TFL occurs is given by

$$V_{\text{TFL}} = \frac{eN_ts}{2K\epsilon_0}$$

In Fig. 31 we have schematically illustrated a typical I-V characteristic for an insulator having a shallow discrete trapping level. At the lower voltages ($V < V_x$), the characteristic is ohmic, because the bulk-generated current exceeds the SCL current. In the voltage range $V_x < V < V_{\text{TFL}}$ the SCL current predominates and $I\alpha V^2$, (56). When

$$V = V_{\text{TFL}}$$

sufficient charge has been injected into the insulator to fill the traps. Hence, as V just exceeds V_{TFL} the current rises rapidly such that for $V > V_{\text{TFL}}$ the I-V characteristic obeys the trap-free law, (52). Clearly, from the structure exhibited in the characteristic, much information about traps in insulators can be deduced from the experimental data.

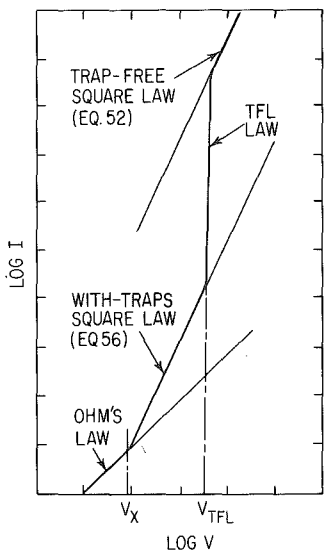

Fig. 31 SCL I-V characteristic for an insulator containing shallow traps.

(2) Exponential-trap Distribution In amorphous or polycrystalline structures, which are characteristic of thin film insulators, a distribution of trap levels is to be expected, rather than a discrete level of traps. Rose[2] has treated the case of SCL conduction in the presence of a distribution of trap levels that decreases exponentially in density with increasing energy below the conduction band; that is,

$$N_t = A \exp(-E/kT_c)$$

where E is the energy measured from the bottom of the conduction band, and T_c is a characteristic temperature greater than the temperature at which the currents are measured. In this case

$$I\alpha V^{(T_c/T+1)}$$

which means that, since $T_c > T$, the current increases more rapidly with applied voltage than in the trap-free or discrete trap-level cases.

e. Double Injection

If the cathode is ohmic to electrons and the anode ohmic to holes, SCL currents of both sign of carriers can flow in an insulator. Because the injected electrons and holes are capable of neutralizing each other, it follows that the number of injected

carriers at any point in the insulator can exceed the space charge existing there. It follows, then, that double-injection currents will be larger than those due to single-carrier injection under the same condition of voltage bias. Furthermore, in the case of double injection, recombination will take place between the two types of carriers, usually through recombination centers situated in the forbidden gap of the insulator. The effect of charge neutralization and recombination results in the double-injection J-V characteristic's being even richer in structure than that of the single-carrier characteristic.[79,80]

The theory of double injection with attendant recombination processes is clearly much more complex than the single-carrier theory. Furthermore, it is not possible to ignore the effect of diffusion currents, as is done in the single-carrier theory. There is also the problem of making suitable contacts to an insulator, for an electrode that is ohmic to electrons will normally be blocking to holes. The problem of obtaining even one type of ohmic contact to a wideband insulator is difficult; making two different types is formidable. Indeed we are not aware of any unequivocal evidence for double injection in the thin film, wideband insulators with which we are primarily concerned. Thus in view of all the above-mentioned difficulties, it is inappropriate to discuss double injection further here. For further details of the theory of double injection and the experimental evidence for it on bulk crystals, we refer the reader to the review by Lampert.[3]

f. Experimental Results

Very few experimental data on SCLC in solids appear in the literature; there are probably two main reasons for this. First, the problem of making ohmic contacts to

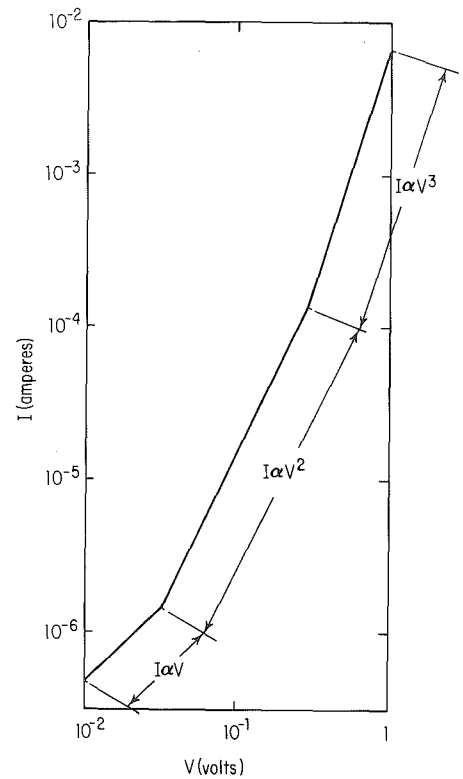

Fig. 32 Space-charge-limited conduction in In-CdS-Au samples.

wideband insulators is formidable. Second, in thin films it is clear that the application of only 1 or 2 V induces fields often sufficient to cause the conduction process to be emission-limited.

Zuleeg[81] has reported SCLC in In-CdS-Au sandwiches, in which the indium electrode made the ohmic contact. His results are shown in Fig. 32. At low voltages the I-V characteristic is ohmic, apparently because of thermally generated carriers. At high voltages $I \alpha V^3$. Over the intermediate-voltage range $I \alpha V^2$, suggesting space-charge-limited conduction. There is no apparent trap-filled limit in this case. Zuleeg and Muller[82] have investigated the effect of the CdS thickness in In-CdS-Au samples in the range where $I \alpha V^2$, by plotting log I (at constant voltage) vs. log s. Their results, which are shown in Fig. 33, were taken from three batches of differing thick-

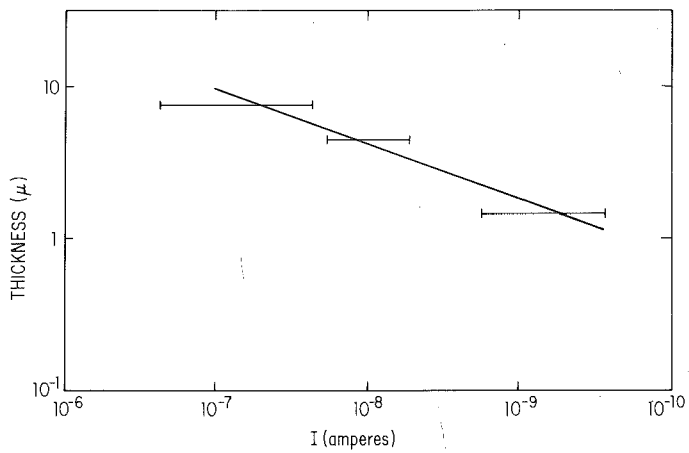

Fig. 33 Logs plotted against log I for three batches of In-CdS-Au samples.

ness; the variation from sample to sample is indicated by the bars. It is seen that a straight line drawn through the average of each batch has a slope of -3, which is in agreement with (56).

7. NEGATIVE RESISTANCE AND MEMORY EFFECTS

It has been observed that when thin insulating films between approximately 200 and 3,000 Å thick have undergone a forming process, the conduction process manifests voltage-controlled negative resistance,[83–96] reversible voltage,[90,94,95] thermal-voltage memory effects,[90,94,95] temperature-independent conductivity,[90,94] and electroluminescence.[91] These effects have been observed in a wide variety of insulators and electrode materials, but we will confine our attention to a series of measurements recently carried out in the Al-SiO-Au system.[94,95]

a. Forming Process and DC Characteristic

The resistance of an unformed sample, when measured at room-ambient pressure, is very high. The V-I characteristic is illustrated in Fig. 34 and is similar to that found in thicker SiO films [see Sec. 5b(2)], that is, log $I \alpha V^{\frac{1}{2}}$ and is very temperature-sensitive. If the sample is now placed in a modest vacuum, a few microns pressure or less, its conductivity is observed to increase sharply when a voltage of about 5 V is applied across the insulator with the gold electrode positively biased. Upon reducing

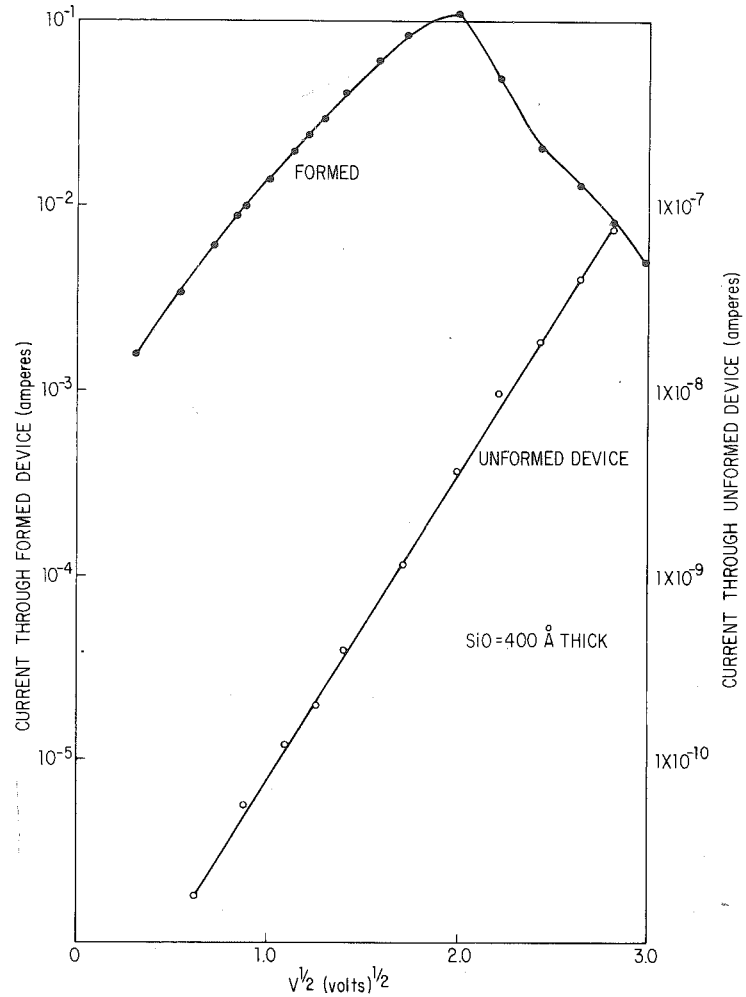

Fig. 34 I-V characteristic of unformed and formed sample, on a ln I vs. $V^{\frac{1}{2}}$ plot, illustrating the large change in conductivity that occurs on forming.

the voltage slowly to zero volts, a V-I characteristic with a pronounced voltage-controlled negative-resistance region is observed, as shown in Fig. 35. Any subsequent dc voltage sweep generates an identical characteristic to that shown in Fig. 35; therefore, the change in conductivity is permanent. The sample has clearly undergone some form of forming process.[96]

The interesting feature of the I-V curves of the formed sample is the pronounced negative-resistance region which is observed for $V \simeq 4$ V. The peak-to-valley ratios of the curves are typically 100:1, although ratios as high as 1,000:1 have been observed. A second interesting point is that the voltage at which the peak current occurs is not only practically independent of insulator thickness but has also been shown to vary little (3 to 5 V) with insulator and electrode material; the peak current is, however, dependent on the nature and thickness of the insulator.

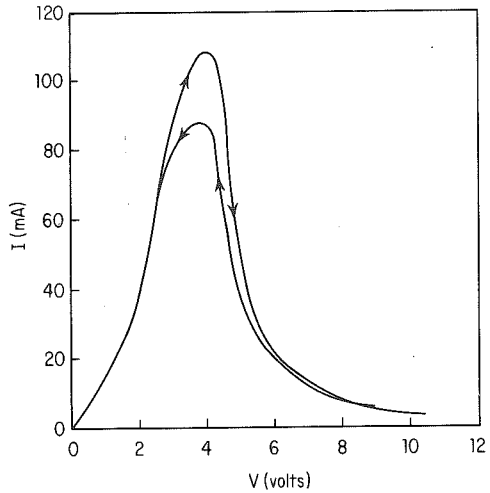

Fig. 35 I-V characteristic of formed sample.

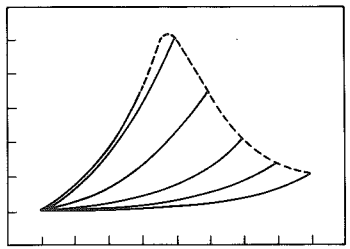

Fig. 36 Oscillogram of V-I traces for 1,000-Hz applied voltages for various voltage amplitudes (y = 10 mA cm^{-1}, x = 1 V cm^{-1}). The dotted line illustrates the dc characteristic.

b. Dynamic Characteristics

In the voltage range 0 to 2.5 V, the ac ($\gtrsim$1,000 Hz) V-I characteristic follows the dc low-impedance characteristic. However, for applied ac voltages in excess of 2.5 V, which is designated the threshold voltage V_T, a new set of V-I characteristics are apparent, as shown by the oscillogram in Fig. 36. In the range $V_T \lesssim V \lesssim 9$ V a distinct ac V-I characteristic can be associated with every voltage amplitude, the overall impedance of the characteristic increasing with increasing voltage amplitude; that is, the sample resistance to alternating current is a function of voltage amplitude. These characteristics do not manifest negative resistance, although the locus of their end points, for increasing values of voltage amplitude beyond 4 V, generate the dc negative-resistance characteristic, as shown by the dotted line in Fig. 36.

c. Memory Characteristics

If the voltage corresponding to any operating point on the negative-resistance characteristic shown in Fig. 37 (for example, the points B, C, E), is reduced to zero in about 0.1 ms, it is found on reapplication of voltage that the low-impedance characteristic OA is not generated, but rather a higher-impedance state prevails. For example, a voltage pulse with a trailing edge faster than 0.1 ms, and of amplitude corresponding to the voltage of the operating points E, C, B, generates, respectively, the new impedance states OE', OC', and OB'. These induced, or *memory*, states appear to prevail indefinitely provided that no peak voltage in excess of V_T is applied while the sample is in a memory state. The original low-impedance state AO can be regenerated, that is, the memory erased, by applying a voltage slightly in excess of V_T.

It will be apparent from a consideration of the above observations that the area enclosed by the perimeter $OAB'C'E'O$ represents the electrical stable characteristic of the sample; that is, any operating point therein is obtainable and stable. Operating points lying in the area enclosed by the perimeter $EC'B'ABCEE'$ are unstable except for points on the curve $ABCE$, which are stable.

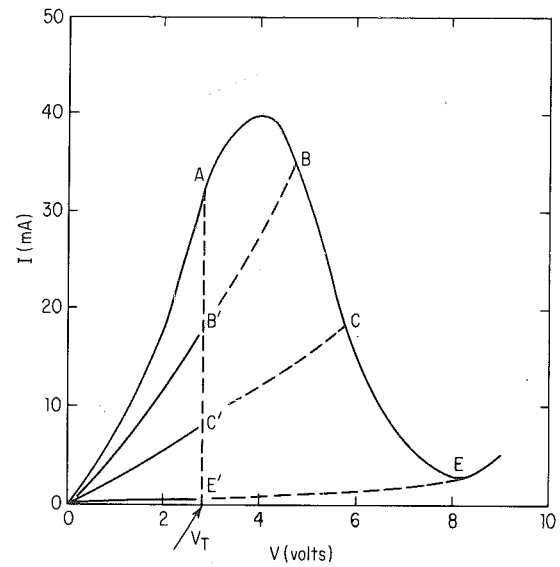

Fig. 37 Typical I-V diagram illustrating several impedance states and the threshold voltage for Al-SiO-Au junction.

Switching from the low- to the high-impedance state, as described above, can be accomplished with a voltage pulse as narrow as 2 or 3 ns wide. Switching from the high-impedance to the low-impedance state can be accomplished by a voltage pulse slightly greater than V_T in height and approximately 100 ns wide. In view of these facts, it would appear that repetitive switching from the high- to low-impedance state and back again should be accomplished in a cycle time of about 100 ns. It transpires, however, that although the sample can be switched from the *high- to the low-impedance* state and back again in about 100 ns, any immediate attempt to repeat the process is inhibited for a period of time which varied from milliseconds to fractions of a second dependent, apparently, upon the insulator preparation. This effect is designated the *dead time*.

It has been shown that these type of samples have potential as nonvolatile, nondestructive readout, analog, or binary memory elements.[95]

d. Connection between DC and Dynamic Characteristics and Memory

The difference between the dc characteristic (Fig. 35) and the dynamic characteristic (Fig. 36) is ascribed to the memory phenomena when the rate of fall of the applied sinusoidal voltage waveform is fast enough to set the sample into an induced-memory state. Thus, supposing the sample has impressed on it a voltage of amplitude corresponding to the voltage of the operating point B in Fig. 37, as the voltage falls rapidly from this value to zero, the characteristic $BB'O$ is generated rather than BAO. On the subsequent rising voltage sweep the curve $OB'B$ is again swept out. The reason why the sample does not switch to its low-impedance state as the voltage exceeds V_T is that the time between the voltage being removed and then reapplied to its maximum value is less than the dead time of the sample.

The curve $BB'O$ will continue to be swept out provided the amplitude remains constant. If the amplitude is increased slightly, the sample will partially switch to, say, the point C in Fig. 37, and the operating characteristic is then $OC'C$. The switching thus described is the case of a partial switch as compared with the complete switch described by the transition from A to E shown in Fig. 37.

e. Thermal Characteristics

(1) Thermal-voltage Memory Figure 38 is an oscillogram of the dc V-I characteristics displayed on a X-Y oscilloscope for an applied sawtooth pulse with a rise time of 10 s and a fall time of 2 s. Curves a and b are the V-I traces for a sample at 300 and 77°K, respectively. The slight difference in the two traces for *increasing* voltage indicates that the V-I characteristic is virtually temperature-independent. The return trace to zero volts for the sample at 300°K does not follow the increasing voltage trace because of partial memory storage (cf. Fig. 35). The return trace at 77°K is generated independent of the rate at which the voltage is removed and never exhibits any partial switching. Any subsequent voltage cycling (fast or slow) of the sample up to 9 V at the low temperatures invariably generates the same high-impedance characteristic. Hence at low temperature the state cannot be erased; the induced state at low temperatures is thus irreversible. Any similar low-temperature irreversible V-I characteristic lying between the low-impedance and high-impedance characteristic OA and $OE'E$ shown in Fig. 37 can be induced simply by limiting the maximum voltage excursion to lie between $V = V_T$ and $V = 9$ V respectively; for example, limiting the maximum voltage excursion to 5 V results in the V-I characteristic $OB'B$.

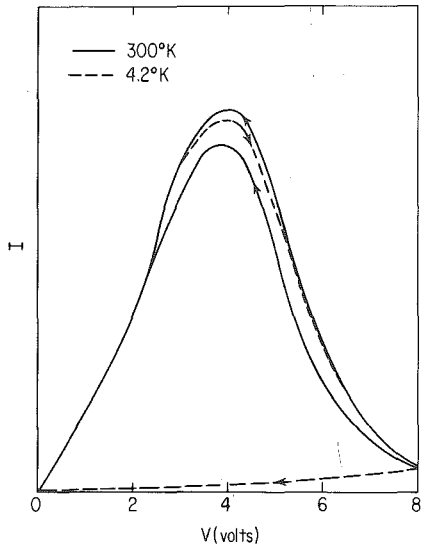

Fig. 38 Typical I-V trace for a complete voltage cycle between 0 and 9 V at (a) 300°K and (b) 77°K.

These low-temperature memory states are irreversible only if the sample is held at low temperature. Raising the sample temperature to 300°K and applying to it a voltage slightly in excess of V_T causes it to switch back into its initial low-impedance state OA. *In fact, the sample will switch back to the low-impedance state even if the sample temperature is lowered first to 77°K and then the voltage ($V > V_T$) is applied.* Thus the sample manifests a thermal-voltage memory in addition to a voltage memory.

(2) Voltages below V_T A somewhat different thermal characteristic is observed if the applied voltage is confined to values less than V_T. In this case one does not observe a thermal memory; that is, increasing and decreasing voltages generate identical I-V characteristics. The sample conductivity at liquid-nitrogen temperature is typically 10% less than the room-temperature characteristic and obeys a relationship of the form

$$\sigma = \sigma_0(1 + \alpha T^2) \tag{57}$$

where α for applied voltages above 1 V is approximately 1×10^{-6} per $(°K)^2$ (see Fig. 39). Equation (57) suggests that the increase in conductivity with increasing temperature is not due to the thermal creation of free carriers, since no activation-energy term is involved.

f. Theory of Operation

Simmons and Verderber[94,95] have postulated that the forming process introduces a broad band of localized levels within the normally forbidden band of the insulator. The electrons are assumed to move through the insulator by tunneling between adjacent sites within the impurity band; it is also assumed that the electrons can, under certain conditions, be trapped within the localized band. A model based on these ideas accounts in a self-consistent manner for all the experimental observations.

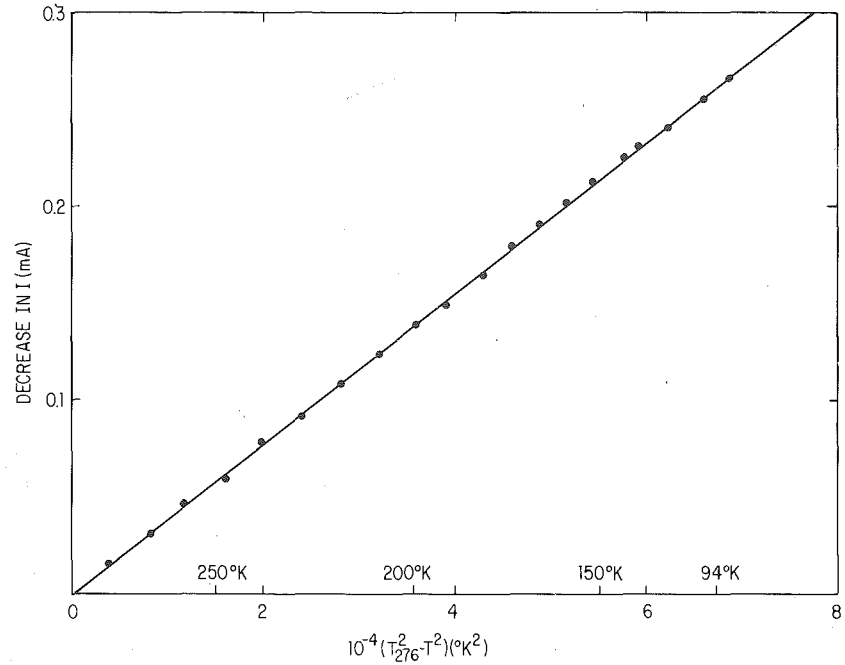

Fig. 39 Plot of decrease in current from 276 to 94°K as a function of the absolute temperature squared. Applied voltage 1.5 V; current at 276°K is 2.25 mA.

A detailed account of this theory is, however, beyond the scope of this chapter, and the reader is referred to the literature for further details.[94]

Hickmott[91] has also presented a theory of operation in order to account for his observations.

8. PRINCIPLES OF HOT-ELECTRON THIN FILM DEVICES

a. Tunnel-cathode Emitter

(1) Principle of Tunnel Cathode The principle of the tunnel cathode, as proposed by Mead,[97] is based on the fact that the electron energy is conserved during the tunnel process. Thus in tunneling from one electrode to the other, such that $eV < \phi_0$, the electrons enter the positively biased electrode with energy eV above the Fermi level of the electrode, as shown in Fig. 40a. The electrons then give up their energy to the lattice and fall into the Fermi sea.

For voltage biases such that $eV > \phi_0$ the electrons first tunnel into the conduction band of the insulator before entering the positively biased electrode. The electron is then assumed to be accelerated by the field within the insulator, without undergoing energy losses, to again enter the positively biased electrode with energy eV above the Fermi level. If eV is less than the positively biased electrode work function ψ, the electron gives up its energy to the electrode lattice as previously described. However, if $eV > \psi$ and the thickness of the positively biased electrode is less than the mean free path of the electrons, Mead[97] suggests that the electrons pass through the electrode to the vacuum interface with little loss of energy, and thus escape into vacuum, as shown by the line abc in Fig. 40b.

In principle, then, with suitable geometry and voltage bias, a large fraction of the tunneling electrons should be able to escape from the tunnel junction into vacuum

and be collected by a suitably biased anode. The tunnel junction then, in principle, constitutes a cold cathode. It transpires, however, that such cathodes are extremely inefficient, having *transfer ratios* (ratio of emission current to circulating current) typically of the order 10^{-4} or 10^{-3}.[98–102] These disappointingly low values have apparently been established to result primarily because the majority of the electrons undergo energy losses while traveling in the conduction band of the insulator, as shown by the stepped line *bd* in Fig. 40*b*, although losses in the metal film are not negligible. The energy losses in the insulator are such that the electron enters the positively biased electrode with energy less than that of the work function of the electrode and are thus unable to escape. Collins and Davies[104] and Handy[105] estimate the electron mean path in Al_2O_3 to be about 4 or 5 Å.

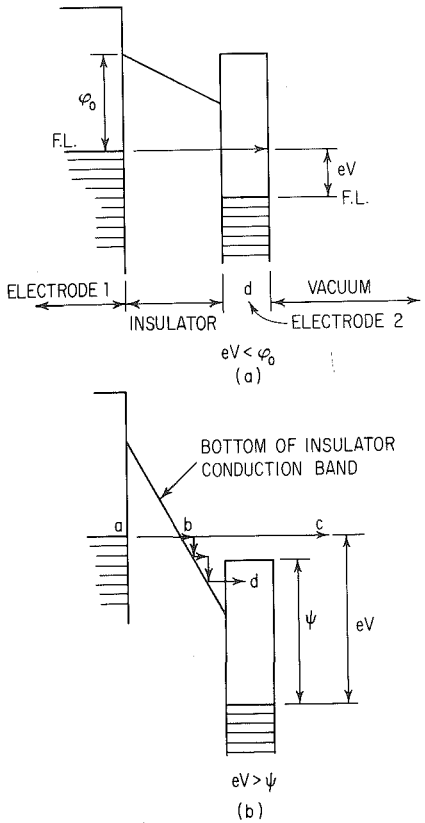

Fig. 40 Energy diagram of cold cathode for (*a*) $eV < 4$, (*b*) $eV > \psi$.

(2) Attenuation of Hot Electrons in Metal Films The use of a cold-cathode emitter to determine the attenuation length of hot electrons in metal films, the metal film under investigation being the exit electrode of the cathode per se, was first used by Mead.[99] In essence the technique

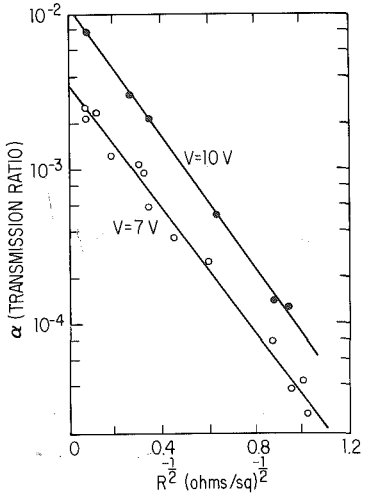

Fig. 41 Emission current vs. electrode sheet resistance for two voltage biases: $V = 7$ V, $V = 10$ V.

consists of measuring the relative number of electrons transmitted through the metal film as a function of the film thickness, all other parameters remaining constant, assuming the following relationship holds:

$$n(d,V) = n(0,V) \exp\left(-\frac{d}{\lambda}\right) \tag{58}$$

where $n(d,V)$ is the number of electrons per second escaping through a film of thickness d at an applied voltage V, and λ is the mean free path of the electrons. A more convenient way of expressing (58) is in terms of the transfer ratio α:

$$\alpha = \frac{n(d,V)}{n(0,V)} = \exp\left(-\frac{d}{\lambda}\right) \tag{59}$$

Mead's[97] results for a series of Be-BeO-Au junctions, in which the BeO thickness was held constant and the Au film thickness varied, are shown in Fig. 41 for two voltage biases: $V = 7$ and 10 V. In this figure α is shown plotted against the reciprocal of the square root of the Au sheet resistance ($R^{-\frac{1}{2}}$), which is assumed proportional to the film thickness. (Mead estimates that $R^{-\frac{1}{2}} = 0.95$ corresponds to a Au film thickness of 400 Å.) From the slopes of the two curves, λ_{Au} is seen to be 190 and 180 Å, respectively, for $V = 7$ and 10 V; λ_{Au} is therefore a function of the electron energy, as is to be expected.

It will be noted that the transfer ratios extrapolated to zero thickness ($R^{-\frac{1}{2}} = 0$) are considerably less than unity for both voltage biases, and this is attributed to the electron's undergoing losses in the insulator. This means, of course, that only a fraction of the electrons are reaching the electrode with the energies (7 and 10 eV) that are assumed to correspond to the two plots in Fig. 41.

b. Tunnel-emission Triode

If a second insulator (assumed to be less than the electronic mean free path) and a third electrode are deposited onto the cold cathode described above, we have the basic structure of the thin film, tunnel-emission triode conceived by Mead.[97] The principle of operation of this device can be seen with the aid of the energy diagram of Fig. 42. The energy of electrons tunneling between the emitter and base elec-

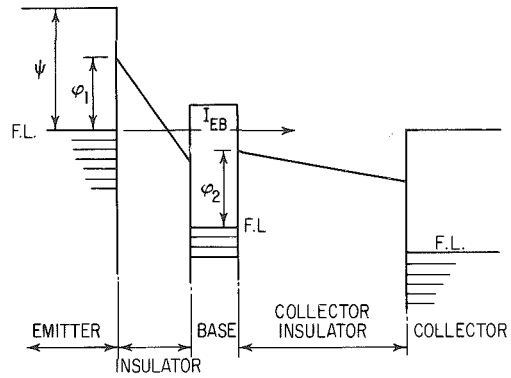

Fig. 42 Energy diagram illustrating proposed operation of tunnel triode.

trodes (the electrodes corresponding to those of the cold cathode described above) is assumed to be conserved; they thus reach the interface existing between the base and the second insulator with energy eV_{EB}, where V_{EB} is the emitter-base bias (base positively biased). For $0 < eV < \phi_2/e$, the electrons do not have sufficient energy to enter the conduction band of the second insulator. If $V > \phi_2/e$ the electrons can enter the conduction band of the second insulator, in which they are then accelerated toward, and collected by, the collector, which is positively biased with respect to the base during operation. Thus the second insulator and collector serve the same function as the vacuum interspace and anode in the cold-cathode emitter.

Unfortunately, the triode suffers from all the problems of the cold cathode, plus additional problems arising from scattering and trapping in the collector insulator, which are not present in the vacuum interspace between the tunnel junction and anode comprising the cold cathode. Thus such a device does not appear viable in light of current state-of-the-art techniques. Nevertheless the principle involved is clearly quite ingenious, and the subsequent research stimulated as a result of the proposal has certainly increased our understanding of the transport properties of thin films.

c. Metal-base Transistor

The operation of the metal-base transistor proposed by Geppert[106] is similar in construction to that of the tunnel transistor, except that the insulators are replaced by

semiconducting films. This device suffers from the same kind of problems that bedevil the tunnel triode, and the feasibility of a commercially viable device does not appear very auspicious at present. For further details the reader is referred to the literature.[106]

d. Cold Cathodes with Thick Insulating Films

The tunnel devices described above rely for their operation on direct electrode-electrode tunneling and thus necessitate the use of an extremely thin (< 50 Å) layer if significant tunnel currents are to be realized. The use of such a thin oxide is undesirable because even a few volts applied to the structures induces intense electric fields within the oxide during operation—5 V across a 50-Å oxide results in a field of 10^7 V cm^{-1}. Fields of this order are perilously close to dielectric-breakdown fields and must therefore severely limit the life of such cathodes.

Cold cathodes having much thicker insulating films, which thus do not suffer from the above restrictions, have recently been reported;[83,88,90,91,107–110] the transfer ratios obtained from these cathodes are usually a good deal higher than those for the tunnel cathodes described above.[108–110] This cathode is identical in physical structure to the negative-resistance and memory device described in Sec. 7; that is, a typical device would be Al-SiO-Au structure which has undergone forming. It is usual, however, to use somewhat thicker insulating films (>1,000 Å) when using the structure as a cold-cathode emitter, if a high α is required.[108–110]

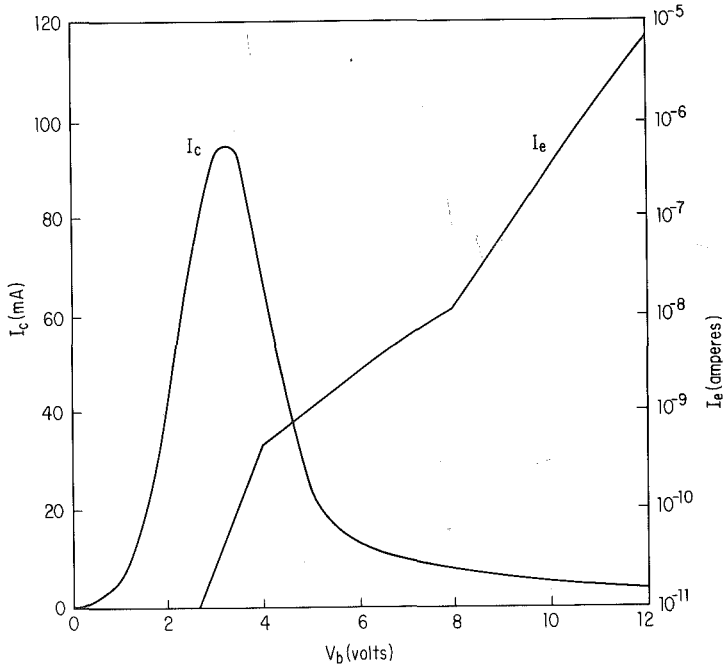

Fig. 43 Emission current I and circulating current I_c vs. applied voltage for a Al-SiO-Au sample.

In Fig. 43 we have plotted the circulating current I_c and the emission current I_E (which was collected by a metal anode biased a few volts positive with respect to the exit electrode) as a function of the applied cathode voltage.[108–110]

The emission process is shown to be due to hot-electron emission rather than thermally stimulated on two counts:[107–110] (1) the emission current in Fig. 43 is observed to increase in the range 3 to 7 V, although the power dissipated in the device is actually decreasing; and (2) on reversing the polarity of bias across the cathode no emission is observed, even though the same power is generated in the device independent of polarity of bias.

Examination of the cathodes always revealed pinholes in the upper electrodes,[107–110] and by forming a magnified image of the emitting surface of the cathode at a phosphor screen by means of an electrostatic lens, it was possible to show that the pinhole edges were the source of the emitted electrons.[107–110]

(1) Distribution of Emitted Electrons The spatial distribution of the emitted electrons was observed at a zinc sulfide phosphor screen. The phosphor-coated anode was positioned parallel to the plane of the cathode and biased at 1 to 2 kV positive with respect to the cathode, in order that the emitted electrons would acquire sufficient energy to excite the phosphor.

With the anode positioned a few centimeters from the cathode and the device biased less than 8 V, the emission pattern consists of several small spots within the 0.1 cm^2 area of the device, which indicate that the electrons are being emitted from small localized regions. If the bias voltage is increased above 7 V, an image much larger than the cathode is observed. The particular interesting aspect of the image is the amount of structure exhibited in it, as shown in Fig. 44. Imposed on a faintly illuminated background are many small brightly illuminated arcs, all of the same radii, none of which subtends an arc greater than 90°.

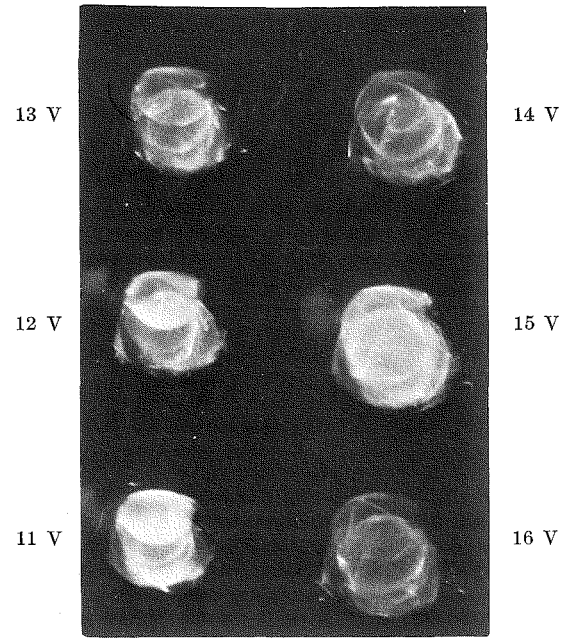

Fig. 44 Full-sized photograph of arcs for various voltage bias. Cathode-screen distance = 4.42 cm; cathode-screen potential 2.0 kV; cathode area $\frac{1}{8}$ by $\frac{1}{8}$ in. (note shape and size of image much different from that of cathode). Exposure times given alongside photographs showing qualitatively that the emission is increasing rapidly with V_b.

All the devices displayed similar images on the screen. It has been shown that these patterns are caused by coherent scattering of hot electrons injected into the gold electrode, followed by their subsequent escape from exposed *edges* of the electrode, such as the circumference of pinholes. Briefly, what this means is that the electrons entering the upper electrode undergo very low energy electron diffraction and are scattered into a hollow cone of diffraction of half angle 2θ (see Fig. 45). The diffracting angle (2θ) is related to the bias voltage V and the material constants

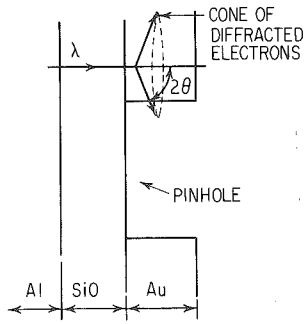

Fig. 45 Schematic diagram of thin film cold cathode, in the vicinity of a pinhole, showing the path of a diffracted electron.

by the Bragg equation

$$\sin\theta = \frac{\lambda}{2d_{hkl}} \tag{60}$$

where λ is the wavelength of the electron in the top electrode expressed by[107–109]

$$\lambda = \frac{12.27}{(V+\eta)^{\frac{1}{2}}} \quad \text{angstroms} \tag{61}$$

d_{hkl} is the distance between the diffracting crystal planes, and η is the Fermi energy of the top metal electrode. For the voltage biases shown in Fig. 43, it can be shown that the component of electron momentum parallel to the plane of the film is much greater than the perpendicular component, which means that the electron can escape only at an exposed edge of electrode.

(2) Scattering Processes[109] The structure in the emission I-V characteristic of Fig. 43 is due to electrons undergoing multiple scattering below 8 V and only a single (coherent) scatter above 8 V. Using $V < 8$ V, $\eta = 5.5$ eV, and d_{111} (the most intense diffracting plane of Au) = 2.35 Å in (61) we find $\sin\theta = 0.71$ and hence $2\theta \simeq 90°$. Thus for $V < 8$ eV the electrons are scattered back toward the insulator and would not normally escape from the electrode unless they underwent further scattering (electron-electron, electron-phonon) to reach an electrode edge. These additional scattering processes lead to substantial energy losses, and the process is obviously an inefficient means of directing electrons toward an edge. For $V > 8$ eV the scattering process is coherent only, and the energy is conserved in this process. In this case, if the scattered electron is within a distance not greater than the electron mean free path from the edge, it will reach the edge without loss of energy and hence will escape into vacuum.

The transmission ratios observed in the cathode are typically 10 to 100 times greater than for tunnel cathodes, yet the insulator and electrode materials are typically ten times thicker. On this basis it is difficult to attribute the low emission currents in tunnel cathodes to small-insulator and electron mean free paths alone. For further details on the transport of the electrons through the exit electrode and insulator the reader is referred to the literature.[109]

(3) Alphanumeric Display[110] A thin film, cold-cathode, alphanumeric display panel, consisting of a 5 × 5 matrix utilizing the cold cathodes described above, has been shown to be feasible. With this simple system the alphabet and all numerals can be displayed quite legibly, as shown in Fig. 46. These characters are clearly visible under normal room-lighting conditions. For fabrication details, response time, etc., the reader is referred to Ref. 110.

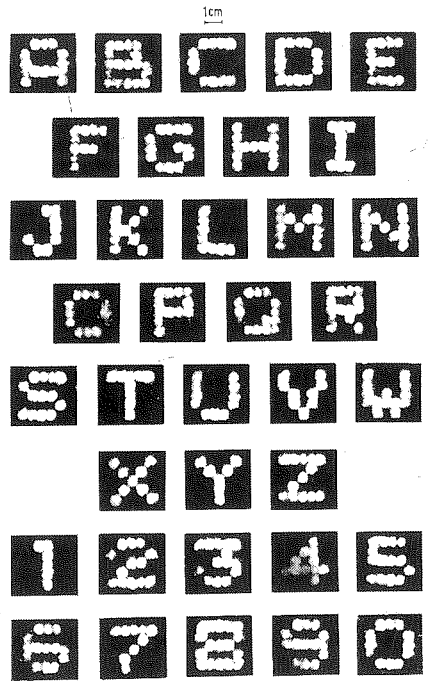

Fig. 46 Characters obtained from an alphanumeric display panel consisting of a 5 × 5 matrix of thin film cold cathodes.

REFERENCES

1. Mott, N. F., and R. W. Gurney, "Electronic Processes in Ionic Crystals," 2d ed., chap. V, Oxford University Press, Fair Lawn, N.J., 1948.
2. Rose, A., *Phys. Rev.*, **97,** 1538 (1955).
3. Lampert, M. A., *Rept. Progr. Phys.*, **27,** 329 (1964).
4. Simmons, J. G., and G. S. Nadkarni, *J. Vac. Sci. & Tech.*, **6,** (12) 1969.
5. Bujatti, M., and R. S. Muller, *J. Electrochem. Soc.*, **112,** 702 (1965).
6. Hass, G., *J. Am. Ceram. Soc.*, **33,** 353 (1958).
7. Ritter, E., *Opt. Acta*, **9,** 197 (1962).
8. Brady, G. W., *J. Phys. Chem.*, **63,** 1119 (1959).
9. Faessler, Von A., and H. Kramer, *Ann. Physik*, **7,** 263 (1959).
10. Simmons, J. G., *Phys. Rev.*, **155,** 657 (1967).
11. Dresner, J., and F. V. Shallcross, *Solid-State Electron.*, **5,** 205 (1962).
12. Schottky, W., *Z. Physik*, **113,** 367 (1939).
13. Henisch, H. K., "Rectifying Semiconductor Contacts," Oxford University Press, Fair Lawn, N.J., 1957.
14. Simmons, J. G., *Phys. Rev. Letters*, **10,** 10 (1963).
15. Bardeen, J., *Phys. Rev.*, **71,** 717 (1947).
16. Frenkel, J., *Phys. Rev.*, **36,** 1604 (1930).
17. Sommerfeld, A., and H. Bethe, in VonGeiger and Scheel (eds.), "Handbuch der Physik," vol. 24/2, p. 450, Springer-Verlag OHG, Berlin, 1933.
18. Holm, R., *J. Appl. Phys.*, **22,** 569 (1951).
19. Stratton, R., *J. Phys. Chem. Solids*, **23,** 1177 (1962).
20. Simmons, J. G., *J. Appl. Phys.*, **34,** 1793 (1963).
21. Simmons, J. G., *J. Appl. Phys.*, **34,** 2581 (1963).
22. Simmons, J. G., *J. Appl. Phys.*, **35,** 2655 (1964).
23. Geppert, D. V., *J. Appl. Phys.*, **34,** 490 (1963).
24. Geppert, D. V., *J. Appl. Phys.*, **33,** 2993 (1962).
25. Pittelli, E., *Solid-State Electron.*, **6,** 667 (1963).
26. Penley, J. C., *Phys. Rev.*, **128,** 596 (1962).
27. Schmidlin, F. W., *J. Appl. Phys.*, **37,** 2823 (1966).
28. Stratton, R., G. Lewiki, and C. A. Mead, *J. Phys. Chem. Solids*, **27,** 1599 (1966).
29. Grundlock, K. H., and G. Heldmann, *Phys. Solid State*, **21,** 575 (1967).
30. Schnup, P., *Solid-State Electron.*, **10,** 785 (1967).
31. Schnup, P., *Phys. Stat. Solidi*, **21,** 567 (1967).
32. Simmons, J. G., *Brit. J. Appl. Phys.*, **18,** 269 (1967).
33. Stratton, R., *Phys. Rev.*, **90,** 515 (1963).
34. Fowler, R. H., and L. W. Nordheim, *Proc. Roy. Soc. London*, **A119,** 173 (1928).
35. Fisher, J., and I. Giaever, *J. Appl. Phys.*, **32,** 172 (1961).
36. Handy, R. M., *Phys. Rev.*, **126,** 1968 (1962).
37. Penley, J., *J. Appl. Phys.*, **33,** 1901 (1962).
38. Simmons, J. G., and G. J. Unterkofler, *J. Appl. Phys.*, **34,** 1838 (1963).
39. Simmons, J. G., G. J. Unterkofler, and W. W. Allen, *Appl. Phys. Letters*, **2,** 78 (1961).
40. Meyerofer, D., and S. A. Ochs, *J. Appl. Phys.*, **34,** 2535 (1963).
41. Hurst, H. G., and W. Ruppel, *Z. Naturforsch.*, **19a,** 573 (1964).
42. Hartman, T. E., and J. S. Chevian, *Phys. Rev.*, **134,** A1094 (1964).
43. Hartman, T. E., *J. Appl. Phys.*, **35,** 3283 (1964).
44. Pistoulet, B., M. Rouzeyou, and B. Bouat, *L'Onde Elec.*, **44,** 1297 (1964).
45. Nakai, J., and T. Miyazaki, *Japan. J. Appl. Phys.*, **3,** 677 (1964).
46. Pollack, S. R., and C. E. Morris, *J. Appl. Phys.*, **35,** 1503 (1964).
47. Pollack, S. R., and C. E. Morris, *Trans. AIME*, **233,** 497 (1965).
48. McColl, M., and C. A. Mead, *Trans. AIME*, **233,** 502 (1965).
49. Miles, J. L., and P. H. Smith, *J. Electrochem. Soc.*, **110,** 1240 (1963).
50. Braunstein, A., M. Braunstein, G. S. Picus, and C. A. Mead, *Phys. Rev. Letters*, **14,** 219 (1965).
51. Simmons, J. G., *Phys. Letters*, **17,** 104 (1965).
52. Esaki, L., and P. J. Stiles, *Phys. Rev. Letters*, **16,** 1108 (1966).
53. Chang, L. L., P. J. Stiles, and L. Esaki, *J. Appl. Phys.*, **38,** 4440 (1967).
54. Dahlke, W. E., *Appl. Phys. Letters*, **10,** 261 (1967).
55. Lewicki, G., and C. A. Mead, *Phys. Rev. Letters*, **16,** 939 (1966).

55a. Franz, W., in S. Flügge (ed.), "Handbuch der Physik," vol. 17, p. 155, Springer-Verlag OHG, Berlin, 1956.

56. Esaki, L., P. J. Stiles, and S. von Molnar, *Phys. Rev. Letters*, **19**, 852 (1967).
57. Schottky, W., *Physik. Z.*, **15**, 872 (1914).
58. Simmons, J. G., *Phys. Rev. Letters*, **15**, 967 (1965).
59. Crowell, C. R., *Solid-State Electron.*, **8**, 395 (1965).
60. Tantraporn, W., *Solid-State Electron.*, **7**, 81 (1964).
61. Simmons, J. G., *J. Appl. Phys.*, **35**, 2472 (1964).
62. Emptage, P. R., and W. Tantraporn, *Phys. Rev. Letters*, **8**, 267 (1962).
63. Mead, C. A., in R. Niedermayer and H. Mayer (eds.), "Basic Problems in Thin Film Physics," p. 674, Vandenhoeck and Ruprecht, Goettingen, 1966.
64. Pollack, S. R., *J. Appl. Phys.*, **34**, 877 (1963).
65. Flannery, W. E., and S. R. Pollack, *J. Appl. Phys.*, **37**, 4417 (1967).
66. Simmons, J. G., *Phys. Rev.*, **166**, 912 (1968).
67. Frenkel, J., *Tech. Phys.*, **5**, 685 (1938).
68. Frenkel, J., *Phys. Rev.*, **54**, 647 (1938).
69. Mead, C. A., *Phys. Rev.*, **128**, 2088 (1962).
70. Hirose, H., and Y. Wada, *Japan. J. Appl. Phys.*, **3**, 179 (1964).
71. Johansen, I. T., *J. Appl. Phys.*, **37**, 449 (1966).
72. Hartman, T. E., J. C. Blair, and R. Bayer, *J. Appl. Phys.*, **37**, 2468 (1966).
73. Simmons, J. G., *Phys. Rev.*, **155**, 657 (1967).
74. Stuart, M., *Brit. J. Appl. Phys.*, **18**, 1637 (1967).
75. Stuart, M., *Phys. Stat. Solidi*, **23**, 595 (1967).
76. Frank, R., and J. G. Simmons, *J. Appl. Phys.*, **38**, 832 (1967).
77. Lampert, M. A., *Phys. Rev.*, **103**, 1648 (1956).
78. Lampert, M. A., A. Rose, and R. W. Smith, *J. Appl. Chem. Solids*, **8**, 484 (1959).
79. Lampert, M. A., and A. Rose, *Phys. Rev.*, **121**, 26 (1961).
80. Lampert, M. A., *Phys. Rev.*, **125**, 126 (1962).
81. Zuleeg, R., *Solid-State Electron.*, **6**, 645 (1963).
82. Zuleeg, R., and R. S. Muller, *Solid-State Electron.*, **7**, 575 (1964).
83. Kreynina, G. S., L. N. Selivanov, and T. I. Shumskaia, *Radio Eng. Elec. Phys.*, **5**, 8 (1960).
84. Kreynina, G. S., L. N. Selivanov, and T. I. Shumskaia, *Radio Eng. Elec. Phys.*, **5**, 219 (1960).
85. Kreynina, G. S., *Radio Eng. Elec. Phys.*, **7**, 166 (1962).
86. Kreynina, G. S., *Radio Eng. Elec. Phys.*, **7**, 1949 (1962).
87. Hickmott, T. W., *J. Appl. Phys.*, **33**, 2669 (1962).
88. Hickmott, T. W., *J. Appl. Phys.*, **34**, 1569 (1963).
89. Hickmott, T. W., *J. Appl. Phys.*, **35**, 2118 (1964).
90. Cola, R. A., J. G. Simmons, and R. R. Verderber, *Proc. Natl. Aerospace Electron Conf.*, 1964, p. 118.
91. Hickmott, T. W., *J. Appl. Phys.*, **36**, 1885 (1965).
92. Lewowski, T., S. Sendecki, and B. Sujak, *Acta Phys. Polon.*, vol. 28, 1965.
93. Uzan, R., A. Roger, and A. Cachard, *Vide*, **137**, 38 (1967).
94. Simmons, J. G., and R. R. Verderber, *Proc. Roy. Soc. London*, **301**, 77 (1967).
95. Simmons, J. G., and R. R. Verderber, *Radio Elec. Eng.*, **34**, 81 (1966).
96. Verderber, R. R., J. G. Simmons, and B. Eales, *Phil. Mag.*, **16**, 1049 (1967).
97. Mead, C. A., *Proc. IRE*, **48**, 359 (1960).
98. Mead, C. A., *J. Appl. Phys.*, **32**, 646 (1961).
99. Mead, C. A., *Phys. Res. Letters*, **8**, 56 (1962).
100. Kanter, H., and W. A. Feibelman, *J. Appl. Phys.*, **33**, 3580 (1962).
101. Cohen, J., *J. Appl. Phys.*, **33**, 1999 (1962).
102. Cohen, J., *Appl. Phys. Letters*, **1**, 61 (1962).
103. Collins, R. E., and L. W. Davies, *Appl. Phys. Letters*, **2**, 213 (1963).
104. Collins, R. E., and L. W. Davies, *Solid-Stat. Electron.*, **7**, 445 (1964).
105. Handy, R. M., *J. Appl. Phys.*, **37**, 4620 (1967).
106. Geppert, D. V., *Proc. IRE*, **50**, 1527 (1962).
107. Simmons, J. G., R. R. Verderber, J. Lytollis, and R. Lomax, *Phys. Rev. Letters*, **17**, 675 (1965).
108. Simmons, J. G., and R. R. Verderber, *Appl. Phys. Letters*, **10**, 1967.
109. Verderber, R. R., and J. G. Simmons, *Radio Elec. Eng.*, **33**, 347 (1967).
110. Lomax, R., and J. G. Simmons, *Radio Elec. Eng.*, **35**, 265 (1968).

Other Recent Reviews on Conduction in Thin Films

Lamb, D. R., "Electrical Conduction in Thin Insulating Films," Methuen & Co., Ltd., London, 1967.
Ekertova, L., *Phys. Stat. Solidi*, **18**, 3 (1966).
Hill, R. A., *Thin Solid Films*, **1**, 39 (1967).
Jonscher, A. K., *Thin Solid Films*, **1**, 213 (1967).

Chapter 15

Piezoelectric and Piezoresistive Properties of Films

NORMAN F. FOSTER

Bell Telephone Laboratories, Allentown, Pennsylvania