REVIEW ARTICLE

The physics of Schottky barriers

The physics of Schottky barriers†

E. H. RHODERICK

Department of Electrical Engineering and Electronics,
University of Manchester Institute of Science and Technology

MS. received 21*st April* 1970

Abstract. A review is given of the physical processes which determine the height of the barrier and the current–voltage relationship in a metal–semiconductor Schottky barrier.

1. Historical

The metal–semiconductor rectifier must be the oldest solid state device used in electronics, because the first paper published on the subject appeared in 1874, when Braun reported the asymmetrical nature of conduction between metal points and crystals such as lead sulphide. Its application as a radio-frequency detector is almost as old as wireless telegraphy itself, for it is now fairly certain that Branly's 'coherer' must have relied for its action on the rectifying properties of metal contacts separated by an oxide film. In 1906 Pickard took out a patent for a point-contact rectifier using silicon, and in 1907 Pierce published rectification characteristics of diodes made by sputtering metals on to a variety of semiconductors. The point-contact rectifier or 'catswhisker' was used extensively in the early days of radio, but the first real scientific study of the device (and indeed the beginning of semiconductor physics) was stimulated by the wartime use of silicon and germanium point-contact rectifiers as microwave detectors. Point-contact rectifiers were very variable and unreliable in their characteristics, and our present understanding of Schottky barriers has come with the realization that metal films evaporated onto single-crystal semiconductor surfaces under conditions of high cleanliness can show almost ideal rectification characteristics. An extensive account of the early history of metal–semiconductor contacts can be found in Henisch's book (Henisch 1957), and an up-to-date review of the properties and applications of Schottky barriers has been given by Atalla (1967).

2. Formation of the barrier

2.1. *Simple Schottky theory*

The rectifying properties of a Schottky barrier arise from the presence of an electrostatic barrier between metal and semiconductor. To see how this arises, consider a piece of metal and a piece of semiconductor, both electrically neutral and separated from each other. The energy band diagram is as shown in figure 1(*a*), the zero of energy being the energy of an electron at rest outside either solid. For definiteness we show the case of an n-type semiconductor whose work function is less than that of the metal, which is in practice the most important case. (The argument can be inverted for the case of p-type semiconductors). If the metal and semiconductor are connected electrically by means of a wire, electrons pass from the semiconductor to the metal and the Fermi levels are forced into coincidence. The energy of an electron at rest outside the surfaces of the two solids is no longer the same, and there is an electric field in the gap (as shown in figure 1(*b*) the field is from right to left). There must be a negative charge on the surface of the metal, balanced by a positive charge on the semiconductor. Since the semiconductor is n-type, this positive charge is provided by conduction electrons receding from the surface, leaving uncompen-

† Based on an invited paper given at the third Solid-State Devices Conference, Exeter, 1969.

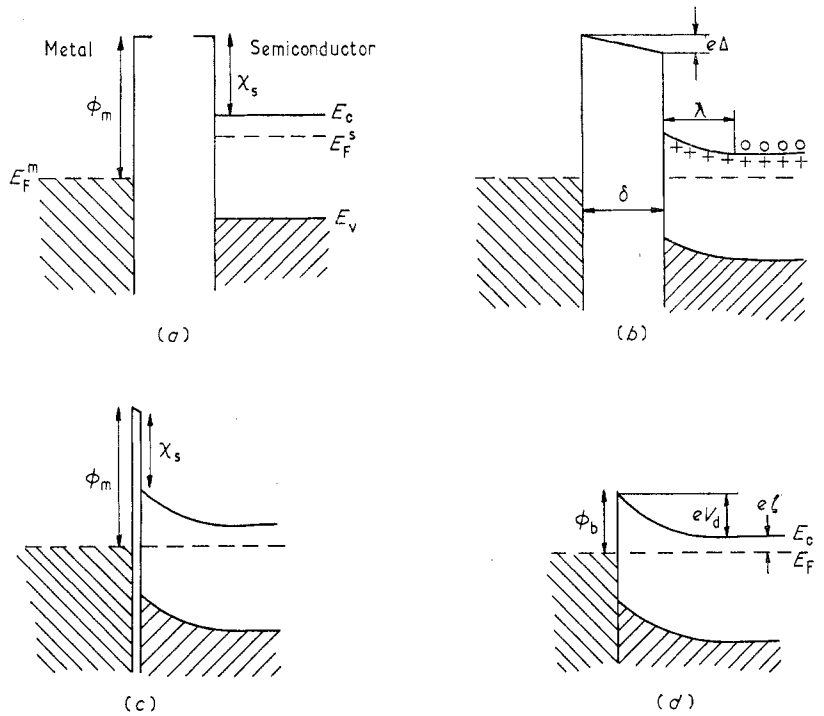

Figure 1. Formation of Schottky barrier from metal and semiconductor. (*a*) neutral and isolated, (*b*) electrically connected, (*c*) separated by narrow gap, (*d*) perfect contact.
○ denotes electron in conduction band; + denotes positive donor ion.

sated donor ions. Because the concentration of these donors is many orders of magnitude less than the concentration of electrons in the metal, the balancing positive charges occupy a layer of appreciable thickness, and the bands in the semiconductor are bent upwards as shown. The work-function of the metal (ϕ_m) and the electron affinity of the semiconductor (χ_s) remain unaltered as shown. As the metal and semiconductor approach each other, the drop in electrostatic potential (Δ) associated with the field in the gap tends to zero if the field is to remain finite, as shown in figure 1(*c*). Finally, if the gap becomes truly zero (corresponding to perfectly intimate contact between metal and semiconductor), the thin potential barrier separating them disappears altogether, as in figure 1(*d*), and we are left with a barrier arising from the band-bending. Even if contact is not perfect and a very thin high barrier separating metal and semiconductor still remains as in 1(*c*), this barrier is usually so thin (~ 10 Å) that electrons can easily pass through it by tunnelling, and the situation shown in figure 1(*c*) is essentially equivalent to 1(*d*). Clearly, in the limit of perfect contact, $\phi_b = \phi_m - \chi_s$, a relationship which was first derived by Schottky (1938).

2.2. *Influence of surface states*

It was soon found that in the case of point contact rectifiers the barrier is often independent of the metal, which is at variance with the simple Schottky theory. Bardeen (1947) was the first to point out the importance of surface states in the semiconductor in determining the barrier height. For our present purpose it is sufficiently accurate to regard surface states as unsaturated (or 'dangling') bonds on the surface of the semiconductor. The surface states are usually continuously distributed in energy within the forbidden gap, and are characterized by a 'neutral level' ϕ_0 such that if the surface states are occupied up to ϕ_0 and empty above ϕ_0 the surface is electrically neutral. If we approximate the Fermi–Dirac distribution by a step function, it follows that when the Fermi level E_F lies above ϕ_0

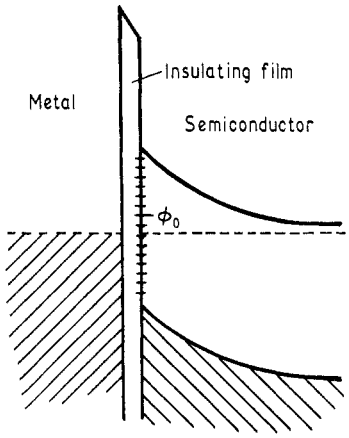

Figure 2. Metal–semiconductor contact with surface states.

the surface states possess a net negative charge, while if E_F lies below ϕ_0 they possess a net positive charge. In other words, the states below ϕ_0 are 'donor-like' (positive when empty) and the states above ϕ_0 are 'acceptor-like' (negative when occupied). In the case shown in figure 2, for which $\phi_0 > E_F$, there is a net positive charge in the surface states, and some of the lines of force terminating on the metal emanate from surface charges rather than from ionized donors; in this case the space charge region is not as thick as when surface states are absent, and the barrier height is reduced. In the same way, if $\phi_0 < E_F$, there is a net negative charge in the surface states and the barrier height is increased. Thus the charge in the surface states has a 'negative feed-back' effect which always tends to reduce the deviation of ϕ_0 from E_F. If the density of surface states is very high, the gain in the feedback loop is very large, and ϕ_0 tends to be locked to E_F so that the barrier height is virtually independent of the metal. Cowley and Sze (1965) have shown that, in the case of an n-type semiconductor, Bardeen's theory can be approximately expressed in the form

$$\phi_b = \gamma(\phi_m - \chi_s) + (1-\gamma)(E_g - \phi_0) \tag{1}$$

where E_g is the energy gap of the semiconductor and the neutral level ϕ_0 is measured from the top of the valence band. The approximations leading to (1) are strictly justified only if the charge in the surface states is very much greater than the charge in the space-charge layer. However, in practice it seems to be a reasonable approximation in all cases. This equation assumes that the metal and semiconductor do not make perfect contact as in figure 1(*d*) but remain separated by a thin oxide film of thickness δ as in figure 1(*c*). γ is given by

$$\gamma = \epsilon_i\epsilon_0/(\epsilon_i\epsilon_0 + e\delta D_s)$$

where ϵ_i is the relative permittivity of the oxide film, ϵ_0 is the permittivity of free space, and D_s is the density of surface states per electron-volt per unit area of semiconductor surface. If there are no surface states, $\gamma = 1$ and (1) reduces to the simple Schottky relationship $\phi_b = \phi_m - \chi_s$. On the other hand, if D_s is very large, γ tends towards zero and ϕ_b approximates to $E_g - \phi_0$, i.e. the barrier becomes stabilized so that the Fermi level coincides with ϕ_0 independently of the value of ϕ_m.

3. Measurements of barrier heights

3.1. *Experimental methods*

There are several methods at our disposal for measuring ϕ_b. The best known relies on the measurement of the differential capacitance C as a function of reverse bias V. It can be shown that C is proportional to $(V+V_d)^{-1/2}$, where V_d is the 'diffusion' voltage (i.e. the band bending in volts), so that a plot of C^{-2} against V gives a straight line with

intercept $-V_d$. The barrier height ϕ_b is equal to $e(V_d+\zeta)$, where $e\zeta$ is the energy of the conduction band relative to the Fermi level (see figure 1). Alternatively we can measure the current–voltage characteristic and, provided it is close to the ideal rectifier relationship $J=J_0\{\exp(eV/kT)-1\}$, we can write $J_0=AT^2\exp(-\phi_b/kT)$, where A is the Richardson constant corrected for the effective mass of the electrons in the semiconductor (Crowell 1965), so that ϕ_b can be obtained from J_0. Finally, we can measure the threshold for optical excitation of electrons from the metal over the barrier into the semiconductor. All these methods are intrinsically capable of good accuracy, but the ultimate error is limited by repeatability from sample to sample and is determined by the technology of specimen preparation.

3.2. *Results for silicon*

Following the pioneering work of Archer and Atalla (1963), the semiconductor we know most about as far as Schottky barriers are concerned is silicon. An extensive study of metal–silicon barriers by Turner and Rhoderick (1968) has shown that

(i) ϕ_b depends on the method of preparation of the surface,

(ii) ϕ_b often ages with time, becoming stable in a matter of days or even weeks (see figure 3),

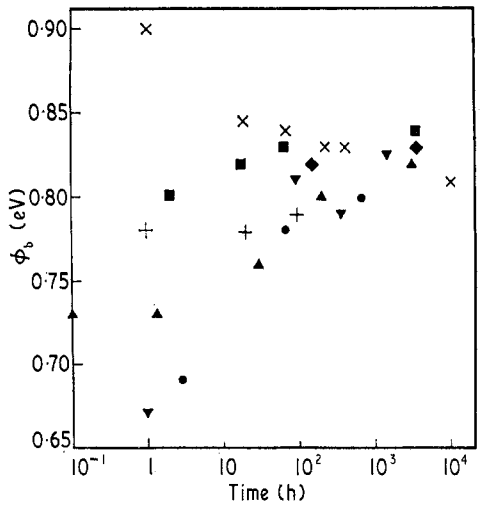

Figure 3. Ageing of gold–silicon Schottky barriers (Turner and Rhoderick 1968). The letters A–E refer to different methods of surface preparation.

Sample	Resisitivity (Ω cm)	Surface	Vacuum (torr)
×	1	A	10^{-5}
+	1	A	10^{-7}
◆	10	B	10^{-5}
●	10	C	10^{-5}
■	10	C	10^{-5}
▲	10	D	10^{-5}
▼	10	E	10^{-7}

(iii) in the case of etched surfaces the barrier depends on the choice of metal, but not to the extent predicted by the Schottky theory, and

(iv) in the case of surfaces cleaved in ultra-high vacuum, the barrier is almost independent of the metal (see figure 4).

It is clear that cleaved surfaces belong to the case where a high density of states locks the barrier so that it is independent of ϕ_m ($\gamma\rightarrow 0$), and that etched surfaces lie somewhere

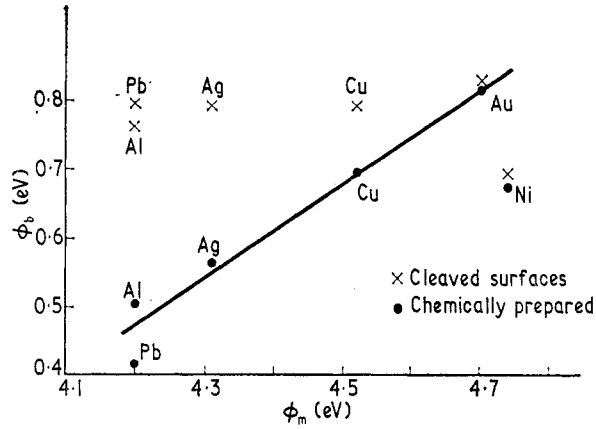

Figure 4. Barrier heights of metal–silicon Schottky barriers (Turner and Rhoderick 1968).

between the two extreme cases with $\gamma \simeq 0{\cdot}67$. The cleaved-surface data are compatible with $D_s \simeq 4 \times 10^{18}$ m^{-2} eV^{-1}, which corresponds to a total density of surface states of about 1 per atom, as we might expect from a dangling bond picture. The etched-surface data are compatible with $D_s \simeq 10^{16}$ m^{-2} eV^{-1}, which is a figure often measured on 'real' surfaces by field effect measurements. One can suppose that the thin oxide layer, which always covers etched surfaces, almost completely saturates the dangling bonds.

Any attempt at a very detailed comparison between theory and experiment is hardly profitable because of the uncertainty in the values of the metal work functions. These are generally taken from published data and usually refer to experimental conditions quite different from those under which the metal film has been evaporated. From time to time the accepted values of work functions may be drastically revised; for example, the work function of gold has apparently increased by 0·5 eV in the last few years! Most of the published analyses of Schottky barrier data should be treated with caution. The ideal procedure would be to make a measurement of the semiconductor work function, then evaporate the metal and measure its work function, then measure the height of the Schottky barrier, all in an ultra-high vacuum system without breaking the vacuum. Such a programme has been recently initiated by Poole (1969) but there is not yet sufficient data for any extensive conclusions to be drawn. Preliminary measurements suggest that it may not prove possible to assign to every metal or semiconductor a unique work function which determines in a consistent manner the height of the barrier formed by that material in conjunction with some other semiconductor or metal.

3.3. *Results for gallium arsenide*

The next most widely studied semiconductor is gallium arsenide, and here the results are different from silicon in two respects (figure 5). Firstly, on cleaved (110) surfaces there is much more variation with choice of metal than in the case of silicon, suggesting that the surface-state density is lower by about an order of magnitude ($D_s \simeq 7 \times 10^{17}$ m^{-2} eV^{-1}), and secondly, while there is more dependence on ϕ_m in the case of etched surfaces than in the case of cleaved surfaces, the difference is not nearly as marked as in the case of silicon. In fact, the data on etched surfaces are compatible with a surface state density $\simeq 3 \times 10^{17}$ m^{-2} eV^{-1}, which is only a factor of two lower. The first result is compatible with the theoretical calculations of Jones (1968), and also with the observation by Van Laar and Scheer (1967) that the Fermi level at the free surface of cleaved gallium arsenide surfaces is *not* clamped by surface states as in the case of Si. The second may be due to the fact that gallium arsenide does not oxidize as easily as silicon.

There is also a dependence of barrier height on surface orientation. In the case of gallium arsenide, the $(\bar{1}\bar{1}\bar{1})$ surface gives a value of ϕ_b about 0·1 eV higher than the (100)

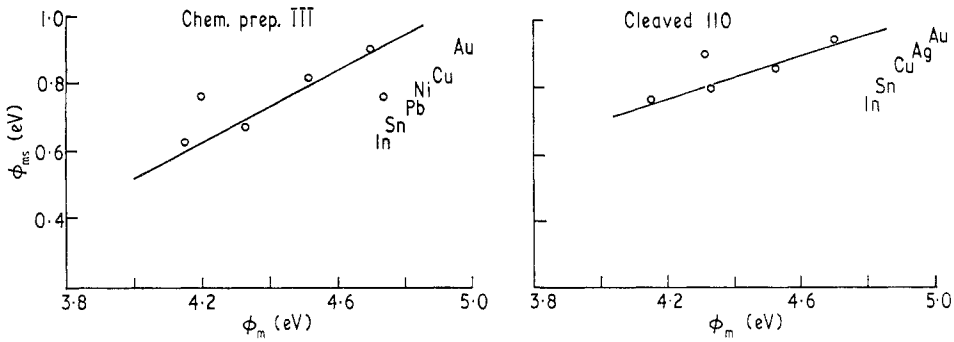

Figure 5. Barrier heights of metal–gallium arsenide Schottky barriers (Smith 1969).

face, and the (111) face is intermediate between the two. The difference between the (111) and $(\bar{1}\bar{1}\bar{1})$ faces is to be expected in polar compounds.

It turns out that, for both silicon and gallium arsenide, $\phi_0 \simeq \frac{1}{3}E_g$. It was suggested by Mead and Spitzer (1964) that this may be true for all semiconductors with the diamond or zincblende structure. Such data as are available suggest that the rule is obeyed quite well—certainly by germanium, silicon, and the III–V compounds.

3.4. *Other semiconductors*

Measurements of varying degrees of completeness on a large number of semiconductors have been summarized by Mead (1966) and Sze (1969). Mead suggests that semiconductors can be divided into two categories—those in which the barrier height is controlled by surface states, and those in which surface states are unimportant. It is clear from the foregoing that this is too simple a view, and that some semiconductors (e.g. silicon) may be in either category according to the method of surface preparation. There are some cases, however, in which it is probably legitimate to characterize semiconductors in this way. Mead reports that in the case of cadmium sulphide surface states are unimportant, while for cadmium selenide the barrier height is effectively locked by surface states. In the case of Cd(S, Se) mixed crystals there is apparently a smooth variation from one type of behaviour to the other as the composition is varied from CdS to CdSe.

3.5. *p-type semiconductors*

The foregoing results refer exclusively to n-type semiconductors. There is relatively little data on p-type semiconductors, partly because the barriers are so low that measurement is difficult, and partly because there is less technical interest in Schottky diodes made on p-type than on n-type semiconductors as a result of the lower barrier height and lower carrier mobility. The most extensive measurements on p-type semiconductors are those of Smith and Rhoderick (1970) on p-type silicon. Smith has shown that, for barriers on etched p-type silicon, the barrier heights can be explained in terms of the same surface parameters (density of surface-states and position of the neutral level) as in the case of Turner's results on n-type silicon.

It might be expected that the surface properties should not depend on the bulk impurity. For the case of p-type semiconductors, the analogue of equation (1) is

$$\phi_b{}^p = \gamma(\chi_s - \phi_m + E_g) + (1-\gamma)\phi_0. \tag{2}$$

From (1) and (2) we find $\phi_b{}^n + \phi_b{}^p = E_g$, i.e. for a given metal the sum of the barriers on n-type and p-type specimens of the same semiconductor should equal the band-gap, provided the surface-state parameters γ and ϕ_0 are the same in both cases. Table 1 shows that the data for silicon satisfy this relationship quite well.

Table 1. Experimental barrier height data on etched p-type and n-type silicon. (n-type from Turner and Rhoderick (1968), p-type data from Smith and Rhoderick (1970).)

Metal	ϕ_m (eV)	ϕ_b^p (eV)	ϕ_b^n (eV)	$\phi_b^p + \phi_b^n$
Al	4·20	0·58	0·50	1·08
Au	4·70	0·35	0·81	1·16
Cu	4·52	0·46	0·69	1·15
Ni	4·74	0·51	0·67	1·18
Pb	4·20	0·55	0·41	0·96

4. I–V characteristics

There are three principal conduction mechanisms in a Schottky junction. For the case of an n-type semiconductor these are:

(i) transport of electrons from the conduction band of the semiconductor over the top of the barrier into the metal,

(ii) injection of holes into the neutral region of the semiconductor, and

(iii) recombination (or generation) of electrons and holes in the depletion region of the semiconductor.

These processes are shown schematically in figure 6. Since (ii) involves hole-electron recombination, it differs from (iii) only in that the recombination takes place in two different regions. There is also in some circumstances a fourth process, namely tunnelling of elec-

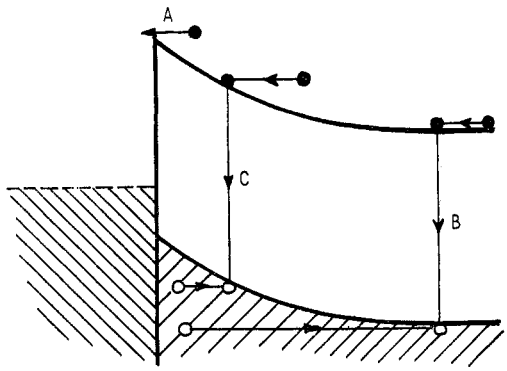

Figure 6. Schottky barrier, showing conduction mechanisms under forward bias. A, emission over barrier; B, recombination in neutral region (hole injection); C, recombination in space-charge region.

trons from the conduction band into the metal. This was at first thought to be the most important transport mechanism until it was realized that it predicted the wrong direction of rectification. It is now known to play a very minor role in the majority of practically important cases.

It is possible to make metal–semiconductor junctions in which processes (ii) and (iii) are insignificant in comparison with (i). Such junctions are said to be 'ideal', and we may think of (ii) and (iii) as causing small departures from ideality.

4.1. *Transport of electrons over the barrier*

Two apparently quite distinct theories have been proposed to describe process 4(i). Historically, the first of these was the so-called diffusion theory of Wagner (1931) and Schottky and Spenke (1939). According to this theory, the current is supposed to be limited by the ordinary processes of diffusion and drift operating within the space-charge region where the bands are bent. But according to the thermionic emission theory, proposed by Bethe (1942), the current is assumed to be limited by the emission of electrons over

the barrier into the metal in a manner similar to the thermionic emission of electrons from a metal into a vacuum. At first sight, the condition for the validity of the thermionic emission theory would seem to be that the mean free path of the electrons should be greater than the thickness of the barrier region, but a more refined argument given by Bethe suggests that all that is necessary is that the mean free path should be greater than the distance in which the barrier falls by an amount kT ftom its maximum value.

The difference between the two theories can be more clearly understood by considering the behaviour of the quasi-Fermi level for electrons. The quasi-Fermi level is a hypothetical energy level which has the significance that, if inserted into the Fermi–Dirac distribution function, it gives the correct concentration of electrons even though the system is not in true thermal equilibrium. Far from the junction the quasi-Fermi level must coincide with the Fermi levels in metal and semiconductor respectively. The assumption tacitly made in the diffusion theory is that the quasi-Fermi level coincides with the Fermi level in the metal *at the junction* (figure 7). This is just another way of saying that the conduction electrons

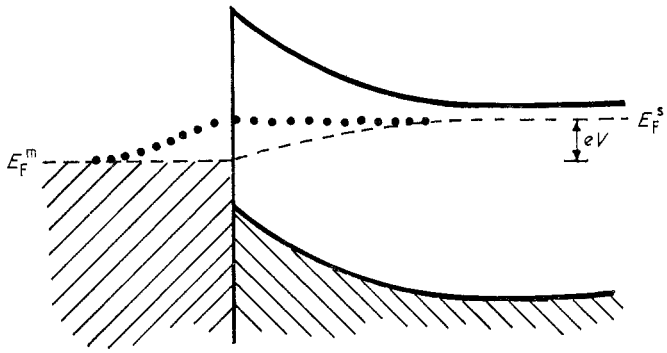

Figure 7. Schottky barrier with forward bias voltage V, showing behaviour of quasi-Fermi level – – – – diffusion theory; • • • • thermionic emission theory.

in the semiconductor immediately adjacent to the metal are in equilibrium with those in the metal, i.e. that the concentration of electrons in the semiconductor at the top of the barrier does not change when a bias is applied. Since the gradient of the quasi-Fermi level is the driving force for electrons, this theory implies that the bottleneck for current flow is provided by the transport of electrons in the space charge region.

The assumption that the quasi-Fermi level is coincident with the Fermi level in the metal at the junction is in sharp contrast with the assumption normally made in p–n junction theory, where it is assumed that the quasi-Fermi level for each type of carrier remains flat through the junction. This assumption has been justified at length by Shockley (1949). Since the transport mechanism in a Schottky barrier is basically very similar to that in a p–n junction, there seems to be a contradiction here. The contradiction was resolved by Gossick (1963), who pointed out that the electrons which cross the barrier from the semiconductor into the metal are not in thermal equilibrium with the conduction electrons in the metal because they have energies exceeding the Fermi-energy by about 1 eV (see figure 8). In a loose sense, they are 'hot' electrons, and in a crude sort of way we may regard them as a different species of electrons from the ordinary 'cool' electrons in the metal, having a quasi-Fermi level which differs from the Fermi-level in the metal. As the hot electrons move into the metal, they lose energy by collisions, and their quasi-Fermi level falls to that of the metal (figure 7). This allows us to relax the requirement that the quasi-Fermi level at the boundary must coincide with the Fermi level in the metal; in particular, we may now go to the other extreme and assume that the quasi-Fermi level remains constant throughout the barrier region, as in a p–n junction. This is precisely the approximation made in the thermionic emission theory. It is equivalent to assuming that the concentration of electrons at the top of the barrier changes by a factor $\exp(eV/kT)$ when a forward bias voltage V is applied.

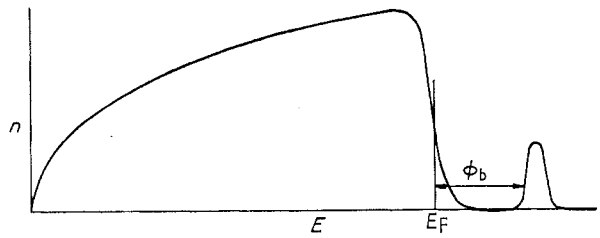

Figure 8. Energy distribution of electrons in metal. The parabolic distribution falling off at E_F represents the ordinary conduction electrons. The small 'pip' (much exaggerated) at an energy $E_F+\phi_b$ represents the 'hot' electrons injected into the metal from the semiconductor.

It can be shown by analysing existing experimental data (E. H. Rhoderick, to be published) that for high mobility semiconductors (silicon, gallium arsenide etc.) the thermionic emission theory is a good approximation and the diffusion theory a very bad one In other words, the bottleneck for current flow is provided by the actual process of emission of electrons into the metal rather than by the transport of electrons through the space charge region. According to the thermionic emission theory, the current density for an applied voltage V is given (see Henisch 1957) by

$$J=J_0\{\exp(eV/kT)-1\} \tag{3}$$

where

$$J_0=AT^2\exp(-\phi_b/kT). \tag{4}$$

Here A is the Richardson constant modified for the effective mass of the electrons in the semiconductor (Crowell 1965). For moderately large forward bias voltages such that $V>3\,kT/e$, the second term in the expression for J is negligible, and we may write

$$J=J_0\exp(eV/kT). \tag{5}$$

Junctions whose J–V characteristics can be represented by (3) or (5) are said to have 'ideal' characteristics.

4.2. *Hole injection*

The position in the literature with regard to hole injection is confusing. The early papers on the subject (e.g. Bardeen and Brattain 1949, Banbury 1953), which were all concerned with point contacts on n-type germanium, predicted that in the forward direction the current should be largely carried by holes, with an injection ratio approaching unity. This conclusion seems to have been based on the purely qualitative argument that, if the barrier height exceeds one-half of the energy gap, the surface of the semiconductor immediately underneath the metal is inverted, so that at the interface the density of holes exceeds the density of electrons. But the downward bending of the bands away from the interface presents a barrier to holes, and the injection ratio is determined not by the ratio of the electron and hole densities at the interface but by the ratio of the electron density at the interface to the hole density in the bulk of the semiconductor. This is generally much greater than unity even though the surface may be inverted.

Scharfetter (1965) has calculated the injection ratio in Schottky diodes by using standard p–n junction theory to calculate the flow of holes in the neutral region and the thermionic emission theory to calculate the electron current. He finds that for small forward bias voltages, for which the holes flow by diffusion, the injection ratio is very small (of the order of 10^{-4} in a typical case), but that above a critical current density ($\sim 10^4$ A m^{-2}) the electric field in the semiconductor assists the flow of holes and the injection ratio increases linearly with current density. Scharfetter's results have been confirmed experimentally by Yu and Snow (1969) for diodes made by evaporating gold films onto silicon.

It now seems firmly established, both experimentally and theoretically, that for large-area evaporated-film Schottky diodes on n-type semiconductors the injection ratio for holes is exceedingly small at low current densities. This makes it hard to understand how a point-contact transistor works; the probable explanation is that in a point-contact emitter the current densities are quite high ($\sim 10^5$ A m^{-2}), and exceed the critical value at which the injection ratio begins to increase.

4.3. *Recombination or generation in the depletion region*

Unlike hole injection, recombination (for forward bias) or generation (for reverse bias) within the depletion region may be very important at comparatively low current levels. The role of these processes was convincingly demonstrated in a classic paper by Yu and Snow (1968), who added a gate electrode which covered the surface of the semiconductor (in their case n-type silicon) immediately adjacent to the metal contact, as shown in figure 9. This enabled them to alter the potential of the semiconductor surface as in a MOS transistor, and to alter the area of the depletion region at will. By putting a negative bias on the gate, the surface of the semiconductor surrounding the metal contact could be strongly inverted, thereby increasing the area of the depletion region (see figure 10(*b*)). Because the p-type skin induced on the surface was in good electrical contact with the metal, this effectively increased the area for process 4(iii) without changing the area for process 4(i). In this way

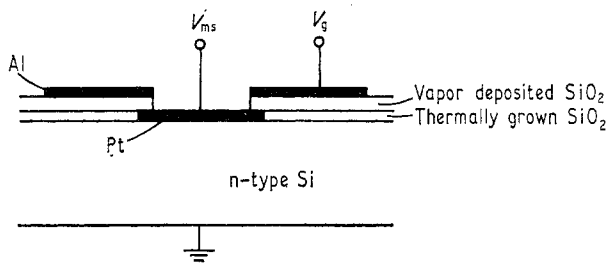

Figure 9. Schottky barrier with gate electrode (Yu and Snow 1968).

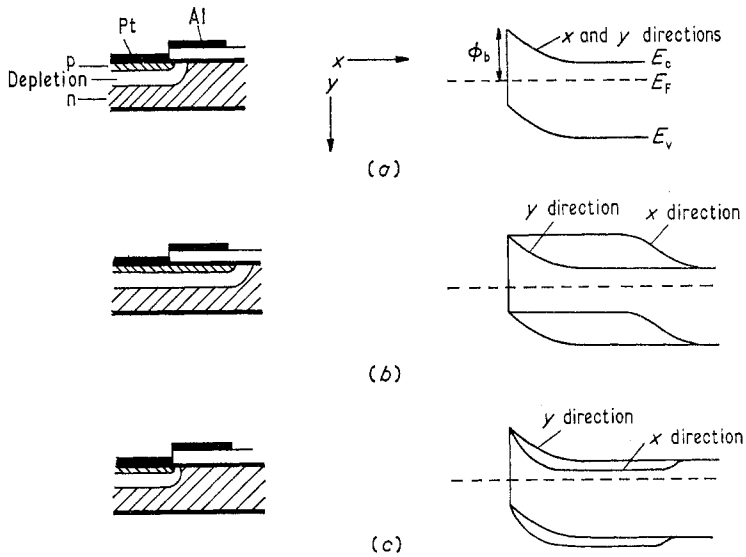

Figure 10. Band diagrams for Schottky barrier with gate (Yu and Snow 1968). (*a*) Flat bands ($V_G = V_{FB}$), (*b*) Surface inverted ($V_G \ll V_{FB}$), (*c*) Surface accumulated ($V_G \gg V_{FB}$).

the two processes could be separated, and it was concluded that for moderately large forward bias voltages the current is of the form

$$J=J_0 \exp(eV/kT)+J_r \exp(eV/2kT)$$

where the first term is the ordinary thermionic emission current given by (5), and the second term is the recombination current. J_r is given by

$$J_r=en_i dA'/\tau,$$

where n_i is the intrinsic electron concentration which is proportional to $\exp(-E_g/2kT)$, d is the thickness of the depletion region, A' its area, and τ the lifetime within the depletion region. If the negative bias on the gate is reduced until the bands adjacent to the metal contact are flat, the area of the depletion region A' becomes equal to the area of the contact, and the recombination current is reduced in proportion (figure 10(*a*)).

For a given semiconductor, the relative importance of thermionic emission and recombination/generation current may vary from one junction to another. For flat band conditions ($A=A'$), the ratio of the two terms depends on ϕ_b, E_g, T, and τ. It is possible in practice to find combinations of these such that the 'ideal' thermionic emission current dominates, but recombination/generation current represents a very common cause of departure from the ideal thermionic emission theory. If recombination current is important, the temperature variation of the forward current shows two activation energies (figure 11). Above room temperature the activation energy tends to the value ϕ_b, characteristic of the thermionic emission term, while below room temperature it approaches the value $E_g/2$ characteristic of the recombination current.

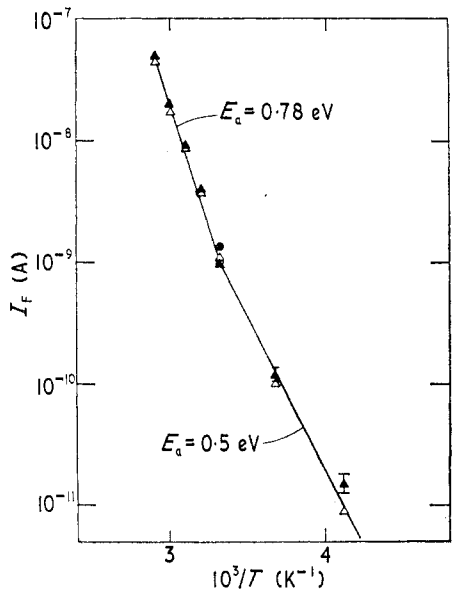

Figure 11. Forward current of platinum–silicon Schottky diode as function of temperature, for flat-band conditions. (Yu and Snow 1968.)

The reverse current under conditions where the surface adjacent to the metal is inverted or exhibits flat bands can be explained in terms of generation current. The current density can be well represented by

$$J=J_0+J_g$$

where the generation current is proportional to (reverse bias)$^{-1/2}$ because of the increase in thickness of the depletion region, and is also proportional to A'. The temperature dependence again shows two activation energies.

4.4. *The effects of image force lowering of the barrier and of funnelling*

Any charged body close to a conducting plane produces an electrostatic field as if there were an equal and opposite charge located at the mirror image of the body in the plane separating the conducting and insulating media. This image charge, being of opposite sign, lowers the potential energy of the body. The effect on a Schottky barrier is as shown in figure 12, and causes a lowering of the barrier by an amount $\Delta\phi$. The barrier lowering

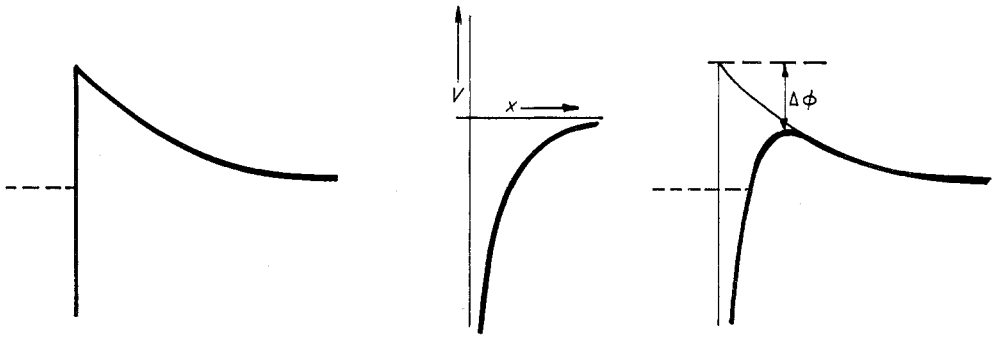

Figure 12. Image-force lowering of barrier. (*a*) Schottky barrier, (*b*) Image potential, (*c*) Resultant.

$\Delta\phi$ depends on the maximum electric field in the barrier (i.e. the field which would exist if there were no image force) and is therefore a function of the applied bias. The effects of this are twofold: (i) In the forward direction, the barrier increases with increasing forward bias so that J increases less rapidly with V than it would otherwise do. This changes the voltage-dependence of the forward characteristic to

$$J=J_0\{\exp(eV/nkT)-1\}$$

where n is a constant slightly larger than unity—typically 1·02 for gold on 1 Ω-cm silicon (Bethe 1942). (ii) In the reverse direction, the barrier decreases with increasing reverse bias, so that J does not saturate at the value J_0 given by (4). Instead, J increases gradually with reverse bias according to $J=J_0\exp(e\Delta V/kT)$, where

$$\Delta V=\{e^3N_d(V_d-V)/8\pi^2\epsilon_0^3\epsilon_s^3\}^{1/4}$$

is the lowering of the barrier due to the image force (Sze *et al.* 1964). Here N_d is the donor concentration and ϵ_s the relative permittivity of the semiconductor.

The effect of quantum-mechanical tunnelling through the barrier is important only for very thin barriers. For moderately doped semiconductors (up to $N_d\sim10^{16}$ cm^{-3} say) as are normally used in practical devices, tunnelling is significant only in the reverse direction where, although the width of the depletion region increases with reverse bias, the width of that part of the barrier through which electrons have to tunnel decreases. The effect of this is to make the reverse current increase rapidly above a certain voltage, as in a Zener diode.

Both image-force lowering and tunnelling may be exaggerated if the surface of the semiconductor outside the metal contact is accumulated. In this case the depletion layer is much thinner in a direction parallel to the surface than it is in a direction normal to the surface (see figure 10(*c*)), with the result that tunnelling and image-force lowering are both enhanced. In consequence, the 'n' value in the forward direction is a function of voltage (Yu and Snow 1968), and in the reverse direction the current increases very quickly with reverse bias, as shown in figure 13. Here the slow increase up to $V=-0{\cdot}3$ V is due to image-force lowering and the fast increase above $V=-1{\cdot}0$ V is due to tunnelling.

The 'ideality' of the I–V characteristics (low n value, good saturation in the reverse direction, high reverse breakdown) is thus very sensitive to the semiconductor surface potential

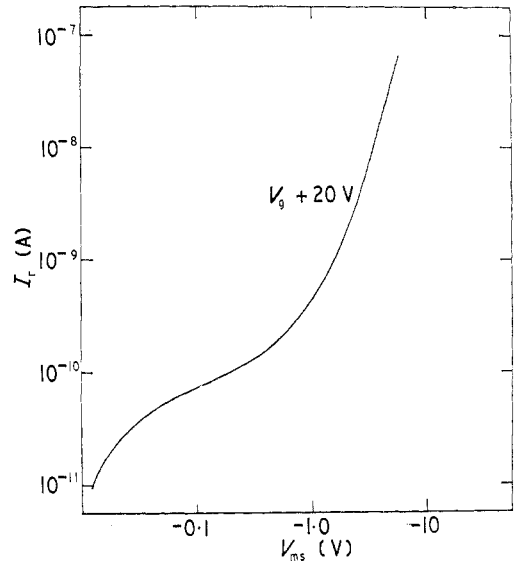

Figure 13. Reverse *I–V* characteristic for Pt–Si Schottky diode with accumulated surface (Yu and Snow 1968).

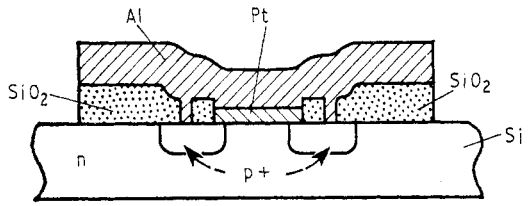

Figure 14. Schottky diode with p-type guard ring (Lepselter and Sze 1968).

adjacent to the metal, and in practice departures from ideality can often be associated with variations in surface potential resulting from surface films, adsorbed atoms or ions, etc. These effects can be largely overcome by using a guard ring consisting of a diffused p-type annulus (Lepselter and Sze 1968) as shown in figure 14. This has the effect of ensuring that the surface of the semiconductor immediately adjacent to the metal is inverted, thereby avoiding any pronounced effects due to image-force lowering or tunnelling. There will, of course, be an increase in recombination–generation current due to an increase in the area of the depletion layer, but this effect is relatively small. The characteristics of diodes prepared in this way are close to ideal.

4.5. *The effect of an insulating interfacial layer*

In practice, a metal–semiconductor junction, unless it is made by cleavage in ultra-high vacuum, will inevitably include a thin insulating layer of oxide between metal and semiconductor. This film is normally so thin ($\sim$10 Å) that electrons can tunnel though it quite freely. However, when a voltage difference exists across the junction, there will be a voltage drop across the film with the result that the change in the barrier height as seen from the semiconductor is less than it would otherwise have been. This makes the forward characteristic increase more slowly than exp (eV/kT). If the interfacial layer is so thin that the occupation of the surface states is determined by the Fermi level in the metal, it can be shown (Crowell and Sze 1966, Smith 1969) that the forward current is approximately proportional to exp (eV/nkt), where

$$n = 1 + \{\delta\epsilon_s\epsilon_0/\lambda(\epsilon_i\epsilon_0 + e\delta D_s)\},$$

ϵ_i being the relative permittivity of the insulating film, ϵ_s the relative permittivity of the semiconductor, and λ the width of the depletion layer. Since λ is proportional to $(V_d - V)^{1/2}$, n is not strictly constant but varies slightly with forward bias.

4.6. *Causes of departure from ideal characteristics*

To sum up, it is possible to make Schottky junctions in which the thermionic emission of electrons over the barrier is far more important than any other current mechanism. Such diodes have J–V characteristics which approximate very closely to (3). Deviations from this ideal J–V relationship may be due to:

(i) Recombination–generation current (composite forward characteristic with two activation energies; non-saturating reverse current).

(ii) Image force lowering of the barrier (n value slightly greater than 1; non-saturating reverse current).

(iii) Tunnelling (important only in reverse direction for moderate doping).

(iv) Insulating interfacial layer (n value greater than 1).

The first three of these are sensitive to the potential of the surface of the semiconductor immediately adjacent to the metal. If surface effects can be minimized, either by using a guard ring or by great attention to surface cleanliness, near-ideal characteristics can be obtained. For example, Lepselter and Sze (1968) obtained an exponential J–V relationship over 8 orders of magnitude and an n value of 1·02 for Pt–Si diodes using a p-type guard ring, and Smith (1968) has obtained similar near-ideal behaviour for Au–GaAs diodes, but with an n value of 1·06, by using a special cleaning procedure.

5. Transient effects

5.1. *Recovery time*

The recovery time which determines the switching speed of a Schottky diode depends on storage times of carriers in various parts of the device. Although the Schottky diode is commonly said to be a 'majority carrier' device, this is not strictly true. The electrons injected into the metal ('hot' electrons) are effectively minority carriers in the metal and can be sucked back by the application of reverse bias until they lose sufficient energy to stop them returning over the barrier. Roughly speaking, this happens when an injected electron collides with a conduction electron within the metal. The mean free path for this process is about 500 Å and the velocity about 10^6 m s^{-1}, so the storage time for injected electrons is about 10^{-13} s or less, which is negligible for most purposes. The storage time for injected holes is longer, and is determined by the recombination time which is typically 10^{-6} s for silicon. However, the contribution of holes to the recovery time is much smaller than this due to the very low injection ratio. The stored charge due to holes is $I_p\tau_p$, where I_p is the hole contribution to the forward current and τ_p the hole lifetime. If the circuit parameters are such as to limit the reverse current after switching to I_r, then the effective recovery time is $I_p\tau_p/I_r$. If I_r is of the order of the forward current, which is predominantly made up of the electron component I_n, then the effective recovery time is of the order of $I_p\tau_p/I_n$, which is reduced by the injection ratio I_p/I_n. In practice, recovery times of Schottky diodes are limited by their RC product rather than by charge storage, and times shorter than 10^{-11} s have been recorded for silicon diodes.

5.2. *Bulk traps*

It is not uncommon for some semiconductors (especially gallium arsenide) to contain deep-lying traps around the middle of the gap, and in gallium arsenide these can often have time constants of the order of seconds or minutes. These bulk traps may give rise to transient effects. For example, if the junction is reverse biased, the quasi-Fermi level for electrons may fall below the trap level and the traps begin to empty and become positively charged. The effect is as if the donor density were increased from N_d to $N_t + N_d$, where N_t is the trap density. The depletion layer gets thinner and the capacitance increases with

the time constant of the traps. If one waits until the capacitance is steady before making the measurement, the C^{-2}–V plot will have two straight line portions which intersect when the quasi-Fermi level for electrons first falls below the trap energy E_t. This gives a means of measuring E_t. The effect may be very undesirable if the C^{-2}–V plot is being used to measure N_d. The difficulty may be overcome if the diode is biased in the forward direction immediately before measurement and the measurement then made before the traps have time to empty. In this way true values of N_d may be obtained even in the presence of traps (Smith and Rhoderick 1969).

6. Conclusion

The present position (1970) seems to be that the basic properties of Schottky barriers are fairly well understood. There are still some areas (e.g. precise dependence of barrier height on work-function of metal, exact causes of departure from ideal J–V characteristics) where the agreement between theory and experiment is imperfect, but in most cases this is due to inadequate knowledge of the nature of the interface or of the surface properties of the semiconductor, and stems from limitations in the technology of fabrication rather than deficiencies in the theory. It is probably fair to say that Schottky junctions are now better understood than p–n junctions.

References

ARCHER, R. J., and ATALLA, M. M., 1963, *Ann. N.Y. Acad. Sci.*, **101**, 697–708.
ATALLA, M. M., 1966, *Proc. Munich Symp. on Microelectronics* (Munich: Oldenburg), pp. 123–57.
BANBURY, P. C., 1953, *Proc. Phys Soc.* B, **66**, 833–40.
BARDEEN, J., 1947, *Phys. Rev.* **71**, 717–27.
BARDEEN, J., and BRATTAIN, W. H., 1949, *Phys. Rev.*, **75**, 1208–25.
BETHE, H. A., 1942, *M.I.T. Radiat. Lab. Rep.*, 43–12.
BRAUN, F., 1874, *Ann. Phys., Lpz.*, **153**, 556–62.
COWLEY, A. M., and SZE, S. M., 1965, *J. appl Phys.*, **36**, 3212–20.
CROWELL, C. R., 1965, *Solid-St. Electron.*, **8**, 395–9.
CROWELL, C. R., and SZE, S. M., 1966, *Solid-St. Electron.*, **9**, 1035–48.
GOSSICK, B. R., 1963, *Solid-St. Electron.*, **6**, 445–52.
HENISCH, H. K., 1957, *Rectifying Semiconductor Contacts* (Oxford: Clarendon Press).
JONES, R. O., 1968, *Phys. Rev. Lett.*, **20**, 992–4.
LEPSELTER, M. P., and SZE, S. M., 1968, *Bell Syst. Tech. J.*, **47**, 195–208.
MEAD, C. A., 1966, *Solid-St. Electron.*, **9**, 1023–33.
MEAD, C. A., and SPITZER, W. G., 1964, *Phys. Rev.*, **134**, A713–16.
PICKARD, G. W., 1906, U.S. Patent No. 836531.
PIERCE, G. W., 1907, *Phys. Rev.*, **25**, 31–63.
POOLE, K. F., 1969, *Ph.D. Thesis*, University of Manchester.
SCHARFETTER, D. L., 1965, *Solid-St. Electron.*, **8**, 299–311.
SCHOTTKY, W., 1938, *Naturwiss.*, **26**, 843–53.
SCHOTTKY, W., and SPENKE, E., 1939, *Wiss. Veroff. Siemens Werke*, **18**, 1–67.
SHOCKLEY, W., 1949, *Bell Syst. Tech. J.*, **28**, 435–89.
SMITH, B. L., 1969, *Ph.D. Thesis*, University of Manchester.
—— 1968, *Electron. Lett.*, **4**, 332–3.
SMITH, B. L., and RHODERICK, E. H., 1969, *J. Phys.* D: *Appl. Phys.*, **2**, 465–7.
—— 1970, *Solid-St. Electron.*, in the press.
SZE, S. M., 1969, *Physics of Semiconductor Devices* (New York: John Wiley).
SZE, S. M., CROWELL, C. R., and KAHNG, D., 1964, *J. appl. Phys.*, **35**, 2534–6.
TURNER, M. J., and RHODERICK, E. H., 1968, *Solid-St. Electron.*, **11**, 291–300.
VAN LAAR, J., and SCHEER, J. J., 1967, *Surf. Sci.*, **8**, 342–56.
WAGNER, C., 1931, *Phys. Z.*, **32**, 641–5.
YU, A. Y. C., and SNOW, E. H., 1968, *J. appl. Phys.*, **39**, 3008–16.
—— 1969, *Solid-St. Electron.*, **12**, 155–60.