Metal-semiconductor contacts

Prof. E.H. Rhoderick, M.A., M.Sc., Ph.D., C.Eng., F.Inst. P., F.I.E.E.

Indexing terms: *Semiconductor devices and materials, Schottky-barrier devices*

Abstract: A review is given of our present knowledge of metal-semiconductor contacts. Topics covered include the factors that determine the height of the Schottky barrier, its current/voltage characteristics, and its capacitance. A short discussion is also given of practical contacts and their application in semiconductor technology, and a comparison is made with *p-n* junctions.

1 Historical

The study of metal-semiconductor contacts goes back to 1874, when Braun reported the asymmetrical nature of conduction between metal points and crystals such as lead sulphide. Their application as radio-frequency detectors is almost as old as wireless telegraphy itself, and in 1906 Pickard took out a patent for a point-contact rectifier using silicon. In 1907 Pierce published rectification characteristics of diodes made by sputtering metals onto a variety of semiconductors, and the first copper-oxide plate rectifiers appeared in the early 1920s. The point-contact rectifier or 'cat's-whisker' 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 contact behaviour has come with the realisation that metal films evaporated onto single-crystal semiconductor surfaces under conditions of high cleanliness can show almost ideal rectification characteristics. The intensive study of contacts in the 1960s and 1970s was largely stimulated by their importance in semiconductor technology, both as rectifying elements and as low resistance or 'ohmic' contacts. This activity continues unabated, with emphasis on the metallurgical and reliability aspects and also on the underlying physics.

An extensive account of the early history of metal-semiconductor contacts can be found in Henisch's book [1], and a more up-to-date review has recently been given by this author [2]. There are also some short reviews that may be helpful to the reader [3–5].

2 Formation of barrier

2.1 Schottky-Mott theory

The rectifying properties of a metal-semiconductor contact arise from the presence of an electrostatic barrier between the metal and the semiconductor. This barrier is due to the difference in work functions of the two materials. If the work function of the metal ϕ_m exceeds that of the semiconductor ϕ_s, electrons pass from the semiconductor into the metal to equalise the Fermi levels, leaving behind a depletion region in the semiconductor in which the bands are bent upwards (Fig. 1*a* for the case of an *n*-type semiconductor). Assuming that the region of the semiconductor where the bands are bent upwards is completely devoid of conduction electrons (the so-called 'depletion approximation'), the space charge is due entirely to the uncompensated donor ions. If these are uniformly distributed, there will be a uniform space charge in the depletion region, and the electric field strength will increase linearly with distance from the edge of the depletion region as the metal is approached, in accordance with Gauss's theorem. The magnitude of the electrostatic potential will increase quadratically, and the resulting potential barrier will be parabolic in shape. This is known as a Schottky barrier.

It can be shown by a straightforward argument (e.g. Reference 2) that the amount by which the bands are bent upwards (the so-called diffusion potential V_{do}) is given by

$$V_{do} = \phi_m - \phi_s \tag{1}$$

All energies are measured in electron-volts, and so the energy of an electron due to its electrostatic potential is equal to the magnitude of the potential expressed in volts. If $\phi_m > \phi_s$, V_{do} is positive and the bands are bent upwards; for the case of an *n*-type semiconductor this produces a barrier which the electrons have to surmount in order to pass from the semiconductor into the metal, as in Fig. 1*a*, which, we shall see, leads to rectifying properties. On the other hand, for a *p*-type semiconductor (Fig. 1*b*), the band-bending causes no impediment to the motion of holes, and no rectification takes place, giving an 'ohmic' contact.

If $\phi_m < \phi_s$, the bands are bent downwards. This gives an ohmic contact for an *n*-type semiconductor (Fig. 1*c*) and, since holes have difficulty in passing underneath a barrier, a rectifying contact for a *p*-type semiconductor (Fig. 1*d*). In nearly all cases $\phi_m > \phi_s$ for *n*-type semiconductors, but $\phi_m < \phi_s$ for *p*-type semiconductors, and so most metal-semiconductor combinations form rectifying contacts. Unless the contrary is clearly stated, all subsequent discussions will centre round the case of *n*-type semiconductors with $\phi_m > \phi_s$, which is the most important case in practice.

What is usually quoted is not the diffusion potential but the barrier height ϕ_b as viewed from the metal (Fig. 1). For an *n*-type semiconductor, this is given by

$$\begin{aligned} \phi_{bn} &= V_{do} + (E_c - E_F) \\ &= \phi_m - \chi_s \end{aligned} \tag{2}$$

where $\chi_s \{= \phi_s - (E_c - E_F)\}$ is the electron affinity of the semiconductor, i.e. the difference in energy between the vacuum level and the bottom of the conduction band. Although usually attributed to Schottky [6], eqn. 2 was first stated implicitly by Mott [7], and will be referred to as the Schottky-Mott approximation. In obtaining it, the assumption is made that the surface dipole contributions to ϕ_m and χ_s do not change when the metal and semiconductor come into contact.

In most practical metal-semiconductor contacts the ideal situation shown in Fig. 1*a* is never reached, because there is usually a thin oxide layer, about 1–2 nm thick, on the surface of the semiconductor. Such an oxide film is referred to as an interfacial layer. A practical contact is, therefore, more like Fig. 1*e*; however, the additional barrier presented by the oxide layer is usually so thin that electrons can penetrate it quite easily by quantum-mechanical tunnelling, and so there is little

Paper 1760I, received in final form 1st December 1981. Commissioned IEE Review

Prof. Rhoderick is Professor of Solid-State Electronics, Department of Electrical Engineering & Electronics, University of Manchester Institute of Science & Technology, PO Box 88, Manchester M60 1QD, England

0143–7100/82/010001 + 14 $01.50/0

GF Exhibit 1023 - 1/14

difference between Fig. 1*a* and *e* as far as the electrical characteristics are concerned.

2.2 Modifications to Schottky-Mott theory

The Schottky-Mott approximation (eqn. 2) assumes that the surface dipole contributions to ϕ_m and χ_s do not change when the metal and the semiconductor are brought into contact. These surface dipole layers arise because at the surface of a solid the atoms have neighbours on one side only. This causes a distortion of the electron cloud belonging to the surface atoms, so that the centres of the positive and negative charge distributions do not coincide. It was soon discovered that the linear dependence of ϕ_{bn} on ϕ_m predicted by eqn. 2 does not occur in practice, and so the assumption of constancy of the surface dipole layers cannot be true.

One of the first explanations for the departures of the experimental data from eqn. 2 was given by Bardeen [8], who pointed out the importance of localised surface states. For our present purpose it is sufficiently accurate to regard surface states as unsatisfied bonds on the surface of the semiconductor. At the surface, the atoms have neighbours on one side only, and on the vacuum side the valence electrons have no partners with which to form covalent bonds. Each surface atom, therefore, has associated with it an unpaired electron in a localised orbital, directed away from the surface. Such an orbital is often spoken of as a dangling bond. It can either give up its electron, acting as a donor, or accept another, acting as an acceptor. The surface states are usually continuously distributed in energy within the forbidden gap, and are characterised 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. In general, the Fermi level does not coincide with the neutral level, and in this case there will be a net charge in the surface states. If, in addition, there is a thin oxide layer between the metal and the semiconductor, as will happen if the surface of the latter has been prepared by chemical polishing, the charge in the surface states together with its image charge on the surface of the metal will constitute a dipole layer. This dipole layer will alter the potential difference between the semiconductor and the metal and will upset eqn. 2. It was shown by Cowley and Sze [9] that, according to the Bardeen model, the barrier height is given approximately by

$$\phi_{bn} = \gamma(\phi_m - \chi_s) + (1-\gamma)(E_g - \phi_0) \qquad (3)$$

where

$$\gamma = \frac{\epsilon_i}{\epsilon_i + q\delta D_s}$$

E_g is the bandgap of the semiconductor, δ the thickness of the oxide layer, and ϵ_i its total permittivity. The surface states are assumed to be uniformly distributed in energy within the bandgap, with a density D_s per electron-volt per unit area. The position of the neutral level ϕ_0 is measured from the top of the valence band.

If there are no surface states, $D_s = 0$ and $\gamma = 1$, and so eqn. 3 gives $\phi_{bn} = \phi_m - \chi_s$, which is the Schottky-Mott approximation. If the density of states is very high, γ becomes very small and ϕ_{bn} approaches the value $E_g - \phi_0$. This is because a very small deviation of the Fermi level from the neutral level can produce a large dipole moment, which stabilises the barrier height by a sort of negative feedback effect. The Fermi level is said to be 'pinned' relative to the bandedges by the surface states.

A similar analysis for the case of a *p*-type semiconductor shows that ϕ_{bp} is approximately given by

$$\phi_{bp} = \gamma(E_g - \phi_m + \chi_s) + (1-\gamma)\phi_0 \qquad (4)$$

Hence if ϕ_{bn} and ϕ_{bp} refer to the same metal on *n*- and *p*-type specimens of the same semiconductor, we should have

$$\phi_{bn} + \phi_{bp} \simeq E_g \qquad (5)$$

if the semiconductor surface is prepared in the same way in both cases, so that δ, ϵ_i, D_s, and ϕ_0 are the same. This relationship holds quite well in practice [10]. It is usually true that $\phi_{bn} > E_g/2$, and so $\phi_{bn} > \phi_{bp}$.

Some experimental measurements of the barrier height obtained by depositing films of various metals by evaporation onto chemically etched surfaces of *n*-type silicon are shown in Fig. 2. These are plotted against the work function of the metal (see Reference 2, Table 2.1). There is considerable scatter in the values obtained by different authors with the same metal; this is largely because, as we have seen, the barrier height depends on the thickness and composition of the thin insulating layer, and this is bound to depend on the method of preparing the surface. The wide scatter in the case of aluminium is probably because aluminium oxidises very easily, and therefore tends to modify the chemical nature of any oxide that may be on the surface of the silicon. It cannot really be said that the data conform to eqn. 3, although metals with high work functions (e.g. gold and platinum) tend to give

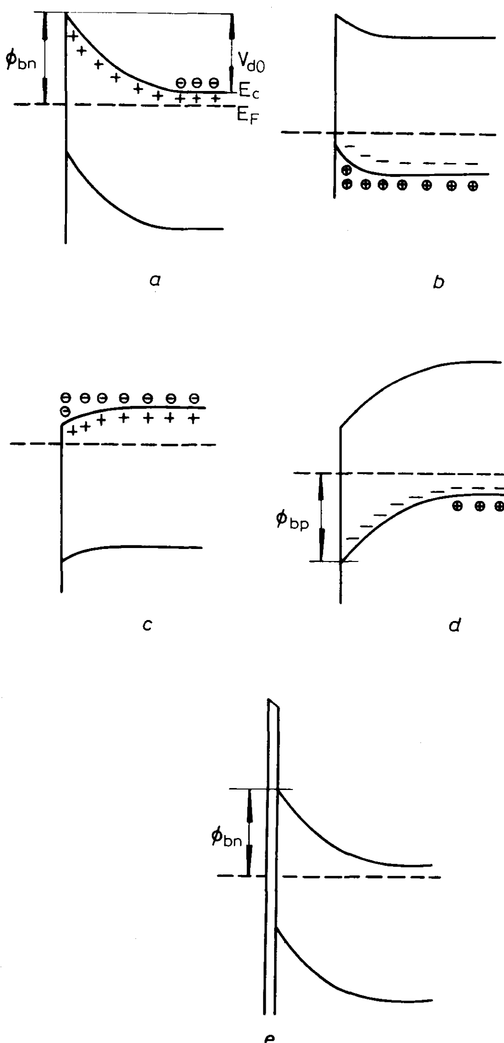

Fig. 1 *Schottky barriers for semiconductors of different types and work functions*

a $\phi_m > \phi_s$, *n*-type *b* $\phi_m > \phi_s$, *p*-type
c $\phi_m < \phi_s$, *n*-type *d* $\phi_m < \phi_s$, *p*-type
e $\phi_m > \phi_s$, *n*-type, with interfacial layer

⊖ electron in conduction band — acceptor ion
⊕ hole in valence band + donor ion

GF Exhibit 1023 - 2/14

large barrier heights, whereas metals with low work functions (e.g. magnesium and titanium) tend to give small barrier heights. Data on barrier heights in other semiconductors are given in Reference 2.

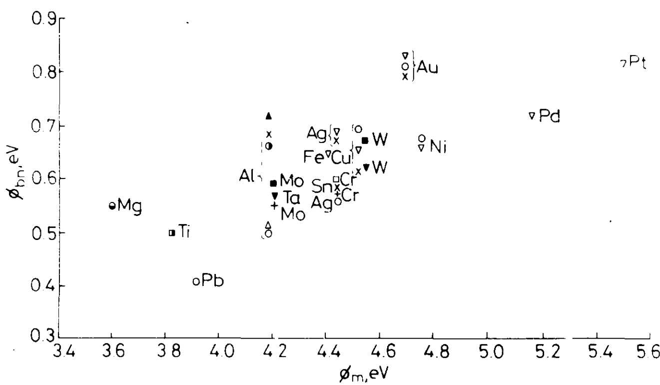

Fig. 2 *Barrier heights of various metals on chemically etched n-type silicon*

Symbols denote results of various authors

2.3 *Intimate contacts*

The Bardeen model assumes the existence of an insulating layer between the metal and semiconductor, as evidenced by the occurrence of δ and ϵ_i in the expression for γ (eqn. 3). This corresponds quite closely to the majority of practical contacts, which are fabricated in such a way that there is a thin oxide layer on the surface of the semiconductor. Occasionally, for research purposes, contacts are made by cleaving a crystal of the semiconductor in an ultra-high-vacuum system and then evaporating a metal film before there is time for an oxide layer to form on the freshly created surface. Such a contact is known as an 'intimate' contact, and is devoid of any interfacial layer. We must now ask whether Bardeen's model of the effect of surface states can be applied to such an intimate contact.

In Bardeen's analysis the interface states are regarded as point charges, and the term $q\delta$ in the expression for γ is simply the dipole moment of a charged interface state together with its balancing charge on the surface of the metal. But the interface states actually extend into the semiconductor to a distance of about 1 nm, and even if there is no interfacial layer there is still a dipole between a charged interface state and the surface of the metal. This point of view has been developed by Heine [13], who regards the interface states as the exponentially decaying tails of the wave functions of the conduction electrons in the metal, which can penetrate into the bandgap of the semiconductor by tunnelling. The situation is much more difficult to analyse theoretically than when there is an interfacial layer present, because the density of the interface states and the position of the neutral level ϕ_0 both depend on the metal as well as on the semiconductor [14].

Our present ideas of barrier formation at intimate metal-semiconductor contacts are influenced very largely by four recent pieces of experimental evidence.

(*a*) It is now well established (e.g. from the work of Thanailakis and co-workers [11, 12]) that ϕ_{bn} is not a monotonically increasing function of ϕ_m. Indeed, there are some groups of metals for which ϕ_{bn} actually decreases as ϕ_m increases.

(*b*) It has recently been discovered that the nature of the barrier produced when a particular metal makes contact with a semiconductor can be correlated with the chemical properties of the metal, i.e. with whether or not a chemical reaction takes place at the interface [15].

(*c*) It is now recognised that ideally sharp boundaries at metal-semiconductor junctions hardly ever occur in practice, even if the metal is deposited at room temperature on a cleaved semiconductor in an ultra-high-vacuum. There may be an alloy phase at the interface, or there may be a gradual transition from metal to semiconductor over a distance of 10 Å or more due to interdiffusion effects [16].

(*d*) It has been found that certain crystal faces of some semiconductors (e.g. the cleaved (110) surface of gallium arsenide) do not possess any intrinsic surface states in the bandgap, yet the Fermi level may be 'pinned' by as little as one-tenth of a monolayer of an evaporated metal [17]. (By 'intrinsic' surface states are meant states which exist at an ideal semiconductor surface exposed to vacuum.)

All these findings can be explained in terms of the following picture:

(i) The short-range dipole at an intimate metal-semiconductor interface is not equal to the difference between the surface dipoles at the metal-vacuum and semiconductor-vacuum surfaces.

(ii) The interface dipole depends on the precise spatial arrangement of the constituent atoms at the interface, and on whether or not they form chemical bonds with each other. This stage depends very much on the chemical reactivity of the metal, and there seems to be an empirical correlation between ϕ_b and the 'heat of reaction' between the metal and semiconductor.

(iii) Barrier heights can be significantly affected by interdiffusion effects, even in the absence of heat-treatment.

(iv) In the case of the (110) surfaces of III–V compounds (e.g. GaAs, InP) which show no intrinsic surface states in the bandgap, the 'pinning' of the Fermi level by a thin overlayer of metal is probably due to the creation of extrinsic surface states. These are thought by Spicer *et al.* [17] to be crystal defects created by the heat of condensation of the metal. Such defects may be missing group III or group V atoms, or 'antisites' in which a pair of atoms have exchanged places.

2.4 *Summary*

Our present understanding of the mechanism of barrier formation is still highly imperfect, especially where intimate contacts are concerned, although progress is being made at quite a rapid rate. 'Real' contacts, i.e. those in which there is an oxide layer between metal and semiconductor, are rather better understood than intimate contacts because the oxide tends to reduce the importance of interdiffusion, and also because, to a certain extent, it 'decouples' the electron states in the semiconductor from the influence of the metal, and so they can be analysed more simply. Nevertheless, it should be emphasised that the Bardeen theory, as extended by Cowley and Sze, uses a very idealised model and should not be expected to explain the finer points of barrier formation, although for many purposes it provides a useful working picture.

Anyone wishing to know what barrier height is likely to occur in a particular metal-semiconductor combination made by a particular technique has little alternative to searching the literature to find out whether that particular combination has been reported using the same recipe. A summary up to 1977 of results for the more common semiconductors is given in Reference 2, together with a description of methods of measuring barrier heights.

3 Current/voltage relationship

The various ways in which current can be transported through a metal-semiconductor contact under forward bias are shown in Fig. 3 and include the following:

GF Exhibit 1023 - 3/14

(*a*) emission of electrons from the semiconductor over the top of the barrier into the metal

(*b*) quantum mechanical tunnelling through the barrier

(*c*) recombination in the space-charge region

(*d*) recombination in the neutral region ('hole injection').

The inverse processes occur under reverse bias.

It is possible to make practical Schottky barrier diodes in which process (*a*) is far and away the most important, and such diodes will be referred to as 'nearly ideal'. Processes (*b*), (*c*), and (*d*) cause departures from this ideal behaviour.

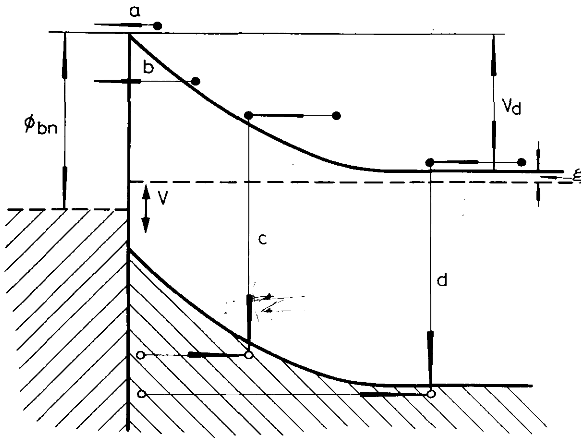

Fig. 3 *Transport processes in forward-biased Schottky barrier on n-type semiconductor*

---- Fermi level according to thermionic-emission theory

3.1 Emission over the barrier

3.1.1 Theory: Before an electron can be emitted over the barrier into the metal, it must first be transported through the depletion region of the semiconductor. In the latter process its motion is determined by the usual mechanisms of diffusion and drift, while the emission process is controlled by the number of electrons that impinge on unit area of the metal per second. These two processes are essentially in series, and the current is determined predominantly by whichever causes the larger impediment to the flow of electrons.

Historically, the first theory of conduction in Schottky diodes was the 'diffusion theory' of Wagner [18] and Schottky and Spenke [19]. According to this theory, the current is limited by diffusion and drift in the depletion region, and the assumption is made that the conduction electrons in the semiconductor immediately adjacent to the metal are in thermal equilibrium with those in the metal. In contrast, the 'thermionic-emission theory' proposed by Bethe [20] assumes that the current is limited by the emission process, and that the quasi-Fermi level for electrons remains horizontal throughout the depletion region, as in a *p-n* junction (see Fig. 3). One can avoid the necessity of postulating a discontinuity in the quasi-Fermi level at the metal-semiconductor interface by regarding the electrons emitted into the metal as 'hot' electrons with their own quasi-Fermi level [21].

For an *n*-type semiconductor, the J/V relationship predicted by the diffusion theory can be shown [22] to be

$$J = qN_c\mu_n \mathscr{E}_{max} \exp(-q\phi_{bn}/kT)\{\exp(qV/kT) - 1\} \tag{6}$$

where J is the current density per unit area, N_c the effective density of states in the conduction band of the semiconductor, μ_n the electron mobility, $\mathscr{E}_{max}$ the maximum electric field, and ϕ_{bn} the barrier height. This is not quite of the form of the ideal rectifier equation $J = J_0\{\exp(qV/kT) - 1\}$ because $\mathscr{E}_{max}$ is voltage dependent.

The I/V characteristic for the case of the thermionic-emission theory can easily be derived if one realises that, under the application of a forward bias V, a flat quasi-Fermi level implies that the electron concentration in the semiconductor just inside the interface is given by

$$n = N_c \exp\{-q(\phi_{bn} - V)/kT\}$$

(see Fig. 3). The flux of these electrons across the interface into the metal can be shown by elementary kinetic theory to be $n\bar{v}/4$, where $\bar{v}$ is the average thermal velocity of electrons in the semiconductor. The flux in the reverse direction, which is independent of V, must exactly balance the flux from semiconductor to metal when no bias is applied, and so the net current density is given by

$$J = \frac{qN_c\bar{v}}{4}\exp(-q\phi_{bn}/kT)\{\exp(qV/kT) - 1\} \tag{7}$$

$$= A^*T^2 \exp(-q\phi_{bn}/kT)\{\exp(qV/kT) - 1\} \tag{8}$$

where $A^* = 4\pi m^* qk^2/h^3$ is the Richardson constant corresponding to the effective mass in the semiconductor. Crowell [23] has shown that, for a semiconductor with ellipsoidal constant energy surface, the appropriate effective mass for a single valley is $(l^2 m_y m_z + m^2 m_z m_x + n^2 m_x m_y)^{1/2}$, where l, m, n are the direction cosines of the normal to the interface relative to the axes of the ellipsoid, and m_x, m_y and m_z are the components of the effective mass tensor. For the case of silicon with the junction parallel to a $\{111\}$ plane

$$m^* = 6\{(m_t^2 + 2m_t m_l)/3\}^{1/2}$$

where m_t and m_l are the transverse and longitudinal effective masses. This can be easily understood because N_c is proportional to the density-of-states effective mass $(m_t^2 m_l)^{1/3}$ raised to the three-halves power, and $\bar{v}$ is inversely proportional to the square root of the conductivity effective mass $\{(2/m_t + 1/m_l)/3\}^{-1}$. The factor 6 comes from the number of valleys.

It is often assumed that the condition for the validity of the thermionic emission theory is that the mean free path of the electrons should exceed the width of the depletion region. This is an unnecessarily stringent requirement, since the theory merely requires that the electron density at the top of the barrier is in equilibrium with the bulk of the semiconductor, and this can be the case even if the electrons make many collisions in negotiating the depletion region. Bethe argued that the condition for the validity of the thermionic-emission theory is merely that the mean free path should exceed the distance within which the barrier falls by kT/q from its maximum value. This criterion has also been derived by Gossick [21] and by Crowell and Sze [24]. Crowell and Sze also take into account the effects of optical phonon scattering in the region between the top of the barrier and the metal† and of the quantum mechanical reflection of electrons which have sufficient energy to surmount the barrier. For the case of the thermionic-emission theory, their combined effect is to replace the Richardson constant A^* with

$$A^{**} = f_p f_q A^*$$

where f_p is the probability of an electron reaching the metal without being scattered by an optical phonon after having passed the top of the barrier, and f_q is the average transmission

†In general the top of the barrier occurs within the semiconductor owing to the effect of the image force (see Section 3.1.2)

GF Exhibit 1023 - 4/14

coefficient. f_p and f_q depend on the maximum electric field in the barrier, on the temperature, and on the effective mass. They have been calculated by Crowell and Sze for silicon, germanium, and gallium arsenide. Generally speaking, the product $f_p f_q$ is of the order of 0.5. The values of A^{**} given by Crowell and Sze for {111}-oriented silicon and gallium arsenide are 96×10^4 $Am^{-2}K^{-2}$ and 4.4×10^4 $Am^{-2}K^{-2}$, respectively.

3.1.2 Effect of bias-dependence of barrier height: There are several reasons why the barrier height ϕ_b may depend on the applied bias. If there is an interfacial layer, the voltage drop in this layer reduces the barrier height by an amount proportional to the maximum field in the barrier, which in turn depends on the applied bias. Furthermore, even if there is no interfacial layer, the barrier height depends on the bias because of the effect of the image force. The image force arises because an electron near to the surface of the metal is attracted to it by the positive image charge. This force has the effect of reducing the barrier height by an amount that depends on the electric field in the semiconductor, and hence on the applied bias (see e.g. Reference 2, p. 37).

Let us consider the general case of either n-type or p-type semiconductors, and suppose that ϕ_b depends linearly on the applied bias V, which is true for small values of V, so that we can write $\phi_b = \phi_{bo} + \beta V$. The coefficient β is positive because ϕ_b always increases with increasing forward bias. We can now rewrite eqn. 8 as

$$J = A^{**}T^2 \exp\{-q(\phi_{bo} + \beta V)/kT\}\{\exp(qV/kT) - 1\}$$
$$= J_0 \exp(-q\beta V/kT)\{\exp(qV/kT) - 1\} \quad (9)$$

where

$$J_0 = A^{**}T^2 \exp(-q\phi_{bo}/kT)$$

Eqn. 9 can be written in the form

$$J = J_0 \exp(qV/nkT)\{1 - \exp(-qV/kT)\} \quad (10)$$

where

$$\frac{1}{n} = 1 - \beta = 1 - \frac{\partial \phi_b}{\partial V} \quad (11)$$

n is often called the 'ideality factor'. If $\partial\phi_b/\partial V$ is constant, n is also constant. For values of V greater than $3kT/q$, eqn. 10 can be written as

$$J = J_0 \exp(qV/nkT) \quad (10a)$$

Eqn. 10 is often written in the literature in the form

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

This form is incorrect, because the barrier lowering must affect the flow of electrons from metal to semiconductor as well as the flow from semiconductor to metal, and so the second term on the right of eqn. 10 must contain n. The correct form, eqn. 10, has the advantage that a plot of $\ln[J/\{1 - \exp(-qV/kT)\}]$ against V should be a straight line, even for values of V less than $3kT/q$. The intercept of the straight line on the vertical axis gives the value of J_0, and a knowledge of A^{**} allows the zero-bias barrier height ϕ_{bo} to be deduced.

If there is no interfacial layer, the bias dependence arises solely from the effect of the image force, and it can be shown that

$$\frac{1}{n} = 1 - \frac{1}{4}\left(\frac{q^3 N_d}{8\pi^2 \epsilon_s^3}\right)^{1/4} \left(\phi_b - V - \xi - \frac{kT}{q}\right)^{-3/4} \quad (12)$$

where $\xi = E_c - E_F$ for n-type material, or $E_F - E_v$ for p-type, and $\epsilon_s(= \epsilon_r\epsilon_0)$ is the total permittivity of the semiconductor. Hence n is not constant but depends on V. If V is restricted to values less than about $\phi_b/4$, n is roughly constant. Taking $\epsilon_s \simeq 10^{-10}$ Fm^{-1}, appropriate to silicon or gallium arsenide, and $\phi_b - \xi \simeq 0.5$ eV, n has the value 1.02 for $N_d \simeq 10^{23}$ m^{-3}. The effect of image-force lowering is therefore negligible for Schottky barriers with $N_d < 10^{23}$ m^{-3}, although, as we shall see, it may be more important under reverse bias.

3.1.3 Comparison with experiment: There are good theoretical reasons for believing that Schottky diodes made from reasonably high mobility semiconductors should conform to the thermionic-emission theory, for moderate forward voltages at least [25], and this conclusion has been confirmed by an analysis of experimental data on silicon and gallium arsenide diodes made by Rhoderick [26]. The only data known to the author which appear to conform to the diffusion theory are those which relate to copper oxide [1] and some recent results on amorphous silicon [27].

3.2 Tunnelling through the barrier

It is possible for electrons with energies below the top of the barrier to penetrate the barrier by quantum mechanical tunnelling. This is of historical interest because it was postulated by Wilson [28] in 1932 that tunnelling is the dominant process in Schottky barriers. It was soon realised, however, that this process alone gives the wrong direction for rectification. It may, nevertheless, modify the ordinary thermionic emission process in one of two ways which may be understood by reference to Fig. 4. In the case of a degenerate

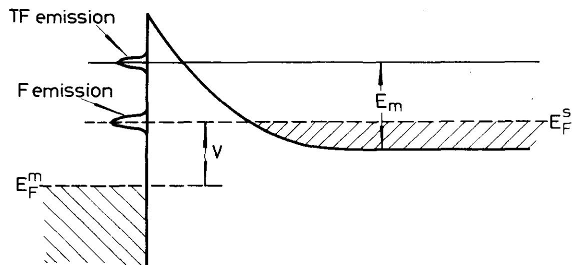

Fig. 4 *Field and thermionic-field emission under forward bias [29]*

semiconductor at low temperature, where the donor density is so high, and the potential barrier so thin, that tunnelling can easily occur, the current in the forward direction arises from electrons with energies close to the Fermi energy in the semiconductor. This is known as field (F) emission. If the temperature is raised, electrons are excited to higher energies, and the tunnelling probability increases very rapidly because the electrons 'see' a thinner and lower barrier. On the other hand, the number of electrons having a particular energy decreases very rapidly with increasing energy, and there will be a maximum contribution to the current from electrons which have an energy E_m above the bottom of the conduction band. This is known as 'thermionic-field' (TF) emission. If the temperature is raised still further, a point is eventually reached at which virtually all of the electrons have enough energy to go over the top of the barrier; the effect of tunnelling is negligible, and so we have pure thermionic emission.

The theory of F and TF emission has been developed by Padovani and Stratton [29] and by Crowell and Rideout [30]. Both these analyses are very mathematical, but the essential features are as follows:

(i) Field emission occurs only in degenerate semiconductors, and because of the very small effective mass, it shows up at lower concentrations in gallium arsenide than in most other semiconductors. The ranges of temperatures and concentrations

GF Exhibit 1023 - 5/14

over which Au-GaAs Schottky barriers exhibit F and TF emission are shown in Fig. 5 [31].

(ii) Except for very low values of V, the forward current-voltage relationship is of the form

$$J = J_s \exp(V/E_0) \quad (13)$$

where, in the notation of Padovani and Stratton.

$$E_0 = E_{00} \coth(qE_{00}/kT)$$

and

$$E_{00} = \frac{h}{4\pi}\left(\frac{N_d}{m^* \epsilon_s}\right)^{1/2}$$

$$= 18.5 \times 10^{-15}\left(\frac{N_d}{m_r \epsilon_r}\right)^{1/2} \text{ electron-volts}$$

Here m^* $(= m_r m_0)$ is the effective mass of electrons in the n-type semiconductor, ϵ_s $(=\epsilon_r \epsilon_0)$ its permittivity, and the donor concentration N_d is expressed in m^{-3}. The pre-exponential term J_s is a complicated function of the temperature, barrier height, and semiconductor parameters, and is given graphically by Crowell and Rideout as a function of kT/qE_{00}.

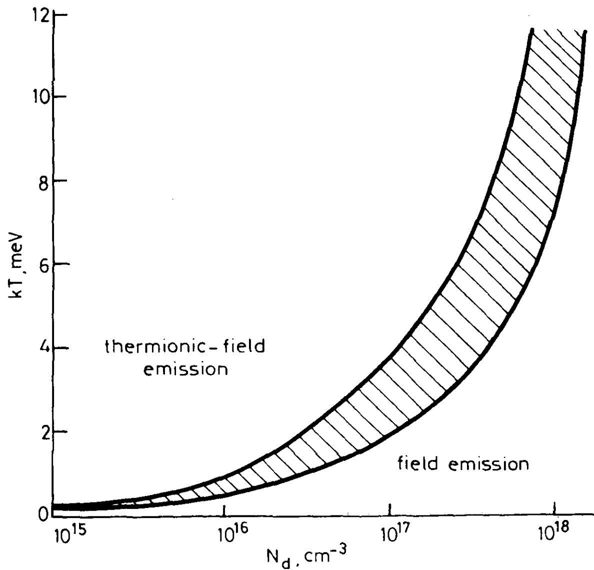

Fig. 5 *Ranges of temperature and donor concentration over which n-type gallium arsenide Schottky diodes exhibit field and thermionic-field emission [31]*

At low temperatures $(kT/qE_{00} \ll 1)$, $E_0 \simeq E_{00}$, and so the slope of the graph of $\ln J$ against V is independent of temperature. This is the case of F emission. At high temperatures $(kT/qE_{00} \gg 1)$, E_0 approaches the value kT/q, which corresponds to pure thermionic emission. For temperatures such that $kT \sim qE_{00}$ we have TF emission, and the slope of the $\ln J$ against V curve can be written as q/nkT, where $n = qE_0/kT = (qE_{00}/kT) \coth(qE_{00}/kT)$. The maximum in the energy distribution of the emitted electrons occurs at an energy $E_m = V_d\{\cosh(qE_{00}/kT)\}^{-2}$ above the bottom of the conduction band in the bulk semi-conductor. Fig. 6 shows the variation of n and of E_m/V_d as a function of kT/qE_{00} in the TF regime. If we take $n > 1.02$ and $E_m/V_d < 0.95$ as criteria for the transition from pure thermionic emission to TF emission, we see that they occur when $kT/qE_{00} < 4$. This corresponds to $N_d > 10^{23}$ m^{-3} and 10^{22} m^{-3} in GaAs at 300 K and 77 K, respectively, whereas in Si the corresponding values of N_d are about a factor of four greater.

Apart from changing the form of the J/V relationship, tunnelling also causes a significant increase in the magnitude of J, since it represents a process in parallel with thermionic emission. For this reason field emission is of considerable importance in connection with ohmic contacts to semiconductors, which often consist of Schottky barriers on very highly doped material. What matters in this case is the specific resistance around zero bias [i.e. $(dV/dJ)_{v=0}$], for which the current/voltage relationship referred to above is not valid. Yu [32] has compared the zero-bias calculations of Padovani [31] with experimental results on silicon with N_d in the range 10^{24} m^{-3} to 10^{26} m^{-3}, and Vilms and Wandinger [33] have done similar work using calculations of their own based on a simplified barrier model. Their results for Si Schottky barriers are shown in Fig. 7.

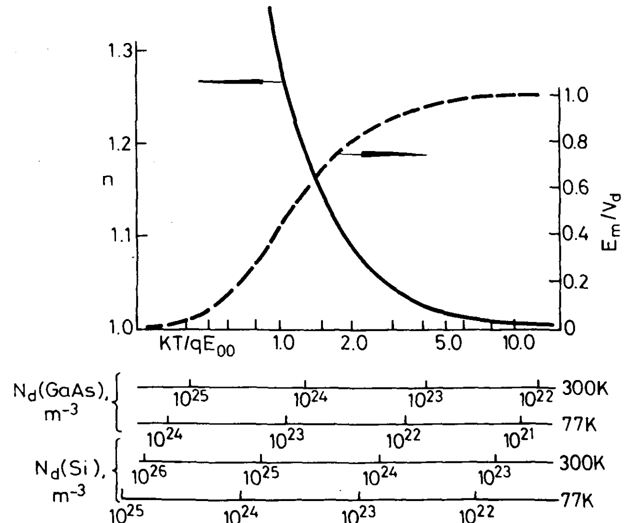

Fig. 6 *Ideality factor n and position of maximum of energy distribution E_m of emitted electrons as function of kT/qE_{00}*

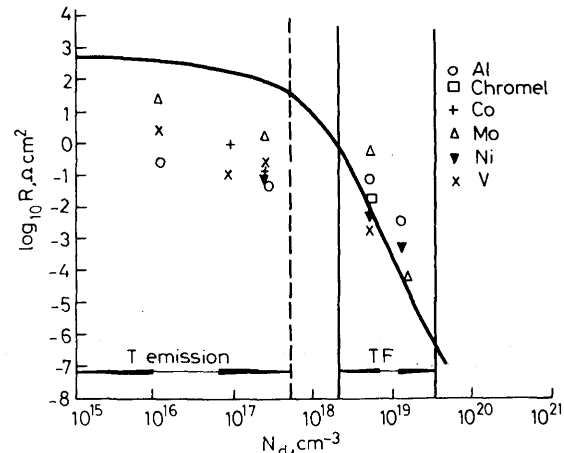

Fig. 7 *Specific contact resistance as function of donor density for contacts to n-type silicon [33]*

3.3 Recombination in depletion region

The importance of recombination in the depletion region has been convincingly demonstrated in a classic paper by Yu and Snow [34]. The current density due to recombination via a Shockley-Read centre near the middle of the gap is given approximately by

$$J = J_r \exp(qV/2kT) \quad (14)$$

where $J_r = qn_i w/\tau$. Here n_i is the intrinsic concentration, which is proportional to $\exp(-qE_g/2kT)$, w is the thickness of the depletion region, and τ the lifetime within the depletion region.

The relative importance of thermionic emission and of recombination in the depletion region depends on ϕ_b, E_g, T and τ. The recombination component is likely to be relatively

more important in high barriers, in material of low lifetime, at low temperature, and at low bias voltage. It is much more important in GaAs than in Si diodes. If recombination current is important, the temperature variation of the forward current shows two activation energies. Above room temperature the activation energy tends to the value ϕ_b, characteristic of the thermionic-emission component, whereas below room temperature it approaches the value $E_g/2$ characteristic of the recombination current.

Recombination current may cause deviations of n from unity and of the pre-exponential term from the ideal value $A^{**}T^2 \exp(-q\phi_b/kT)$. The importance of recombination current in causing small departures from ideal behaviour has been frequently overlooked in the literature. Such departures become more pronounced at low temperatures.

3.4 Hole injection

Hole injection in forward biased Schottky diodes made from n-type semiconductors has been considered theoretically by Scharfetter [35]. In essence, the electron current (assuming pure thermionic emission) is proportional to $\exp(-q\phi_{bn}/kT)$, and the hole current (assuming a horizontal hole quasi-Fermi level) is proportional to $\exp(-q\phi_h/kT)$, where $\phi_h = E_g - \xi$ is the difference in energy between the Fermi level and the top of the valence band in the bulk semiconductor. The injection ratio $J_h/J_e\ (=\gamma)$ is therefore proportional to $\exp\{-q(\phi_h - \phi_{bn})/kT\}$, and the constant of proportionality is much less than unity. The hole barrier ϕ_h normally exceeds ϕ_{bn}, and so the injection ratio is very small, typically of the order of 10^{-4} for Schottky barriers on Si. Scharfetter also shows that there is a critical current density (of the order of $10^4\ \mathrm{Am}^{-2}$) above which γ increases linearly with current as a result of the electric field in the quasi-neutral region. A more detailed numerical analysis suggests that this increase in γ is limited by the rate of supply of holes from the contact.

Schaffetter's predictions have been confirmed experimentally by Yu and Snow for Au-Si diodes. It is reasonable, therefore, to regard hole injection as negligible for low-current operation, as in the case of low-power switching diodes, but in power rectifiers the injection ratio may become appreciable, and can lead to conductivity modulation [36]. Hole injection is also important when Schottky diodes are used under 'punch-through' conditions, as in BARITT diodes [37].

3.5 Reverse characteristics

According to the thermionic-emission theory, the reverse current density of an ideal Schottky diode should saturate at the value $J = A^{**}T^2 \exp(-q\phi_b/kT)$. There are several causes of departure from this ideal behaviour.

3.5.1 Field dependence of barrier height: If for any reason ϕ_b is dependent on the electric field strength in the barrier, the reverse characteristics will not show saturation. There are several possible mechanisms, all of which predict that ϕ_b should be a decreasing function of $\mathscr{E}_{max}$, the maximum field strength in the barrier. Since $\mathscr{E}_{max}$ increases with reverse bias V_r, it follows that ϕ_b decreases with increasing V_r, and the current does not saturate, but increases in proportion to $\exp(\Delta\phi_b/kT)$, where $\Delta\phi_b$ is the barrier lowering due to the field.

The simplest form of barrier lowering is that due to the image force. Arizumi and Hirose [38] have described silicon Schottky diodes in which the reverse characteristics can be completely explained in terms of image-force lowering, and we must regard these diodes as almost ideal.

More usually, the barrier lowering which is necessary to explain the lack of saturation is considerably in excess of that due to the image force, and Andrews and Lepselter [39] have invoked an empirical dependence of the form $\Delta\phi_b = \alpha\,\mathscr{E}_{max}$ to explain their results on metal-silicide Schottky diodes, with experimentally determined values of α in the range 1.5–3.5 nm. Coe [40] has reported results on Al-Si diodes made from 25 Ω-cm silicon, which show reverse characteristics consistent with this simple model up to reverse voltages of 1 kV. The $\alpha\,\mathscr{E}_{max}$ term may be due to the presence of an interfacial layer, or to interdiffusion of the metal and semiconductor which can produce an effect equivalent to an interfacial layer.

3.5.2 Effect of tunnelling: Tunnelling through the barrier becomes significant at lower doping levels and at higher temperatures in the reverse direction than in the forward direction because the application of moderately high reverse bias causes the potential barrier to become thin enough for electrons in the metal to tunnel into the semiconductor at energies below the top of the barrier. At room temperature, tunnelling causes departures from pure thermionic emission at reverse biases of 3–4 V for $N_d > 10^{23}\ \mathrm{cm}^{-3}$ in the case of Si, and for $N_d > 2 \times 10^{22}\ \mathrm{cm}^{-3}$ in the case of GaAs. The corresponding figures for low forward bias are $N_d > 5 \times 10^{23}\ \mathrm{cm}^{-3}$ for Si and $N_d > 10^{23}$ for GaAs.

At still higher concentrations field emission may appear at room temperature. At low bias voltages, the reverse current may actually exceed the forward current, as was originally predicted by Wilson [28].

Tunnelling is one of the most common causes of 'soft' reverse characteristics. It is particularly important near to the edge of the metal contact because the distortion of the field can cause a big increase in field strength, resulting in a decrease in the barrier width. This effect is accentuated if the surface of the semiconductor adjacent to the metal is accumulated due to the presence of surface charges, because this makes the barrier even thinner [41]. Edge effects can be minimised by using a surface preparation which does not cause accumulation [42], and almost totally eliminated by using a diffused p-type guard ring [43].

3.5.3 Generation in depletion region: If the effect of tunnelling is reduced to negligible proportions, there may be an appreciable reverse current due to the generation of electron-hole pairs in the depletion region. This process is the inverse of mechanism (c) of Section 3, and gives rise to a current component $J_g = qn_iw/\tau$, where w is the width of the depletion region and τ the lifetime. J_g increases with reverse bias because w is proportional to $(V_{do} + V_r)^{1/2}$. Like the inverse process, generation current is most important in high barriers and in low-lifetime materials like gallium arsenide, and is more pronounced at low temperatures than at high because it has a lower activation energy than the thermionic-emission component. It is a common cause of lack of saturation of the reverse current in gallium arsenide Schottky diodes.

4 Capacitance of Schottky barrier

Like a p-n junction, a Schottky barrier exhibits capacitance. With one exception, the capacitance of a Schottky barrier in an n-type semiconductor is virtually identical with that of an abrupt p^+-n junction, since the p^+ side of the latter behaves in many respects like a metal. The exception concerns the diffusion capacitance. If an alternating voltage (superimposed on a DC bias) is applied to a p^+-n junction, the injection of holes into the n side during the forward half-cycle is followed by extraction during the negative half-cycle, provided the AC component swings the total voltage into forward bias, and this contributes a capacitance – the so-called diffusion capacitance. The diffusion capacitance is very closely linked with the

GF Exhibit 1023 - 7/14

phenomenon of minority carrier storage, and as we shall see in Section 5, there is no minority carrier storage in a Schottky diode, and so there is no diffusion capacitance either.

4.1 Depletion capacitance

The capacitance of a Schottky barrier is associated with its depletion region, which in some respects resembles a parallel plate capacitor the separation between whose plates increases if reverse bias is applied. The capacitance is usually measured by superimposing a small alternating voltage on a reverse DC bias which yields the differential capacitance $C(= dQ/dV)$. In the absence of surface states there are three sources of charge in a Schottky barrier in an n-type semiconductor: the charge due to the uncompensated donors Q_d, the charge on the surface of the metal Q_m, and the charge due to any holes which may exist in the semiconductor immediately adjacent to the metal Q_h. (All charges refer to unit area.) The hole charge will not be negligible if the barrier height is significantly greater than one half of the bandgap, since in this case the surface of the semiconductor will be p-type. Because there is no electric field in the bulk of the semiconductor or in the metal, electrical neutrality requires that $Q_d + Q_m + Q_h = 0$. If the reverse bias is increased, electrons move into the bulk of the semiconductor so as to increase Q_d, and holes move into the metal so as to reduce Q_h. The holes 'belong' to the metal, so to speak, and the differential capacitance per unit area can be calculated either from $C = \partial Q_d/\partial V_r$ or from $C = -\partial(Q_h + Q_m)/\partial V_r$. According to the depletion approximation (see Appendix) the charge due to the donors is given by eqn. 18 as

$$Q_d = (2\epsilon_s q N_d V_d)^{1/2}$$

where V_d is the diffusion potential under the application of a reverse bias V_r. V_d is equal to $V_{d0} + V_r$, where V_{d0} is the diffusion potential at zero bias. Hence

$$C = \frac{\partial Q_d}{\partial V_r} = \frac{\partial Q_d}{\partial V_d} = \left(\frac{\epsilon_s q N_d}{2V_d}\right)^{1/2} \tag{15}$$

A more accurate calculation, which takes into account the transition region where the electron concentration falls from a value equal to N_d to a value negligible compared with N_d, gives the slightly different result

$$C = \left[\frac{\epsilon_s q N_d}{2\{V_d - (kT/q)\}}\right]^{1/2}$$

$$= \left[\frac{\epsilon_s q N_d}{2\{\phi_b - \xi + V_r - (kT/q)\}}\right]^{1/2} \tag{15a}$$

since $V_{d0} = \phi_b - \xi$. A plot of C^{-2} against V_r should therefore be a straight line with a slope of $2/\epsilon_s q N_d$ and an intercept $-V_0$ on the voltage axis equal to $-(V_{d0} - kT/q)$. Hence $V_{d0} = V_0 + (kT/q)$, and $\phi_b = V_0 + \xi + (kT/q)$. This gives a valuable method for measuring ϕ_b and N_d. The bias dependence of a reverse-biased Schottky barrier can also be exploited as a means of achieving a voltage-controlled variable capacitance.

If N_d is not constant, the graph of C^{-2} against V_r is not linear, but the slope at any point is still given by $2/\epsilon_s q N_d$, where N_d is now the donor density at the edge of the depletion region. The width of the depletion region w can be obtained from $C = \epsilon_s/w$, where C refers to unit area, and so N_d can be found as a function of w. This forms the basis of a very convenient method of measuring impurity distributions in semiconductors, and there are several commercial instruments on the market that rely on this principle [44]. One technique that is widely used relies on contact between the semiconductor and a column of mercury in a capillary tube [67]. This method avoids the need for an evaporated contact, is simple to apply, and does not interfere with the surface of the semiconductor.

It must be appreciated that when the reverse bias is changed, it is the conduction electrons, not the donor ions, that move and eqn. 15 assumes that the conduction electron density n is equal to the donor density N_d. This is strictly true only if N_d is constant throughout the semiconductor. If N_d varies with position, the space charge within the semiconductor will not be zero, and n will not be equal to N_d. The difference between n and N_d will be particularly significant if N_d varies appreciably over a distance equal to the Debye length $L_D = (\epsilon_s kT/q^2 n)^{1/2}$. Hence abrupt variations in N_d over a distance L_D cannot be easily detected by the capacitance-profiling technique, and so L_D sets a limit to the spatial resolution of the method [45, 46].

A fairly thick interfacial layer may have a significant effect on the capacitance, because of the potential drop in the layer. In the presence of an interfacial layer, the value of the barrier height obtained from C/V measurements is equal to the 'flat-band' barrier height ϕ_b^0 which occurs under the application of a forward bias sufficient to eliminate the band-bending in the semiconductor. This result can easily be understood because the intercept of the C^{-2} against V_r graph on the voltage axis correponds to an infinite value of C, i.e. to the disappearance of the depletion region, which occurs when the bands are flat. This result is true whether or not surface states are present. Hence if the interfacial layer is thick enough to cause a significant difference between ϕ_b^0 and the zero-bias barrier height ϕ_{b0}, there will be a discrepancy between the barrier heights measured by the C/V and I/V methods, since it was seen in Section 3.1.2 that the latter measures ϕ_{b0}.

4.2 Effect of deep traps

Sometimes the depletion region of a Schottky barrier may contain traps associated with crystal defects, or with impurities other than the donors or acceptors with which the semiconductor is intentionally doped. These traps usually have energy levels which are near to the middle of the bandgap, and are therefore called 'deep levels', in contrast with the shallow levels associated with the ordinary donors or acceptors. These deep traps may have time constants for capturing or emitting electrons which are as long as minutes or even tens of minutes. The occupation of the traps can be changed by external stimuli, such as light or changes of temperature, or by a change in bias. This change in occupation results in a change in the space-charge density, and consequently in the depletion width, which shows up as a change in capacitance. If the capacitance is measured as a function of reverse bias, the effect of the traps is quite complicated, and depends on the relative magnitudes of the reciprocals of their time constants, the frequency of the measuring signal, and the rate of change of the DC bias. In the simplest case, where the trap time constants are so short that their occupancy can follow almost instantaneously both the measuring signal and the bias variation, the effect is to substitute $N_d + N_t$ for N_d in the expression for C (eqn. 15). However, whereas N_d is the donor density at the edge of the depletion region, N_t is the trap density at the position where the trap energy coincides with the Fermi level in the interior of the semiconductor, and so the C/V profiling data do not have a simple interpretation.

The general case where the inverse of the trap time constant is less than the frequency of the measuring signal is quite complex and not easily interpreted; a discussion will be found in Reference 2. If the agency responsible for changing the occupancy of the traps is altered abruptly, e.g. by switching on

GF Exhibit 1023 - 8/14

a light source or by suddenly changing the applied bias, the resulting change in capacitance is transient in nature, and observaton of the time and temperature dependence of the capacitance can yield not only the time constant of the traps, but also their concentration and energy relative to the band edges. These techniques form the basis of a very important branch of deep-level studies known as deep-level transient spectroscopy (DLTS) [47]. The use of Schottky barriers for semiconductor assessment has been reviewed by Heime [48].

5 Transient behaviour

Each steady-state condition of a Schottky barrier, corresponding to a particular value of applied bias, is characterised by a particular charge distribution, and the high-frequency performance of Schottky barrier devices depends on the time taken for the charge distribution to adjust when the applied voltage changes. In the case of a detector or mixer we are concerned with the response of the diode when the applied voltage is switched from the forward to the reverse direction, whereas in a varactor we are concerned with small changes in reverse bias. We shall suppose there are no deep traps, and so the transient behaviour is determined entirely by the mobile carriers.

In a *p-n* junction, minority carriers injected across the junction in forward bias must be removed, after the bias is reversed, before the current can reach a new steady-state value, and it is this process which controls the high-frequency limit of operation of a mixer or detector. A Schottky barrier in forward bias contains two kinds of minority carriers. First, there are electrons injected from the conduction band of the semiconductor into the metal, where their energy initially exceeds the metal Fermi energy by an amount equal to the barrier height. These electrons are often referred to as 'hot' electrons. If the applied bias were suddenly reversed these electrons could be pulled back into the semiconductor, giving an additional component of reverse current, analogous to that due to the injection of electrons into the *p* side of a *p-n* junction. However, the duration of this current is very short, because hot electrons in a metal are strongly scattered by electron-electron collisions for which the mean free path is about 50 nm and the mean free time about 10^{-14} s. At the first collision, each hot electron loses, on average, about half its energy, so that it will no longer be able to surmount the barrier and return to the semiconductor, and so hot-electron storage in the metal is not an important factor in the response time of a Schottky barrier.

Next, we consider the holes injected into the semiconductor from the metal. As we have seen in Section 3.4 they constitute a fraction γ of the total current – typically 10^{-4} at low current densities – and the recombination time of these holes in *n*-type silicon τ_{rh}, might be about 10^{-6} s. The effect of these holes on the transient response of the device can be estimated by the following semiquantitative argument. The injected hole charge per unit area can be shown to be equal to $\gamma J \tau_{rh}$, where J is the total current density crossing the junction, and the effective recovery time when the bias is reversed is equal to the stored charge divided by the total current, which we shall assume to be limited by the impedance of the external circuit and equal to the forward current, in which case the recovery time will be equal to $\gamma\tau_{rh}$. For the data given above we therefore have a recovery time of about 10^{-10} s, which is clearly longer than that due to hot-electron injection, although it could be reduced by reducing either τ_{rh} or γ. In this connection it is important to notice that γ will be much smaller for gallium arsenide than for silicon because of the larger energy gap of GaAs, and τ_{rh} will also be smaller.

Finally, we must consider the charge motion associated with the charging and discharging of the depletion layer capacitance, which determines the frequency response of a varactor diode. If the reverse bias is increased, electrons will recede into the bulk of the semiconductor with a consequent widening of the depletion layer. It is assumed that the barrier is formed in epitaxial material grown on a high-conductivity substrate, so that the series resistance R_s is due to that part of the epilayer under the Schottky contact. Let C be the depletion layer capacitance and R_p the parallel resistance associated with the reverse current of the Schottky diode. When the reverse bias is changed, C charges through R_s and R_p in parallel, and the equivalent circuit is as shown in Fig. 8.

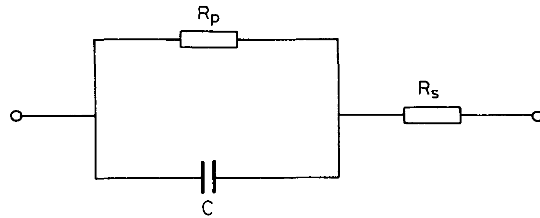

Fig. 8 *Equivalent circuit of Schottky diode*

Now R_p is a function of applied voltage and may vary from around 100 Ω to 10^7 Ω, but in almost all circumstances it is much greater than R_s and may therefore be neglected.

We can then write

$$R_s \simeq \frac{\rho d}{S} = \frac{d}{qn\mu_n S} = \frac{d}{qN_d\mu_n S}$$

where d is the epilayer thickness, ρ its resistivity, S is the area of the junction, and

$$C = \left\{\frac{qN_d\epsilon_s}{2(V_{d0} + V_r)}\right\}^{1/2} S$$

where V_{d0} is the diffusion potential at zero bias, and V_r the applied reverse bias (eqn. 15).

The time constant CR_s is a function of voltage, through V_r, but we can immediately see from

$$CR_s = \frac{d}{\mu_n}\left\{\frac{\epsilon_s}{2qN_d(V_{d0} + V_r)}\right\}^{1/2}$$

that $\mu_n N_d^{1/2}$ should be as large as possible and can be used as a figure of merit when choosing a semiconductor for high-frequency Schottky barrier manufacture. In a typical example of a mixer diode made with 5×10^{-4} Ωm silicon, having $N_d = 3 \times 10^{23}$ m^{-3}, $\mu_n = 0.05$ m^2 V^{-1} s^{-1}, $\epsilon_s = 10^{-10}$ Fm^{-1}, $d = 10^{-6}$ m, and $(V_{d0} + V_r) = 0.2$ V.

$$CR_s \simeq 1.7 \times 10^{-12} \text{ s}$$

Thus we see that the time constants associated with all three processes can be sufficiently short for microwave operation.

6 Practical contacts

6.1 General considerations

In practice the metal-semiconductor contact must be considered as an integral part of a device structure. In addition to having the desired electrical properties it must be possible to bond wires to it, and the whole structure must be mechanically and chemically stable during bonding and encapsulation, and during subsequent use. Clearly this puts a number of additional constraints on the choice of metal in a Schottky barrier, and leads to one of two results. It may be necessary to choose a particular metal to obtain the desired

GF Exhibit 1023 - 9/14

electrical characreristics, in which case some limitation of operating conditions or device life will have to be accepted. Alternatively, it may be possible to obtain both the required electrical characteristics and the necessary reliability by more elaborate device processing. We shall now examine in detail some possible contact systems.

6.2 Effects of heat treatment

Most contacts used in semiconductor devices are subjected to heat treatment. This may be deliberate, to promote adhesion of the metal to the semiconductor, or unavoidable, because high temperatures are needed for other processing stages which occur after the metal is deposited. It is important to avoid the melting of rectifying contacts, because if this happens the interface may become markedly nonplanar, with sharp metallic spikes projecting into the semiconductor. When this occurs, tunnelling through the high-field region at the tip of the spike may severely degrade the electrical characteristics [49]. Unless alloying of the contact is desired (e.g. in the formation of ohmic contacts), it is necessary to keep the temperatures to which contacts are subjected below the eutectic temperature of the metal-semiconductor system. For example, the eutectic temperatures of alloys of silicon with the three common contact metals gold, aluminium, and silver, are, respectively, 370°C, 577°C, and 840°C.

Even at temperatures substantially below the eutectic temperature (as low as 200°C in the case of Au-Si contacts), migration of the semiconductor through the metal may occur. As a result of the migration, the electrical characteristics become very nonideal. The change in I/V characteristic cannot normally be explained simply in terms of a change in barrier height, but as a rule the whole shape of the characteristic alters to such an extent that it is clear that we no longer have a simple Schottky barrier. In some cases the characteristic can be explained by supposing that atoms which behave as donors or acceptors diffuse into the semiconductor, or electrically active defects are created, so that the effective density of dopant in the semiconductor is changed. If the dopant density increases, the barrier gets thinner and thermionic-field emission (or even field emission) may occur. If the diffusing atoms or defects introduced into the semiconductor have the opposite polarity to that of the original dopant, the effective dopant density decreases, and it occasionally happens that a *p-n* junction may form. A good example of this is the case of aluminium-silicon contacts which have been extensively studied because of their technological importance. Chino [50] first reported that Al-Si contacts heated above 450°C showed a significant change in I/V characteristics which, in the case of *n*-type silicon, could be described in terms of increases in the barrier height and ideality factor. This change in characteristics was accompanied by pronounced pitting of the silicon surface. Basterfield, Shannon and Gill [51] have given a convincing explanation of these observations in the following terms. At temperatures around 500°C, silicon is taken up into solid solution by the aluminium, to an amount determined by the solubility limit at the particular temperature. On cooling, the dissolved silicon recrystallises onto the *n*-type silicon as an aluminium-doped *p*-type layer, because aluminium is an acceptor. The concentration of aluminium in the recrystallised silicon at 500°C is about 5×10^{24} m^{-3}. The net space-charge density is therefore negative near the interface, and the bands become bent downwards as shown in Fig. 9. If the distance from the maximum of the barrier to the interface is less than the electron mean free path, the structure behaves like a Schottky barrier whose height is ϕ_b'. The aluminium concentration quoted above is of the right order of magnitude to cause an increase in barrier height of about 0.1 eV, which is what was observed in practice.

Another metal semiconductor combination which has been extensively investigated because of its technological importance is gold on gallium arsenide. The results of these investigations are that above 350°C there is extensive migration of gallium into the gold, up to a limiting concentration set by the solid solubility, followed by a rapid increase in the gallium concentration at the Au-GaAs eutectic temperature of 450°C. At this temperature there is also substantial migration of gold into the gallium arsenide. The arsenic remains relatively stationary up to 450°C, but diffuses rapidly into the gold as the temperature approaches 500°C. The diffusion of gallium into the gold is accompanied by the formation of Ga vacancies in the gallium arsenide. Above 400°C there is a considerable degradation of the I/V characteristics, which, before heat treatment, are those of an almost ideal rectifying contact with a barrier height of about 0.9 eV. The effect of heat treatment is to increase the current density and ideality factor in a manner that suggests the onset of thermionic-field emission [52]. This could be explained if there were a significant increase in donor density in that part of the depletion region closest to the metal (see Section 3.2), though both gallium vacancies and gold atoms are believed to act as acceptors in gallium arsenide. (It is possible, however, that gold might be either an acceptor or a donor in gallium arsenide, as it is in silicon).

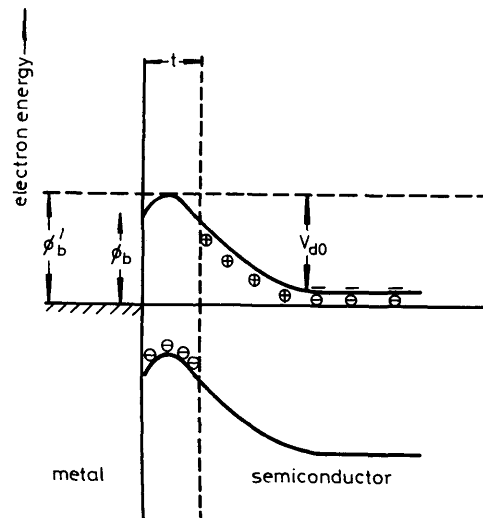

Fig. 9 *Band diagram for aluminium-silicon contact after heat treatment [51] (Basterfield, Shannon and Gill)*

6.3 Silicides

The effect of heat on metal-silicon contacts is particularly important if the metal is capable of forming a silicide which is a stoichiometric compound. Most metals, including all the transition metals, form silicides after appropriate heat treatment. These silicides may form as a result of solid-state reactions at temperatures of about one-third to one-half the melting point of the silicide in degrees Kelvin [53]. The vast majority of silicides exhibit metallic conductivity so that, if a metallic silicide is formed as a result of heat treatment of a metal-silicon contact, the silicide-silicon junction behaves like a metal-semiconductor contact, and may exhibit rectifying properties. Moreover, because the silicide-silicon interface is formed some distance below the original surface of the silicon it is free from contamination and very stable at room temperature; contacts formed in this way generally show stable electrical characteristics which are very close to ideal [43, 54]. They also exhibit very good mechanical adhesion.

Silicides that are particularly useful for practical contacts include those of platinum and nickel. The latter forms Ni_2Si, NiSi, and finally $NiSi_2$ as the reaction temperature is varied

GF Exhibit 1023 - 10/14

from 230°C to 850°C [55]. The barrier height remains remarkably constant at 0.70 ± 0.01 eV, which is indistinguishable from that of the unreacted Ni-Si. This strongly suggests that the barrier height is determined by a thin transition layer, about 10 Å or so thick, which is always present adjacent to the silicon, no matter what temperature is used for the heat treatment.

6.4 *Ohmic contacts*

An ohmic contact is one for which the I/V characteristic is determined by the resistivity of the semiconductor specimen or by the behaviour of the device of which the contact forms part, rather than by the characteristics of the contact. It is not essential that the I/V characteristic of the contact itself is linear, provided its resistance is very small compared with the resistance of the specimen or device. In addition, the contact should not inject minority carriers and should be stable both electrically and mechanically.

An important property of an ohmic contact is its specific resistance (resistance multiplied by area). A good ohmic contact should have a specific resistance of less than 10^{-9} Ωm^2. The specific resistance R_e can be found by measuring the total resistance of a circular contact of diameter d on a slice of semiconductor of thickness t and resistivity ρ. The total resistance R_{tot} is given [56] by

$$R_{tot} = \frac{\rho}{\pi d} \tan^{-1}\left(\frac{4t}{d}\right) + \frac{4R_c}{\pi d^2} + R_0$$

where R_0 is the resistance of the back contact.

The fabrication of ohmic contacts is still more of an art than a science, and every laboratory tends to have its own favourite recipes which involve particular metal or alloy systems, particular deposition methods, and particular forms of heat treatment. A comprehensive list of recipes for ohmic contacts to germanium, silicon, III–V and II–VI compounds has been given by Milnes and Feucht [57], and to III–V compounds, specifically, by Rideout [58].

Most of the recipes appear to depend on one of the following three principles:

(*a*) If the semiconductor is one which conforms approximately to the simple Schottky-Mott theory (eqn. 2), it should be possible to create an ohmic contact by finding a metal with a work function less than the work function of an n-type semiconductor, or greater than the work function of a p-type semiconductor, as discussed in Section 2.1. Unfortunately, there are very few metal-semiconductor combinations with this property. If the inequality is nearly, but not quite, satisfied, the result should be a rectifying contact with a very low barrier height which in practice may serve effectively as an ohmic contact.

(*b*) The vast majority of ohmic contacts depend on the principle of having a thin layer of very heavily doped semiconductor immediately adjacen to the metal. The depletion region is then so thin that field emission takes place, and the contact has a very low resistance at zero bias. This mechanism has already been discussed in Section 3.2.

(*c*) If the surface of the semiconductor is damaged (e.g. by sand-blasting), crystal defects may be formed near the surface, which act as efficient recombination centres. If the density of these is high enough, recombination in the depletion region will become the dominant conduction mechanism, and will cause a significant decrease in the contact resistance.

The most widely used of these methods is the one described in (*b*). The heavily doped layer may be formed separately from the deposition of the metal, say by diffusion, by ion implantation [59], or by molecular beam epitaxy [60]. This may also result from the deposition and subsequent heat treatment of an alloy containing an element which acts as a donor or acceptor in the semiconductor. On heating, the semiconductor dissolves in the metal, and on subsequent cooling it crystallises out with a high concentration of the electrically active element in solid solution. For silicon and germanium, the alloys most commonly used are gold-antimony for n-type and gold-indium for p-type. For gallium arsenide, which has received much attention as far as ohmic contacts are concerned because of its microwave applications, the most common alloys are those of gold or silver with germanium or tin (for n-type material) or with manganese (for p-type material).

One of the most important properties of a suitable alloy for ohmic contacts is that it should 'wet' the surface of the semiconductor when it melts [61]. For instance, gold-germanium does not wet the surface of gallium arsenide very well, and tends to 'ball up' when heated. This can be overcome by depositing a thin film of nickel on top of the gold-germanium which has the effect of keeping the latter continuous. A metallurgical study of the Ni/Au-Ge/GaAs system has been made by Robinson [62] and by Wittmer *et al.* [63]. A review of ohmic contacts to gallium arsenide has been given by Yoder [64]. Sometimes alloys will form equally good ohmic contacts on either p- or n-type material; e.g. gold-germanium will work on both p- and n-type gallium arsenide because germanium is an amphoteric impurity (i.e. it can behave either as a donor or as an acceptor) in gallium arsenide.

7 Comparison between Schottky diodes and *p-n* junctions

Although the choice between Schottky diodes and p-n junctions for device applications is usually dictated by purely technological considerations, it sometimes happens that both of them are technologically feasible and the choice between them has then to be made by reference to their electrical properties. When this happens there are two major factors to take into account. The first concerns transient response, and this has been fully discussed in Section 5. The second, considered below, is the difference in current-transport mechanisms, which give rise to very different saturation current densities.

We shall compare a Schottky diode made from an n-type semiconductor with a p-n junction having the same barrier height. The Schottky diode has the property that the current is carried almost entirely by electrons, although the semiconductor may be only moderately doped, and to ensure that this is also true of the p-n junction we must assume that the latter is made with the p-side more lightly doped than the n-side. (By the barrier height of such a junction we mean the difference in energy between the bottom of the conduction band on the p-side and the Fermi level.)

It can be shown [65] that the ratio of the saturation current of the Schottky diode J_{0s} to the saturation current of the p-n junction J_{0pn} is given by

$$\frac{J_{0s}}{J_{0pn}} = \frac{(3r_e)^{1/2}}{4} \simeq r_e^{1/2}$$

where r_e is the ratio of the lifetime of electrons in the p-side τ_{re} to their mean time between collisions τ_{ce}.

For a silicon p-n junction, a typical value of τ_{re} is $\sim 10^{-6}$ s, while $\tau_{ce} \sim 10^{-13}$ s, and so $(r_e)^{1/2} > 10^3$. Hence, for the same barrier height, the saturation current density of a Schottky diode exceeds that for a silicon p-n junction by a factor of 10^3 or more. For a short-lifetime semiconductor like gallium arsenide the ratio is smaller but is still of the order of 10^2.

In practice, the barrier height of a Schottky diode is likely

GF Exhibit 1023 - 11/14

to be appreciably less than that of a *p-n* junction made from the same semiconductor. This fact, taken in conjunction with the argument of the preceding paragraph, has the result that the saturation current density of a Schottky diode is likely to exceed that for a *p-n* junction made from the same semiconductor by a factor of 10^7 or more.

Expressed in a different way, for the same current density, the forward voltage drop across a *p-n* junction will exceed that across a Schottky diode by at least 0.4 V. This makes a Schottky diode particularly suitable for use as a low-voltage high-current rectifier. Conversely, since the reverse saturation current is also larger, a Schottky diode is not as suitable as a *p-n* junction for use as a high-voltage low-current rectifier.

Another application for which the high value of saturation current in a Schottky diode is a disadvantage is the solar cell, where the large value of J_{0s} results in a low open-circuit voltage. This disadvantage can be partially overcome by interposing a thin insulating layer between the metal and the semiconductor so as to make a MIS (metal-insulator-semiconductor) structure [66]. The insulating layer is thin enough for electrons to be able to tunnel through it, and has the effect of reducing J_{0s} by a factor equal to the tunnelling probability.. On the other hand, the insulating layer has little effect on the photocurrent, since this involves the movement of holes towards the metal and the hole current is limited by diffusion rather than by the probability of tunnelling.

8 References

1 HENISCH, H.K.: 'Rectifying semiconductor contacts' (Clarendon Press, Oxford, 1957)

2 RHODERICK, E.H.: 'Metal-semiconductor contacts' (Clarendon Press, Oxford, 1977)

3 RIDEOUT, V.L.: 'A review of the theory, technology and applications of metal-semiconductor rectifiers', *Thin Solid Films*, 1978, **48**, pp. 261–291

4 NORTHROP, D.C., and RHODERICK, E.H.: 'Variable impedance devices', (John Wiley & Sons 1978) Chap. 2

5 RHODERICK, E.H.: 'The physics of Schottky barriers', *J. Phys. D: Appl. Phys.*, 1970, **3**, p. 1153

6 SCHOTTKY, W.: 'Theory of blocking layer and point rectifiers', *Zeit. f. Phys.*, 1939, **113**, pp. 367–414

7 MOTT, N.F.: 'Note on the contact between a metal and an insulator or semiconductor', *Proc. Camb. Phil. Soc.*, 1938, **34**, pp. 568–572

8 BARDEEN, J.: 'Surface states and rectification at a metal semiconductor contact', *Phys. Rev.*, 1947, **71**, pp. 717–727

9 COWLEY, A.M., and SZE, S.M.: 'Surface states and barrier height of metal-semiconductor systems', *J. Appl. Phys.*, 1965, **36**, pp. 3212–3220

10 SMITH, B.L., and RHODERICK, E.H.. 'Schottky barriers on *p*-type silicon', *Solid-State Electron.*, 1971, **14**, pp. 71–75

11 THANAILAKIS, A.: 'Contacts between simple metals and atomically clean silicon', *J. Phys. C.*, 1975, **8**, pp. 655–668

12 THANAILAKIS, A., and RASUL, A.: 'Transition-metal contacts to atomically clean silicon', *ibid.*, 1976, **9**, pp. 337–343

13 HEINE, V.: 'Theory of surface states', *Phys. Rev.* A, 1965, **138**, pp. 1689–1696

14 TEJEDOR, C., FLORES, F., and LOUIS, E.: 'The metal-semiconductor interfaces Si (111) and zincblende (110) junctions', *J. Phys. C*, 1977, **10**, pp. 2163–2117

15 ANDREWS, J.M., and PHILLIPS, J.C.: 'Chemical bonding and structure of metal-semiconductor interfaces', *Phys. Rev. Lett.*, 1975, **35**, pp. 56–59

16 HIRACKI, A.: 'A model for the mechanism of room temperature interfacial intermixing reactions in various metal-semiconductor couples', *J. Electrochem. Soc.*, 1980, **127**, pp. 2662–2665

17 SPICER, W.E., LINDAU, I., SKEATH, P., SU, C.Y., and CHYE, P.: 'Unified mechanism for Schottky-barrier formation and III–V oxide interface states', *Phys. Rev. Lett.*, 1980, **44**, pp. 420–423

18 WAGNER, C.: 'Theory of current rectifiers', *Phys. Z.*, 1931, **32**, pp. 641–645

19 SCHOTTKY, W., and SPENKE, E.: 'Quantitative treatment of the space-charge and boundary-layer theory of the crystal rectifier', *Wiss. Veroff. a.d. Siemens-Werken*, 1939, **18**, pp. 225–291

20 BETHE, H.A.: 'Theory of the boundary layer of crystal rectifiers'. MIT Radiation Laboratory Report 43–12, 1942

21 GOSSICK, B.R.: 'Metal-semiconductor rectifiers and transistors', *Solid-State Electron.*, 1963, **6**, pp. 445–452

22 SPENKE, E.: 'Electronic semiconductors' (McGraw-Hill, New York, 1958)

23 CROWELL, C.R.: 'The Richardson constant for thermionic emission in Schottky-barrier diodes', *Solid-State Electron.*, 1965, **8**, pp. 395–399

24 CROWELL, C.R. and SZE, S.M.: 'Current transport in metal-semiconductor barriers', *ibid.*, 1966, **9**, pp. 1035–1048

25 CROWELL, C.R., and BEGUWALA, M.: 'Recombination velocity effects on current diffusion and IMREF in Schottky barriers', *ibid.*, 1971, **14**, pp. 1149–1157

26 RHODERICK, E.H.: 'Transport processes in Schottky diodes'. Proceedings of Manchester conference on metal-semiconductor contacts (Institute of Physics, London, 1972), pp. 3–19

27 WRONSKI, C.R., CARLSON, D.E., and DANIEL, R.F.: 'Schottky-barrier characteristics of metal-amorphous-silicon diodes', *Appl. Phys. Lett.*, 1976, **29**, pp. 602–605

28 WILSON, A.H.: 'A Note on the theory of rectification', *Proc. R. Soc. London A*, 1932, **136**, pp. 487–498

29 PADOVANI, A., and STRATTON, R.: 'Field and thermionic-field emission in Schottky barriers', *Solid-State Electron.*, 1966, **9**, pp. 695–707

30 CROWELL, C.R., and RIDEOUT, V.L.: 'Normalized thermionic-field (TF) emission in metal-semiconductor (Schottky) barriers', *ibid.*, 1969, **12**, pp. 89–105

31 PADOVANI, F.A.: 'Semiconductors and semimetals' (Academic Press, New York, 1971), **6A**, Chap. 2

32 YU, A.Y.C.: 'Electron tunnelling and contact resistance of metal-silicon contact barriers', *Solid-State Electron.*, 1970, **13**, pp. 239–247

33 VILMS, J., and WANDINGER, L.: 'Ohmic contacts to semiconductors'. (Electrochemical Society, New York, 1969), pp. 31–47

34 YU, A.Y.C., and SNOW, E.H.: 'Minority carrier injection of metal-silicon contacts', *Solid-State Electron.*, 1969, **12**, pp. 155–160

35 SCHARFETTER, D.L.: 'Minority carrier injection and charge sorage in epitaxial Schottky barrier diodes', *ibid.*, 1965, **8**, pp. 299–311

36 JÄGER, H., and Kosak, W.: 'Modulation effect by intense hole injection in epitaxial silicon Schottky-barrier-diodes', *ibid.*, 1973, **16**, pp. 357–364

37 SZE, S.M. COLEMAN, D.J., and LOYA, A.: 'Curent transport in metal-semiconductor-metal (MSM) strucrures', *ibid.*, 1971, **14**, pp. 1209–1218

38 ARIZUMI, T., and HIROSE, M.: 'Transport properties of metal-silicon Schottky barriers', *Jpn. J. App. Phys.*, 1969, **8**, pp. 749–754

39 ANDREWS, J.M., and LEPSELTER, M.P.: 'Reverse current-voltage characteristics of metal-silicide Schottky diodes', *Solid-State Electron.*, 1970, **13**, pp. 1011–1023

40 COE, D.J.: 'Multiple guard-ring voltage Schottky barrier diodes', European semiconductor device research conference, Munich, 1971, paper 10.3

41 YU, A.Y.C., and SNOW, E.H.: 'Surface effects on metal-silicon contacts', *J. Appl. Phys.*, 1968, **39**, pp. 3008–3016

42 SMITH, B.L.: 'GaAs Schottky diodes with linear log I/V behaviour over eight decades of current ', *Electron. Lett.*, 1968, **4**, (16), pp. 332–333

43 LEPSELTER, M.P. and SZE, S.M.: 'Silicon Schottky barrier diode with near-ideal I/V characreristics', *Bell Syst. Tech. J.*, 1968, **47**, pp. 195–208

44 BAXANDALL, P.J., COLLIVER, D.J., and FRAY, A.F.: 'An instrument for the rapid determination of semiconductor impurity profiles', *J. Phys. E.: Sci. Inst.*, 1971, **4**, pp. 213–221

45 KENNEDY, D.P., MURLEY, P.C., and KLEINFELDER, W.: 'On the measurement of impurity atom distributions in silicon by the differential capacitance technique', *IBM J. Res. & Dev.*, 1968, **12**, pp. 399–409

46 NISHIDA, M.: 'Depletion approximation analysis of the differential capacitance-voltage characteristics of an MOS structure with non-uniformly doped semiconductors', *IEEE Trans.*, 1979, **ED-26**, pp. 1081–1085

47 LANG, D.V.: 'Deep level transient spectroscopy: A new method to characterize traps in semiconductors', *J. Appl. Phys.*, 1974, **45**, pp. 3023–3032

48 HEIME, K.: 'Nondestructive evaluation of semiconductor materials and devices', (Plenum Press, 1979), pp. 201–256

49 ANDREWS, J.M.: 'The role of the metal-semiconductor interface in silicon integrated circuit technology', *J. Vac. Sci. & Technol.*, 1974, **11**, pp. 972–984

50 CHINO, K.: 'Behaviour of Al-Si Schottky barrier diodes under heat treatment', *Solid-State Electron.*, 1973, **16**, pp. 119–121

GF Exhibit 1023 - 12/14

51 BASTERFIELD, J., SHANNON, J.M., and GILL, A.: 'The nature of barrier height variation in alloyed Al-Si Schottky barrier diodes', *ibid.*, 1975, **18**, pp. 290–291
52 MADAMS, C.J., MORGAN, D.V., and HOWES, M.J.: 'Outmigration of gallium from Au-GaAs interfaces', *Electron. Lett.*, 1975, **11**, (24), pp. 574–575
53 MAYER, J.W., and TU, K.N.: 'Analysis of thin-film structures with nuclear backscattering and X-ray diffraction', *J. Vac. Sci. & Technol.*, 1974, **11**, pp. 86–93
54 ANDREWS, J.M., and LEPSELTER, M.P.: 'Reverse current-voltage characteristics of metal-silicide Schottky diodes', *Solid-State Electron.*, 1970, **13**, pp. 1011–1023
55 COE, D.J.: 'A study of metal silicide-silicon Schottky barriers'. Ph.D. thesis, University of Manchester, 1974
56 COX, R.H., and STRACK, H.: 'Ohmic contacts for GaAs devices', *Solid-State Electron.*, 1967, **10**, pp. 1213–1218
57 MILNES, A.G., and FEUCHT, D.L.: 'Heterojunctions and metal-semiconductor junctions' (Academic Press, 1972)
58 RIDEOUT, V.L.: 'A review of the theory and technology for ohmic contacts to group III–V compound semiconductors', *Solid-State Electron.*, 1975, **18**, pp. 541–550
59 MEYER, D.E.: 'Ohmic contacts to semiconductors' (Electrochemical Society, New York, 1969), pp. 227–242
60 ZETTLEMOYER, A.C.: 'Ohmic contacts to semiconductors' (Electrochemical Society, New York, 1969), pp. 48–65
61 BARNES, P.A., and CHO, A.Y.: 'Nonalloyed ohmic contacts to *n*-GaAs by molecular beam epitaxy', *Appl. Phys. Lett.*, 1978, **33**, pp. 651–653
62 ROBINSON, G.Y.: 'Metallurgical and electrical properties of alloyed Ni/Au-Ge films on *n*-type GaAs', *Solid-State Electron.*, 1975, **18**, pp. 331–341
63 WITTMER, M., PRETORIUS, R., MAYER, J.W., and NICOLET, M.-A.: 'Investigation of the Au-Ge-Ni system used for alloyed contacts to GaAs', *ibid.*, 1977, **20**, pp. 433–439
64 YODER, M.N.: 'Ohmic contacts in GaAs', *ibid.*, 1980, **23**, pp. 117–119
65 RHODERICK, E.H.: 'Schottky diodes or *p-n* junctions?'. Proceedings of 5th European solid-state conference, 1976, pp. 103–118
66 PULFREY, D.L.: 'MIS solar cells. A Review', *IEEE Trans.*, 1978, **ED-25**, pp. 1308–1317
67 LEDERMAN, A.: 'Vacuum operated mercury probe for *CV* plotting and profiling', *Solid-State Technol.*, 1981, **24**, pp. 123–126

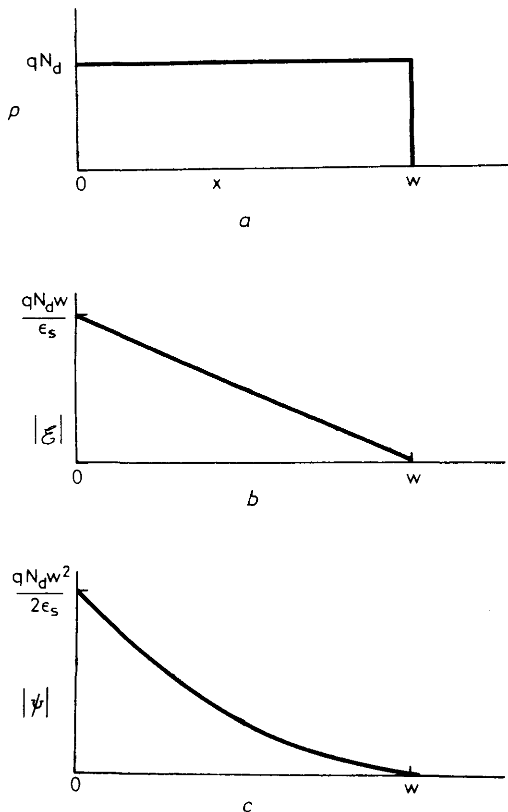

Fig. 10 *Variation of charge density, electric field strength, and electrostatic potential with distance according to depletion approximation*

a Charge density *b* Electric field strength
c Electrostatic potential

9 Appendix: Depletion approximation

It is frequently necessary to know how the electrostatic potential and electric field strength in a Schottky barrier depend on the barrier height, bias voltage, and impurity concentration, and for most purposes it is sufficiently accurate to use an approximation known as the depletion approximation. In this approximation the free-carrier density is assumed to fall abruptly from a value equal to the concentration in the bulk of the semiconductor to a value which is negligible compared with the donor or acceptor concentration, whereas in reality the transition occurs smoothly over the distance in which the bands bend by about $3\,kT/q$.

Let us consider the case of an n-type semiconductor. Outside the depletion region the semiconductor is neutral, and within the depletion region the charge density is equal to qN_d, so that the variation of charge density with distance is as shown in Fig. 10*a*. With no applied bias there can be no electric field within the neutral region of the semiconductor, otherwise a current would flow; and even when a bias voltage is applied, the current is usually so small in comparison with the conductivity of the semiconductor that the electric field in the neutral region is negligible.

The electric field is related to the charge density by Gauss's theorem, so that $\partial \mathscr{E}/\partial x = qN_d/\epsilon_s$. The magnitude of $\mathscr{E}$ (which is negative since it is directed in the negative x-direction) increases linearly as the metal is approached, rising from zero at $x = w$ to a value $qN_d w/\epsilon_s$ at the interface (Fig. 10*b*). If we take the electrostatic potential ψ to be zero within the neutral region, so that $\psi(w) = 0$, the potential at x is given by

$$\psi(x) = \int_x^w \mathscr{E}\,dx$$

$$= -\int_x^w \frac{qN_d(w-x)}{\epsilon_s}\,dx$$

$$= -\frac{qN_d}{2\epsilon_s}(w-x)^2$$

$\psi(x)$ is negative throughout the depletion region, and so the potential energy $-q\psi$ is positive and the bands are bent upwards. The magnitude of ψ rises quadratically as the metal is approached, and has the value $qN_d w^2/2\epsilon_s$ at the interface (Fig. 10*c*). This result can be easily remembered because the difference in potential across the depletion region is simply equal to the average electric field strength multiplied by w, and if $\mathscr{E}$ varies linearly the average field strength is equal to $\frac{1}{2}\mathscr{E}_{max} = \frac{1}{2} qN_d w/\epsilon_s$. The magnitude of ψ at the interface is equal to the diffusion potential V_d, and so

$$V_d = \frac{qN_d w^2}{2\epsilon_s} \tag{16}$$

Making use of the relationships

$$\mathscr{E}_{max} = \frac{qN_d w}{\epsilon_s} = \frac{Q_d}{\epsilon_s}$$

where Q_d is the total charge per unit area due to the

GF Exhibit 1023 - 13/14

uncompensated donors in the depletion region, we can write eqn. 16 in the alternative forms

$$V_d = \frac{\epsilon_s |\mathscr{E}_{max}|^2}{2qN_d} \tag{17}$$

and

$$V_d = \frac{Q_d^2}{2\epsilon_s q N_d} \tag{18}$$

According to eqn. 18, the differential capacitance per unit area is given by

$$C = \frac{\partial Q_d}{\partial V_d} = \left(\frac{\epsilon_s q N_d}{2V_d}\right)^{1/2} = \frac{\epsilon_s}{w}$$

Hence C is equal to the capacitance per unit area of a parallel-plate capacitor having a dielectric of permittivity ϵ_s and thickness w.

GF Exhibit 1023 - 14/14