Conductance-bias characteristics of metal-oxide-silicon tunnel junctions

1991 J. Phys.: Condens. Matter 3 2917

(http://iopscience.iop.org/0953-8984/3/17/009)

J. Phys.: Condens. Matter 3 (1991) 2917–2936. Printed in the UK

Conductance–bias characteristics of metal-oxide–silicon tunnel junctions

P M Owen†, W A Phillips† and G R Fisher‡
† Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK
‡ MEMC, 28100 Novara, Viale Gherzi 31, Casalla Postale 66, Italy

Received 20 September 1990, in final form 28 January 1991

Abstract. Thin oxide films on silicon and their interfaces, increasingly important to device physics, may be characterized using the transport properties of metal-oxide–silicon tunnel junctions. A successful numerical model of the transport properties of metal–insulator–semiconductor junctions has been developed for the first time. In particular, this model clarifies the effects of surface states, workfunction difference and doping density on the electrical characteristics. The effects of varying these parameters on the conductance–bias curves generated by the model are computed. The model predictions are in good agreement with experimental results presented in this paper from degenerate ⟨100⟩ n-type silicon–silicon oxide–metal structures, using a consistent set of parameters whose magnitudes are close to those found by previous workers. Possible applications of the model to the study of the silicon/oxide interface are outlined.

1. Introduction

Thin oxide films on silicon are becoming increasingly important to device physics as the required dimensions for gate oxides shrink. Information about these films and their interfaces cannot be obtained using the usual tools of surface science, developed for crystalline surfaces, or the standard techniques of MOS physics, developed for much thicker oxide layers. The transport properties of semiconductor–insulator–metal tunnel junctions contain information allowing the characterization of the silicon/oxide interface and the oxide during the initial stages of oxide growth. Extracting this commercially useful information from measurements of the transport properties involves the basic physics of the tunnelling process, in order to clarify the nature of the barrier and the effects, for example, of surface states and silicon doping density.

The conductance–bias characteristics of metal–insulator–metal junctions have been extensively measured and modelled, usually with the aim of deducing the shape of the potential barrier experienced by a tunnelling electron (see for example Brinkman *et al* 1970). In contrast, the conductance–bias characteristics of semiconductor–insulator–metal junctions (figure 1) have received little attention other than general qualitative statements (for silicon see for example Busmann *et al* 1985). Several simple analytical models, of rather uneven quality, have been proposed (for a review see Duke 1969), using various approximate methods, which provide some insight into the new features which arise when one metal is replaced by a semiconductor. However, no systematic investigations, either theoretical or experimental, have been performed to test the

0953-8984/91/172917 + 20 $03.50

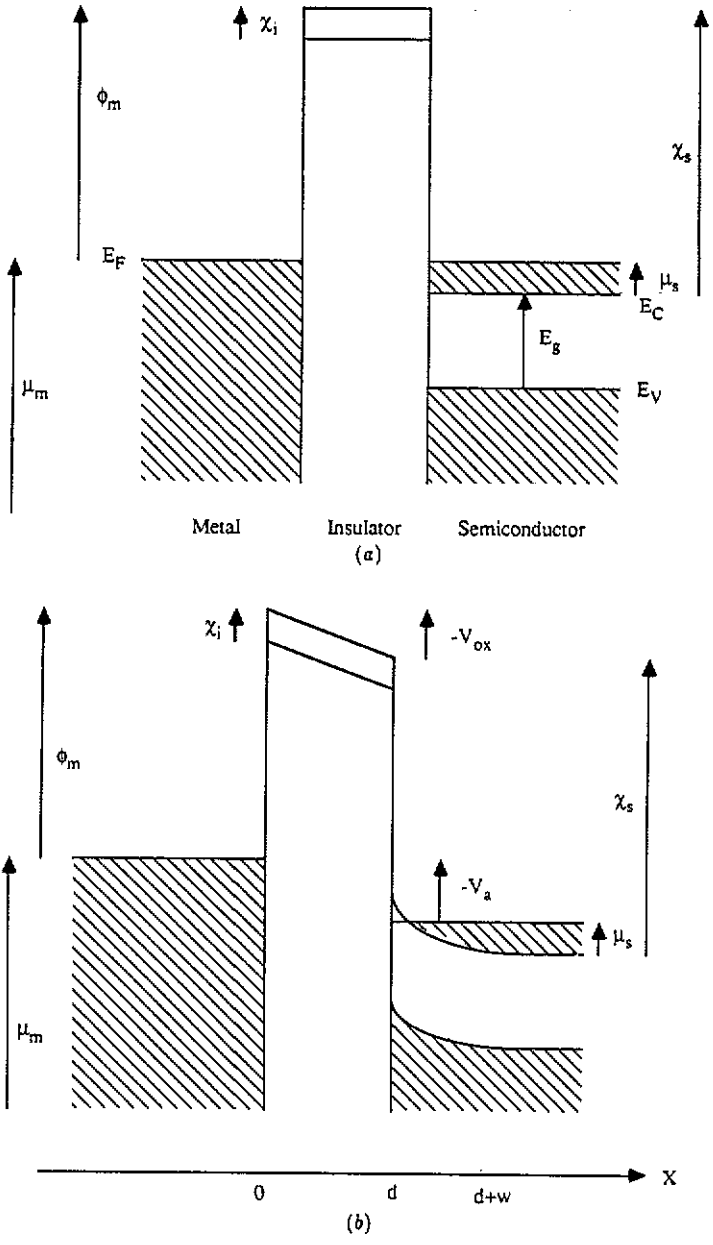

Figure 1. Energy diagram of a metal–insulator–semiconductor structure: (*a*) zero bias (*b*) negative bias. There is no workfunction difference or trapped charge. The figures are drawn for zero temperature.

applicability of the ideas developed for metal–insulator–metal junctions to these systems.

Our aim in this work was twofold: firstly, to measure the conductance–bias characteristics of n-type degenerate ⟨100⟩ silicon-oxide–metal tunnel junctions and to investigate carefully the effects of important parameters such as silicon doping density, metal

work function and surface-state density; and secondly, to develop a theoretical model of the transport properties which explains the effects observed in our experimental system, and also provides a way of characterizing and understanding other semiconductor systems. As well as contributing to the basis physics of tunnelling, this would provide a framework for studies of the initial stages of oxide growth on silicon.

The sample fabrication techniques developed to produce samples with reproducible electrical characteristics and the way in which junction parameters are varied in a controlled manner are presented in section 2. Measurement techniques are also described in this section. Section 3 describes the experiments and presents conductance–bias plots from a variety of junctions. The development of a theoretical model and the assumptions and approximations are decribed in section 4. Section 5 presents a careful comparison of the predictions of the model with the results from section 3.

2. Experimental technique

Silicon wafers (⟨100⟩ oriented, degenerately doped with phosphorus) supplied by MEMC were used as one electrode for the tunnel junctions. Residual oxide on the silicon surface was removed with a dilute solution of hydrofluoric acid, buffered with ammonium fluoride. The wafers were then thoroughly rinsed in water and a thin oxide grown by the techniques described in section 3. Metal dots (1.5 mm diameter, 200 nm thick) were deposited on top of the oxide in a thermal evaporator at 10^{-6} mbar to form a top electrode. Magnesium top electrodes were corrosion protected by evaporating aluminium on top, through the same mask. Ohmic contacts to the silicon were formed by removing the oxide from areas of the wafer adjacent to the metal dots and evaporating aluminium onto these areas. This procedure gave specific contact resistances of less than $10^{-1}\ \Omega\ \mathrm{cm}^2$ at 4.2 K even for the silicon with the lowest doping concentration used in this work.

Individual samples (junction and ohmic contact) were separated by cleaving using a diamond scriber and attached to the end of a dipstick for immersion in liquid helium. Fine copper wires were connected to the electrodes using a suspension of silver particles in an organic solvent. On warming gently (60 °C) with hot air, the solvent is evaporated leaving the silver to form a robust and reliable electrical contact which will tolerate thermal cycling between liquid helium and room temperatures. This procedure gave poor results for contacts to indium, and in this case a contact was formed by pressing the freshly cut surface of an indium pellet onto the wire and the electrode.

The conductance $\mathrm{d}I/\mathrm{d}V$ was measured using standard AC bridge techniques. The DC bias was set with a Keithley 230 voltage source. The bridge voltage was measured with a PAR 5301 lock-in amplifier which also produced the 0.8 mV modulation voltage at a frequency of 323 Hz. Data collection, processing and storage were controlled by an Acorn BBC microcomputer via an IEEE bus.

Two-terminal conductance measurements were made throughout this work. The conductance actually measured therefore is that of the series combination of the tunnel junction, the lead and contact resistances and the resistance of the silicon. By measuring the resistance between two adjacent ohmic contacts, the sum of the last three resistances was found to be around 2 Ω, with the bulk of this resistance contributed by the leads. This (constant) resistance can in principle be corrected for, but in view of the uncertainties involved, only data from junctions with zero-bias resistances greater than 100 Ω are presented. Four-terminal measurements, which would have eliminated these resistances, were not carried out because of the additional experimental difficulty of putting

four contacts on the samples, and because most of the data presented here are from samples having zero-bias resistances of around 1 kΩ.

The only conclusive test that the dominant conduction mechanism in these structures is tunnelling is the observation of an ideal superconducting structure near zero bias, for superconducting electrodes, below the appropriate transition temperature. For this reason, lead top electrodes, superconducting at 4.2 K, were normally employed and only data from samples showing such structure are presented here. A subsidiary test for tunnelling is to compare the conductance–bias characteristics of a junction measured at different temperatures. Tunnel currents are usually only weakly temperature dependent, in contrast to alternative conduction mechanisms such as thermal electron emission which are strongly enhanced by increasing temperature. The conductance as a function of bias was measured at several temperatures between 4.2 K and 200 K for several different samples, and the variation was always found to be small. Applying biases of several hundred millivolts at temperatures above 200 K resulted in changes to the conductance curves. For these reasons, all the measurements were performed at 4.2 K. Where no lead superconducting structure is visible in the plots in section 3 it has been magnetically quenched.

3. Experimental results

Various oxidation techniques were investigated in the early stages of this work, such as wet thermal oxidation, plasma oxidation and liquid-phase anodization. Oxidation in dry oxygen at room temperature or at 150 °C was found to give the most reproducible characteristics and this technique was adopted for the rest of this work. Optical measurements (Archer 1957) suggest that oxide layers of 1–2 nm thickness are produced. The conductance–bias curves from samples oxidized by these techniques, with lead top electrodes, are compared in figure 2. For both techniques the zero-bias resistance increased steadily with oxidation time, suggesting that conduction through any pin holes present in the oxide is not significant. Also shown in figure 2 is a characteristic from an aluminium–aluminium oxide–lead junction. The curves are all normalized to the conductance at +50 mV. These normalized plots are used throughout this work, rather than the absolute measured or computed values, as the actual tunnelling area (as opposed to the top electrode area) is not known and may vary from sample to sample. The +50 mV conductance has been chosen for normalization in preference to the zero-bias value to avoid rapidly varying superconducting structure. Throughout this paper, a positive bias corresponds to raising the semiconductor Fermi level with respect to the metal Fermi level. All the semiconductor curves show a large rise with increasing negative bias, a broad minimum at a positive bias of around the semiconductor Fermi degeneracy, and a rise of variable size for positive biases greater than this. In contrast, the aluminium curve is almost independent of bias on this scale.

During the development of the theory presented in the next section, it became apparent that important parameters for this system are the silicon/oxide interface state density, the semiconductor doping density and the work function of the top metal. These quantities were therefore varied experimentally and the effects are illustrated in figures 3–5. Figure 3 shows the effect of exposing the sample to a hydrogen plasma (0.1 mbar 5% H_2 in N_2, 400 V) for 45 s before the deposition of the lead top electrode, a treatment known to reduce the surface-state density. The normalized conductance change for both bias polarities is reduced. Exposing the oxide to a higher voltage plasma or for a longer time resulted in degraded superconducting structure. Figure 4 shows the effect of altering

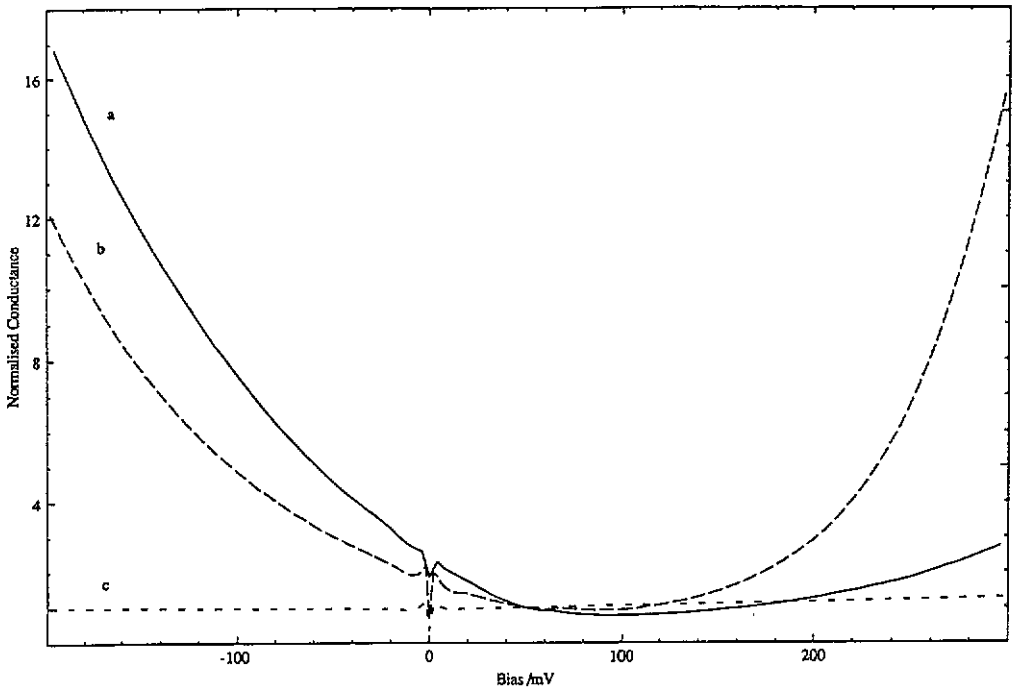

Figure 2. Comparison of characteristics from samples with different oxidation techniques and bottom electrodes. Curve a is for a silicon sample oxidized in dry oxygen at room temperature for 4 h; curve b is for a silicon sample oxidized in dry oxygen at 150 °C for 2 h 30 min; curve c is for an aluminium sample with a plasma oxide. All three samples have lead top electrodes and similar conductances at zero bias. The silicon doping density is 4×10^{25} m^{-3}.

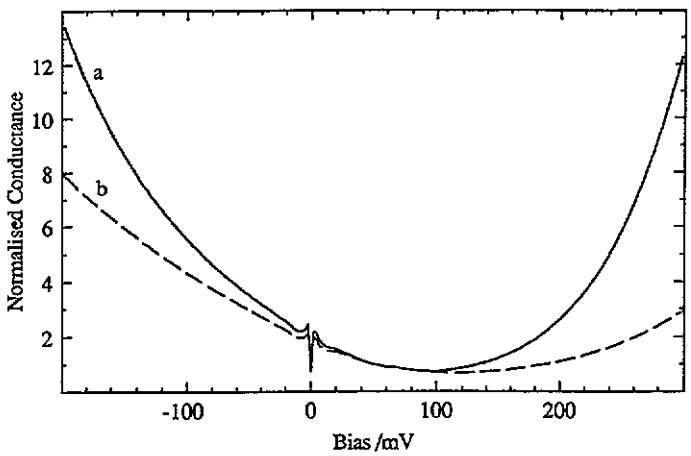

Figure 3. The effect of exposing a sample to a hydrogen plasma before lead top-electrode deposition. Curve a: without plasma; curve b: with plasma. The silicon doping density is 4×10^{25} m^{-3}.

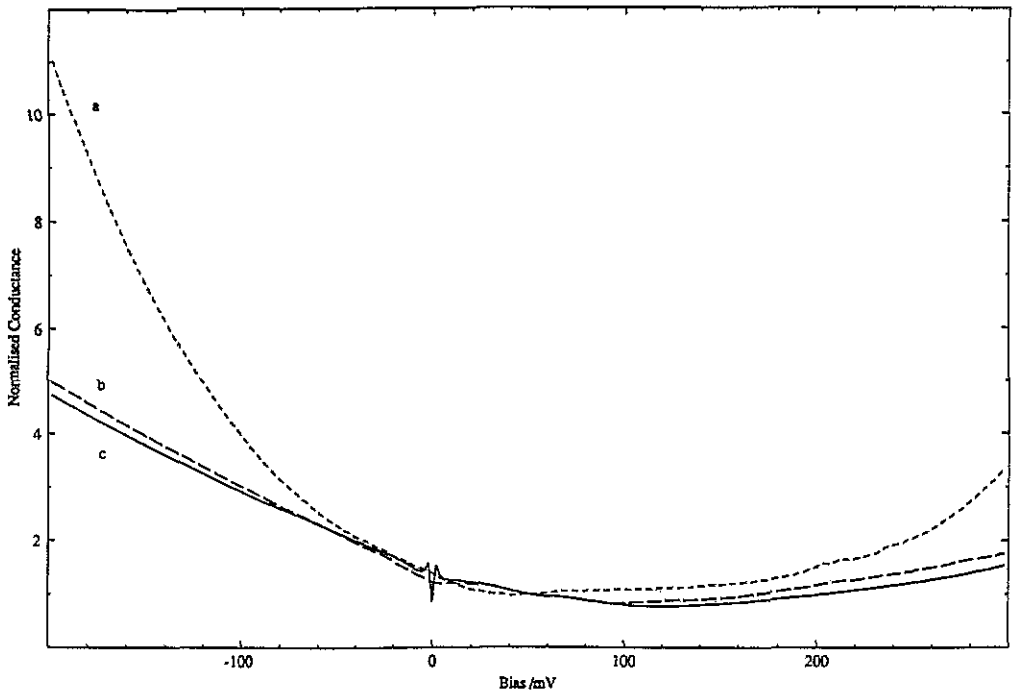

Figure 4. The effect of altering the semiconductor doping density. a: $1 \times 10^{25}\ m^{-3}$; b: $3 \times 10^{25}\ m^{-3}$; c: $4 \times 10^{25}\ m^{-3}$.

the semiconductor doping density, determined from measurements of the spreading resistance. Increasing the doping density decreases both the magnitude and the spatial extent of the band-bending in the silicon and reduces the variation in conductance with bias. Figure 5 compares identical junctions fabricated with different top electrode materials: figure 5(a) compares junctions with lead and indium top electrodes, using a 150 °C oxidation, whilst figure 5(b) compares samples with lead and magnesium top electrodes, having a room temperature oxidation. All other parameters, including oxidation time have been held constant for each pair of curves. The workfunctions for lead, magnesium and indium are quoted as 4.25 V, 3.66 V and 4.12 V, respectively (Weast 1988). In both cases, reducing the metal workfunction reduces the normalized conductance change for both bias polarities.

4. Theory

4.1. Introduction

There are three main differences between metal–insulator–metal and metal–insulator–(degenerate) semiconductor tunnel junctions (Duke 1969). Firstly, the energy gap in the semiconductor band structure must be considered in any explanation of the electrical properties. Consider an n-type degenerate semiconductor. For positive biases (raising the Fermi level of the semiconductor) greater than the semiconductor Fermi degeneracy, the metal Fermi level is opposite the energy gap. Further increase in the bias allows no new occupied extended states in the semiconductor to contribute to the current (figure

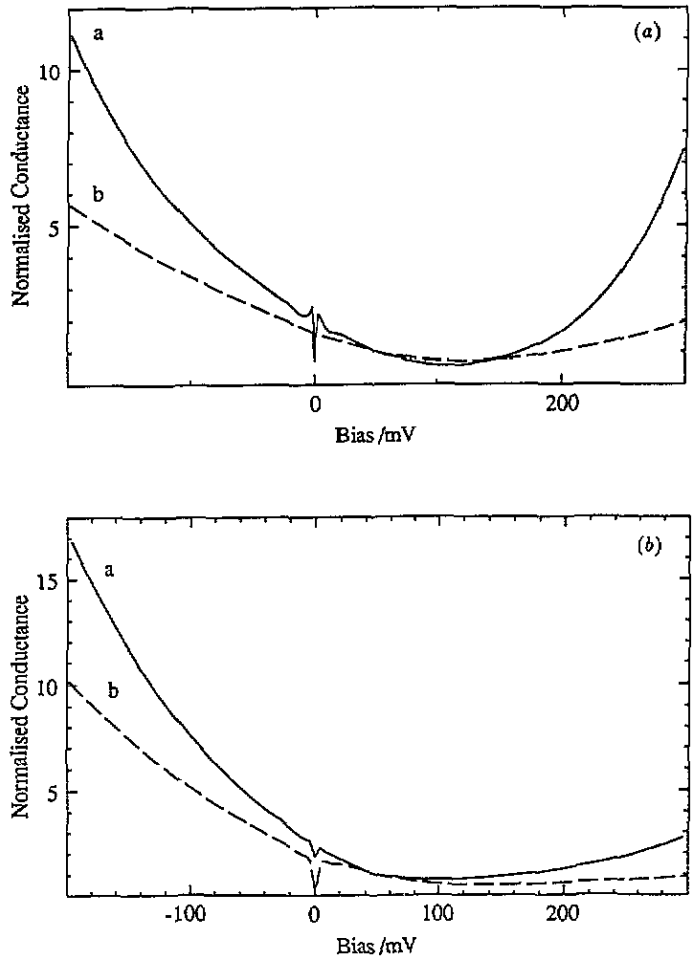

Figure 5. Comparison of samples with different top-electrode materials. (*a*) compares samples with a 150 °C oxidation with lead (curve a) and indium (curve b) top electrodes. (*b*) compares samples with a room temperature oxidation with lead (curve a) and magnesium (curve b) top electrodes. The silicon doping density is 4×10^{25} m^{-3}.

1(*a*)) although a large increase in bias will eventually allow tunnelling from extended states in the semiconductor valence band. However, the contribution from this process will be small because of the high barrier to tunnelling from these states. Secondly, within the band-bending approximation, varying the bias can change the energy of the conduction band relative to the Fermi level close to the semiconductor surface. Applying a large enough negative bias (or having a large workfunction difference or trapped charge density) can produce a depletion layer at the semiconductor surface (figure 1(*b*)). The tunnelling barrier is now the insulator in series with the depletion layer, and varying the bias changes the width as well as the overall shape of the barrier. Since the tunnelling probability is very sensitive to the barrier thickness, this has a large effect on the tunnelling current. Thirdly, surface states located within the band gap of the semiconductor may provide an additional tunnelling path for positive bias. These localized

states may be filled by electrons from the conduction band and then emptied either by tunnelling into the metal or by recombination with a hole from the valence band.

4.2. Formalism and assumptions

The zero-temperature model developed in this section is based on the usual three-dimensional steady-state tunnelling formalism (Duke 1969) in which the tunnel current density, as a function of bias, is given by

$$J(V) = \frac{e}{2\pi^2 h} \int \mathrm{d}E_x \int \mathrm{d}^2 k_\parallel \, D(E_x)(f(E) - f(E + eV)) \tag{1}$$

in which E is the electron energy, $k_\parallel$ is the component of its momentum parallel to the junction, f is the Fermi function, $E_x = E - h^2k^2/8\pi^2 m$ and D is the barrier transmission factor.

The integral over $k_\parallel$ may be transformed to an integral over E giving, in the zero-temperature limit,

$$J(V) = \frac{4me}{h^3} \int_{E_{\min}}^{E_F} \mathrm{d}E \int_0^E \mathrm{d}E_x \, D(E_x). \tag{2}$$

For negative bias, E_F is the metal Fermi level and $E_{\min}$ is $E_F + eV$. For positive bias, E_F is the semiconductor Fermi level and $E_{\min}$ is the energy corresponding to the lowest point, in energy, of the conduction-band edge (figure 1). $E_F - E_{\min}$ is then always at least the semiconductor Fermi degeneracy and may be larger if the conduction-band edge is bent downwards. For computational convenience, this two-dimensional integral may be rearranged to give (Floyd and Walmsley 1978):

$$J(V) = \frac{4me}{h^3} \left(eV \int_0^{E_{\min}} \mathrm{d}E_x \, D(E_x) + \int_{E_{\min}}^{E_F} \mathrm{d}E_x \, (E_F - E_x) D(E_x) \right). \tag{3}$$

The conductance is then calculated as the derivative of this quantity with respect to bias.

The WKB approximation (Merzbacher 1970) for the barrier transmission factor $D(E_x)$ was used throughout this analysis:

$$D_{\mathrm{WKB}}(E) = \exp\left(-2 \int_{x_1}^{x_2} \mathrm{d}x \, \frac{2\pi\sqrt{2m(V(x) - E)}}{h} \right) \tag{4}$$

where x_1 and x_2 are the classical turning points and $V(x)$ is the potential barrier. The integrals in (3) were evaluated numerically. The shortcomings of the WKB approximation have been much discussed (see for example, Froman and Froman 1965). The approximation is really only valid for slowly varying potentials. A comparison of the WKB tunnelling probability with the exact result for a discontinuous potential such as a square or trapezoidal barrier (Grundlach 1966) shows that the WKB result seems to mimic the exponential dependence of the exact result, but neglects a multiplicative factor. This factor is only weakly bias dependent, at least for the barriers considered here, and so is unimportant in a normalized plot.

In all the models discussed below the metal and semiconductor conduction bands are taken to be parabolic. There is assumed to be a definite energy separating extended states in the semiconductor from localized states at the surface or at the bottom of the conduction band in undoped silicon. At the high doping levels used in this work, the donor impurity band overlaps with the conduction band possibly introducing localized

states within the original conduction band. This energy is taken as the bottom of an effective conduction band and the localized states are treated separately (section 4.5). The effective electron mass in the semiconductor is taken as $0.33m_0$ obtained from $(m_l^* m_t^{*2})^{1/3}$ where m_l^* and m_t^* are the density of states effective masses along, and perpendicular to, the principal axes of the ellipsoidal energy surfaces of silicon (Sze 1981). Further work will examine the effects of using different crystallographic orientations of silicon which will affect the choice of effective electron mass. The effective electron mass in the oxide is taken as $0.5m_0$, the effective mass for an electron in the conduction band of SiO_2 obtained from field-emission experiments using thicker oxides (Weinberg 1982). The mass in the metal is taken as m_0. The relative permittivities and electron affinities of silicon and silicon oxide are assumed to be 12 (ε_s) and 3 (ε_{ox}), and 4.15 eV (χ_s) and 0.9 eV (χ_i), respectively (Sze 1981). The normalized results were found to be rather insensitive to the assumed values of the effective masses, although the absolute values are strongly affected because of the exponential dependence in the tunnelling probability (4).

The silicon is assumed to remain in thermal equilibrium regardless of the current flowing so that the electron energy level occupation function is the usual equilibrium Fermi function. This assumption seems to be valid for degenerate semiconductors (Green and Shewchun 1974), at least for reasonable oxide thicknesses (>0.5 nm). Carrier freeze-out is assumed not to occur.

The metal Fermi energy is set arbitrarily as 5 eV throughout all the calculations. The actual value has no measurable consequences since the only important tunnelling contributions are from electrons close to the Fermi surface.

4.3. *Models neglecting band bending*

The simplest model barrier is that of a trapezoidal barrier formed by the oxide. Band bending in the semiconductor is neglected, and any workfunction difference or applied bias appears across the oxide. Figure 6 shows the effect of varying the oxide thickness and the metal workfunction on the computed conductance characteristic. All the curves have been normalized to their values at +50 mV; the actual conductances vary by several orders of magnitude, at a given bias. None of the curves presented in this figure resembles the experimental curves: the change in normalized experimental conductance with bias shown in figures 5(*a*) and 5(*b*) is much larger than for the theoretical curves. Such a large change could be produced by a changing trapezoidal barrier transmission factor only for an unrealistically low barrier height. The experimental curves also do not show the sharp drop at the semiconductor Fermi degeneracy present in the theoretical curves, but rather a shallow minimum at around this bias.

Consider curves a and b for which the metal workfunction has been chosen to make the barrier square at zero bias. The conductance has a minimum at zero bias and rises for increasing bias in both directions. This reflects the increase in the barrier transmission probability as the bias effectively lowers the far side of the barrier. Superimposed on this slow variation is an abrupt fall in the conductance at a positive bias equal to the semiconductor Fermi degeneracy. Above this point, further increases in bias bring no new occupied states in the semiconductor opposite unoccupied states in the metal. Further increase in conductance is again due to the slow variation in tunnelling probability. The main effect of altering the metal workfunction (curve c) is to make the zero-bias barrier trapezoidal and to shift the local conductance minimum away from

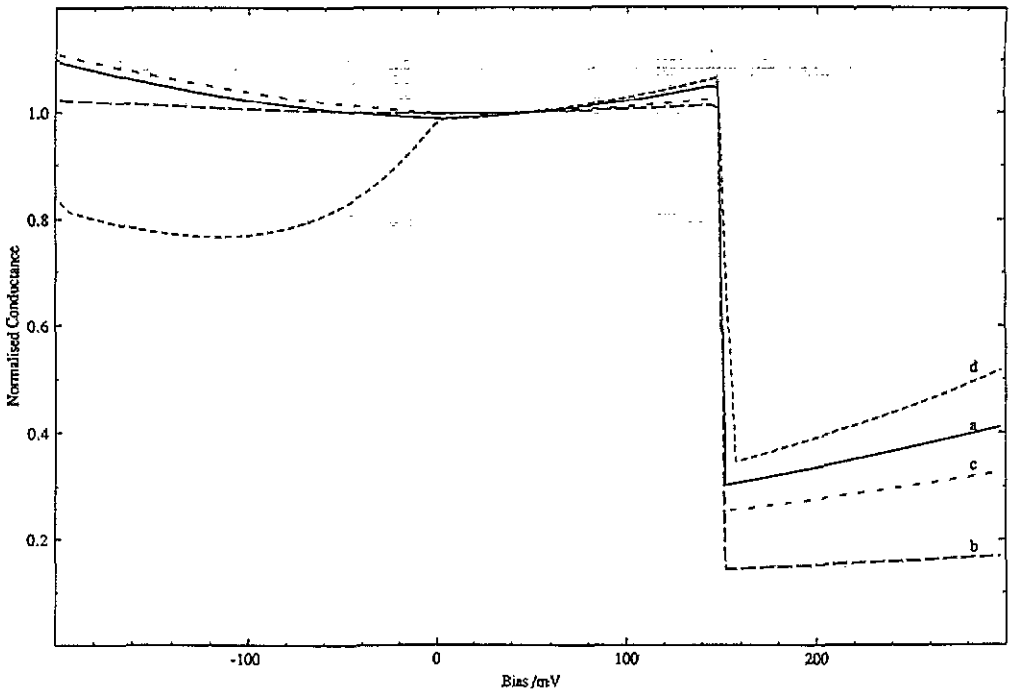

Figure 6. Theoretical conductance–bias plots. Curves a and b are for oxides of thickness 2 nm and 1 nm, respectively. The metal workfunction has been chosen equal to the silicon workfunction so that the barrier is flat at zero bias. Band bending in the silicon has been neglected. Curve c is for a 2 nm oxide but with the metal workfunction chosen to be 2 eV larger than before. Curve d is for the same conditions as a but including the effects of field penetration. The silicon doping density is $5 \times 10^{25}\ m^{-3}$.

zero bias. The simple trapezoidal barrier model sufficient to explain most of the characteristics of metal–insulator–metal junctions is inadequate here.

4.4. Models including band bending

In this section the consequences of including field-penetration into the electrodes are explored. The conduction-band edge close to the semiconductor surface is assumed to have a parabolic form. This neglects the charge density of the free carriers and is therefore only accurate where the semiconductor is depleted. However, the space-charge region in this highly doped material is in any case not expected to resemble a traditional 'Schottky barrier' as the Debye length is short compared with the mean distance between ionized donors. In addition, lattice distortion, donor clustering, a markedly discontinuous charge distribution, and image force effects contribute to the difficulty of modelling the barrier. A parabolic profile for the conduction-band edge is probably the best simple functional form which reproduces the gross tunnelling features of the real barrier. Band bending may be expected to have a significant effect on the conductance as the presence of a depleted region, increasing the effective width of the tunnelling barrier, will strongly reduce the current. If a negative bias is applied to the

ideal structure shown in figure 1(a) part of the bias will be dropped across the oxide and part across a surface layer in the semiconductor. Electrostatics gives

$$V_{ox} = -\frac{eN_D wd}{\varepsilon_{ox}\varepsilon_0} \quad \text{and} \quad w = -\frac{d\varepsilon_s}{\varepsilon_{ox}} + \sqrt{\frac{d^2\varepsilon_s^2}{\varepsilon_{ox}^2} + \frac{2\varepsilon_s\varepsilon_0|V_a|}{eN_D}} \tag{5}$$

where the symbols are explained in figure 1(b). (N_D is the semiconductor doping density.) For a positive bias, w is unchanged and V_{ox} is identical but positive. For this bias the energy range between the bottom of the conduction band and the Fermi level is widened at the semiconductor surface. Surface quantization effects are neglected.

Curve d in figure 6 shows the effect of varying the oxide thickness for an ideal structure in which there is no band bending at zero bias (the workfunction difference has been chosen to be zero and there is no trapped charge). For positive bias, the results are the same as in the last section except that the sharp drop in conductance has been pushed out to higher bias. At negative bias, the conduction-band edge bends upwards producing an initial decrease in conductance as the effective barrier thickness increases, at least for electrons tunnelling with a low energy. At higher bias, the conductance increases again, largely because an increasing proportion of electrons are tunnelling with energies in excess of the bias dropped across the depletion layer. The oxide transmission factor also varies as before.

Workfunction differences and/or a trapped charge density at the silicon/oxide interface may be included in (5) by the substitutions

$$V_a \rightarrow V_a - (\phi_m - \chi_s + \mu_s) + Q/C \qquad V_{ox} \rightarrow V_{ox} - Q/C \tag{6}$$

where Q is the charge density and C the insulator capacitance, both per unit area.

Figure 7 shows the effect of altering the metal workfunction. A decrease in this quantity bends the conduction-band edge downwards, whilst an increase bends it up even at zero bias and, for a large enough increase, depletes the surface. With the surface in this condition, the conductance is a very sensitive function of bias. A small increase in *negative* bias has two effects. Firstly, it produces a small increase in the voltage dropped across the depletion region and a consequent small increase in the width. However, an electron at the metal Fermi surface now tunnels with an increased energy and so has a reduced distance to tunnel because of the parabolic shape of the conduction-band edge. Secondly, the tunnelling probability is increased by the larger voltage drop across the oxide lowering the effective barrier presented by the oxide. In contrast, a small increase in *positive* bias both decreases the width of the depleted region and decreases the electron tunnelling distance. However, the effective barrier presented by the oxide may be increased and the conductance change at positive bias is the result of a balance between these effects. The sharp drop in conductance at a bias equal to the semiconductor Fermi degeneracy is still present, but less visible because of the change in scale. A set of similar characteristics may be obtained (figure 8) by supposing that charge is localized at the silicon/oxide interface (identified with interface states). The effects of varying the doping density on a structure with a large metal workfunction are shown in figure 9. The band bending at any given bias is reduced by increasing the doping density, which decreases the conductance variation with bias. Figure 10 compares the effect of a large metal workfunction difference with and without a trapped charge density. The presence of the charge density produces further band bending and increases the effect of the work-function difference. The effects of varying oxide thickness and the inclusion of image charge for a structure with the surface partially depleted by

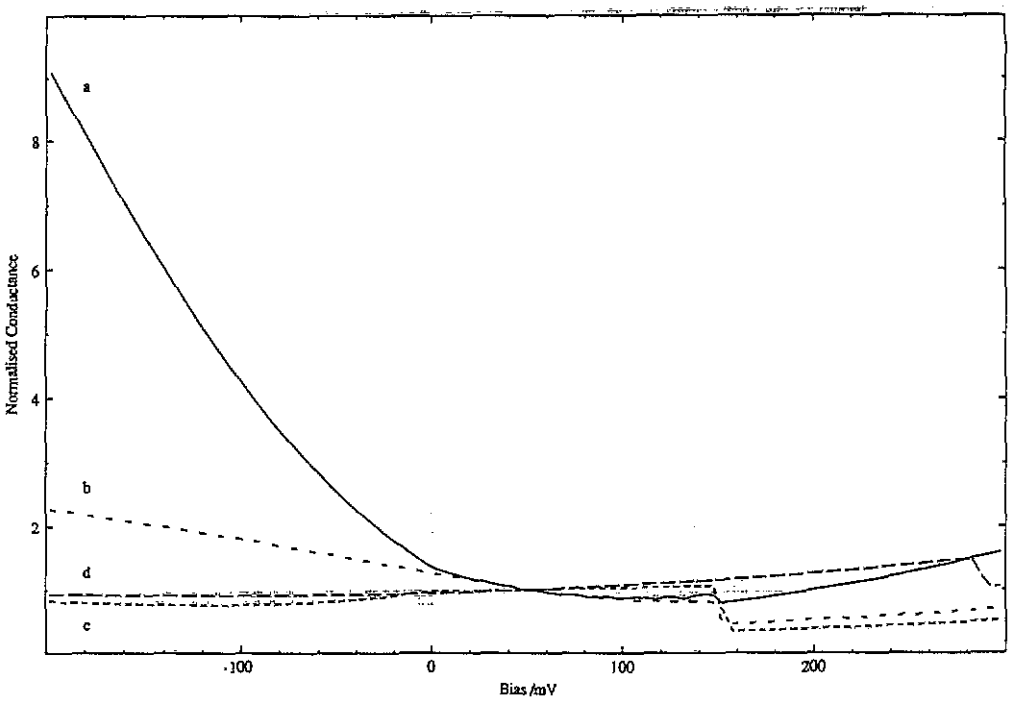

Figure 7. The effect of altering the metal workfunction, including the effects of band bending, for a 2 nm oxide. Curves a and b are for metal workfunctions 2 eV and 1 eV greater than the semiconductor workfunction, curve c for an equal workfunction and curve d for a workfunction 1 eV less. The silicon doping density is 5×10^{25} m^{-3}.

interfacial charge are shown in figures 11 and 12. A single image charge may be included in the calculation by making the substitution in the barrier potential

$$V(x) \rightarrow V(x) - e^2/16\pi\varepsilon_0\varepsilon x \tag{7}$$

where x is the distance from the metal electrode: only the metal image is included as the combination of the roughness of the electrode/barrier interface and the low electron density in the semiconductor probably prevent multiple reflections from being set up. The present of the image charge reduces the height of the barrier and in particular the size of the depleted layer in the semiconductor.

The effects of a large metal workfunction or trapped negative charge, or a combination of the two, can produce changes in the conductance with negative bias of the same magnitude as those observed. At positive bias, the theoretical curves show a minimum at the semiconductor Fermi degeneracy and sometimes a local minimum between zero bias the the main minimum. However, none of the effects described in this section can account for the large increases in conductance with increasing positive bias, above the semiconductor Fermi degeneracy, seen in some of the experimental characteristics.

4.5. Tunnelling from surface states

Tunnelling out of surface states at the silicon surface provides a parallel tunnelling path and may account for the sharp increase in conductance at positive bias. The steady-state

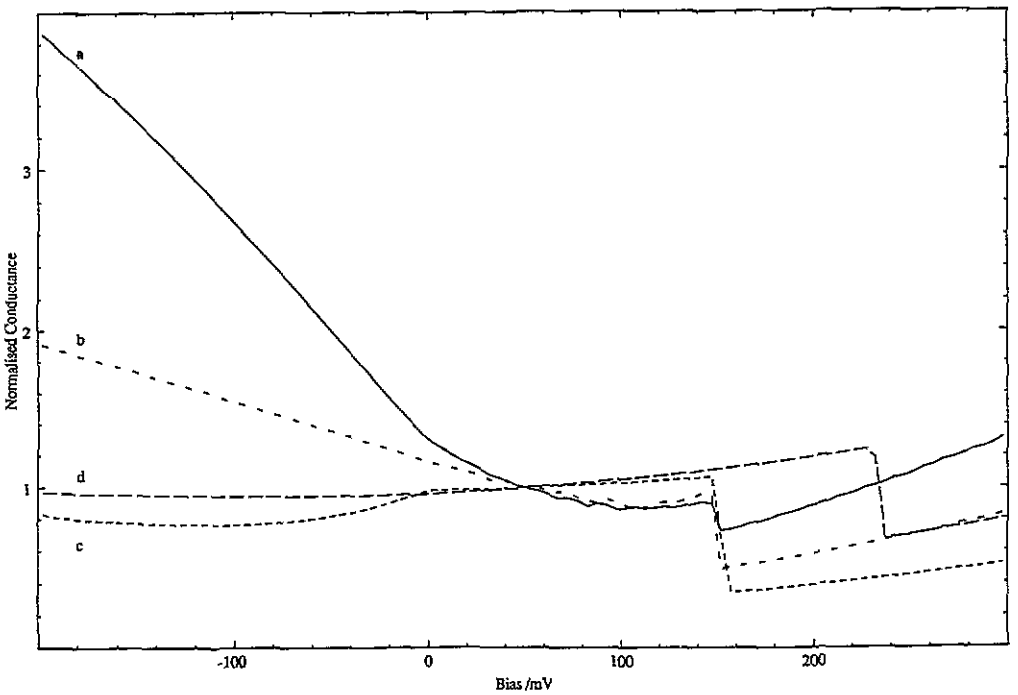

Figure 8. The effect of altering the charge density at the silicon/oxide interface, including the effects of band bending, for a 2 nm oxide. Curves a, b, c and d are for densities of $-0.02\ \mathrm{C\ m^{-2}}$, $-0.01\ \mathrm{C\ m^{-2}}$, zero and $+0.01\ \mathrm{C\ m^{-2}}$, respectively. The silicon doping density is $5 \times 10^{25}\ \mathrm{m^{-3}}$.

tunnelling theory used in the previous sections cannot be extended to this problem: an electron localized at the silicon surface has no well-defined wave vector to match to an extended state in the metal. Instead, a quasi-classical argument similar to that used in the simple theory of radioactive decay (Gamow 1928) may be used to estimate a decay time for the localized electron, before it tunnels into the metal. The contribution to the tunnel current from this process may be determined from this life time. The electron may tunnel out of a localized state at the semiconductor surface through a barrier composed of the Coulomb potential from the electron trap and the trapezoidal oxide barrier into an extended state in the metal. The transmission factor $D(E)$ for this process may be estimated by evaluating the probability that the electron is found in the region of space occupied by the metal. This is easily calculated from the wavefunction for the localized state if it is assumed that neither the presence of the metal nor any potential across the oxide affects the wavefunction. The probability is given by

$$D(E) = \int_0^{\pi/2} \mathrm{d}\theta\, 2\pi \sin\theta \int_{d/\cos\theta}^{\infty} \mathrm{d}r\, r^2 |\psi(r)|^2 \tag{8}$$

where $\psi(r)$ is the wavefunction and θ is the angle measured from a normal to the interface. Evaluating the integrals for a wavefunction of the form $\psi(r) \sim \exp(-r/a_0)$ for r within the oxide gives

$$D(E) = \tfrac{1}{2} \exp(-2d/a_0)(1 + d/a_0). \tag{9}$$

A pure Coulomb potential has $a_0 = \varepsilon_0 \varepsilon_{ox} h^2/\pi m_0 m_{ox} e^2 = 0.31$ nm, whilst a rectangular

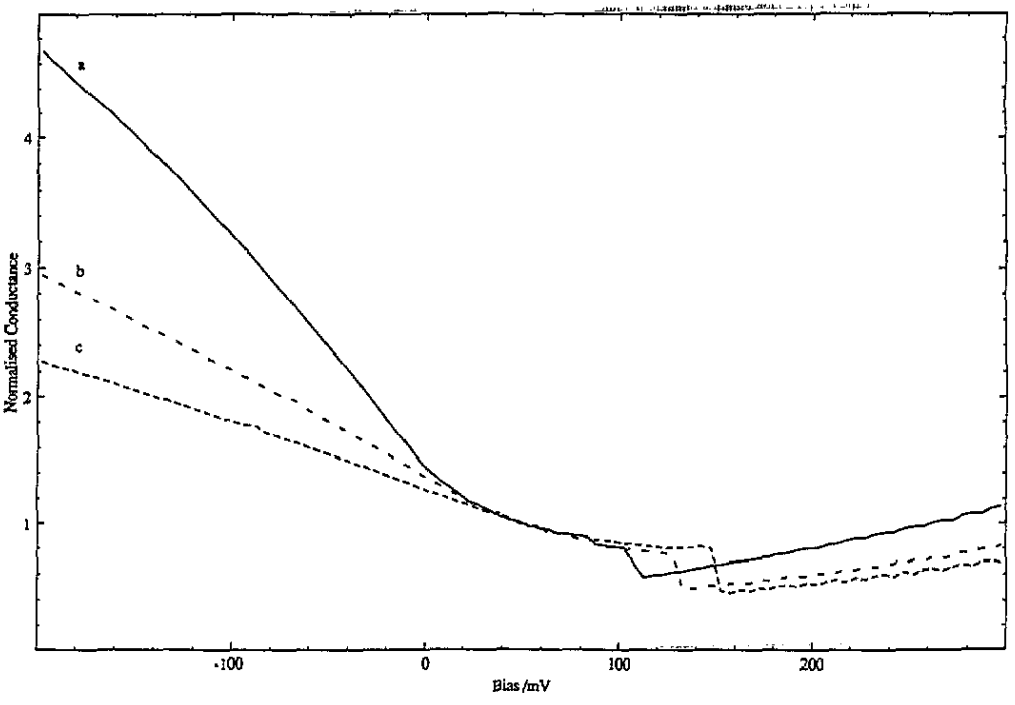

Figure 9. The effect of altering the doping density for a structure with a 2 nm oxide and a metal with a workfunction 1 eV greater than the silicon workfunction. Curves a, b and c are for doping densities of $3 \times 10^{25}\,m^{-3}$, $4 \times 10^{25}\,m^{-3}$ and $5 \times 10^{25}\,m^{-3}$, respectively.

potential (V_0) has $a_0 = 2\pi(2m_0 m_{ox}(V_0 - E))^{1/2}/h \approx 0.20$ nm (taking $V_0 - E \approx 2$ eV), so that the true decay length lies between these values. A more accurate calculation would reveal the true energy dependence of D, increased for electrons tunnelling out of traps with energies nearer to the top of the barrier.

Classically, the trapped electron vibrates within the potential well until it escapes; the 'vibration frequency' $\nu(E)$ may be estimated from the electron kinetic energy and the distance between the classical turning points for the trapped electron. For the pure Coulomb potential

$$\nu = \pi m_0 m_{ox} e^4 / 8\varepsilon_0^2 \varepsilon_{ox}^2 h^3. \tag{10}$$

For an assembly of similar models, of which $N(E, t)$ remain in the trapped state, this quasi-classical argument gives

$$dN(E, t)/dt = -N(E, t)\nu D(E) \tag{11}$$

or $N(E, t) = N(E, 0)\exp(-\lambda t)$ with $\lambda(E) = \nu D(E)$. The empty traps are filled by capturing electrons from the conduction band at a rate given by

$$\frac{dN(E, t)}{dt} = -(N_S(E) - N(E, t))\upsilon N_D \sigma(E) \tag{12}$$

where N_S is the total surface-state density per unit energy and per unit area, υ is the electron velocity estimated as 3/5 of the Fermi velocity and $\sigma(E)$ is the trap capture cross

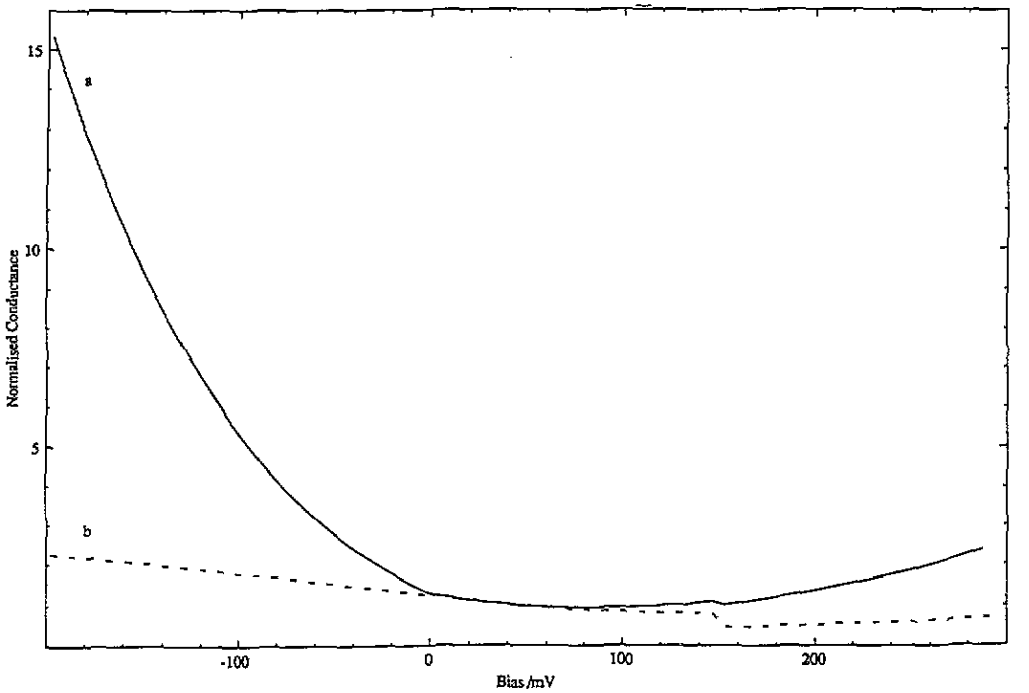

Figure 10. The effect of combining a large metal workfunction with a trapped charge density. Curve b is from a structure with a 2 nm oxide and a metal workfunction of 1 eV larger than the semiconductor workfunction. Curve a has, in addition, a $-0.03\ \mathrm{C\,m^{-2}}$ interfacial charge density. The silicon doping density is $5 \times 10^{25}\ \mathrm{m^{-3}}$.

section. Equating (11) and (12) gives the equilibrium trapped-electron population $N(E)$ and hence the equilibrium electron tunnelling rate as

$$-\mathrm{d}N(E)/\mathrm{d}t = N_S(E)[\nu D(E)/(1 + \nu D(E)/\upsilon N_D \sigma(E))]. \tag{13}$$

The current is given by

$$J(V) = \int \mathrm{d}E\, eN_S(E)\left(\frac{\nu D(E)}{1 + \nu D(E)/\upsilon N_D \sigma(E)}\right)(f(E) - f(E + eV)). \tag{14}$$

The main energy dependence of the integrand in (14) comes from the capture cross section, $\sigma(E)$, rather than from the transmission factor $D(E)$ or the surface-state density $N_S(E)$, at least for energies near to the conduction-band edge (for example Tredwell and Viswanathan 1980) and for the trapping rate much lower than the tunnelling rate. As a simple model we evaluate (14) at zero temperature and neglecting the energy dependences of $D(E)$ (using (9)) and $N_S(E)$. We model the capture cross section as $\sigma(E) = A[\exp(\alpha E) - 1]$, measuring energy downwards from the conduction-band edge, giving the observed roughly exponential increase with energy and forcing the cross section to be zero at the bottom of the conduction band. Evaluating (14) under these conditions and then differentiating with respect to bias yields the conductance as

$$G(V) = N_S e^2 \nu D\{1 - \exp[-\alpha(eV - \mu_s)]\}/\{1 + \gamma \exp[-\alpha(eV - \mu_s)]\} \tag{15}$$

where $\gamma = \nu D/\upsilon N_D A - 1$ and V is the bias defined as before. For $\gamma \gg 1$ the conductance

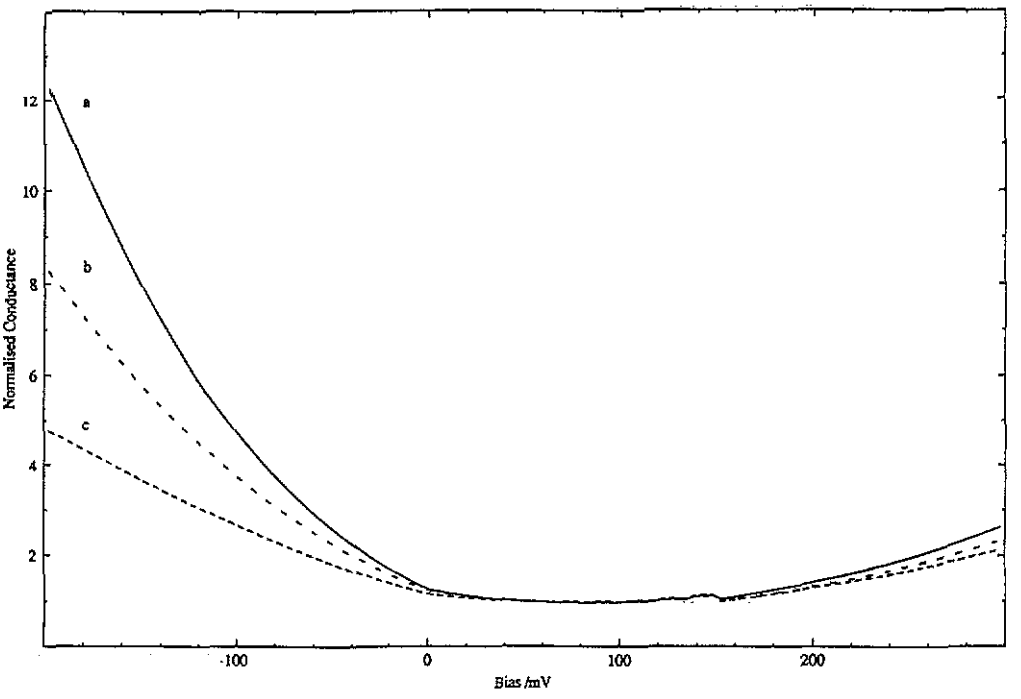

Figure 11. The effect of varying the oxide thickness for a structure with a $-0.03\ \mathrm{C\,m^{-2}}$ interfacial charge density. The silicon doping density is $5 \times 10^{25}\ \mathrm{m^{-3}}$. The curves a, b and c are for thicknesses of 2 nm, 1.5 nm and 1 nm, respectively.

is controlled by the trapping rate and the conductance increases exponentially with bias as shown in figure 13 curve a. Choosing a larger value for d/a_0 strongly reduces D and gives $\gamma \ll 1$. Under these circumstances the conductance is controlled by tunnelling and has a very different bias dependence, tending to a constant value as the bias is increased (15).

In our experiments all the conductance characteristics showed an exponential increase with bias, indicating that the conductance was limited by the trapping rate and not by tunnelling. To generate figure 13 curve a, which resembles the experimental curves, values of $\gamma = 250$ together with $N_S = 3 \times 10^{15}\ \mathrm{m^{-2}\,eV^{-1}}$, $A = 4 \times 10^{-23}\ \mathrm{m^2}$, $\alpha = 20.9\ \mathrm{eV^{-1}}$ (from Tredwell and Viswanathan 1980) were chosen. Using (10) to estimate ν suggests that the ratio d/a_0, which determines D (9), is around 6.45. The ratio is difficult to calculate, as both d and a_0 are uncertain, but this value is plausible as the oxide layer is rather thinner in these junctions than in the usual metal–insulator–metal junctions; tunnelling occurs through a depletion layer in series with the oxide layer. Experimentally, applying a sufficiently large positive bias (~1 V) will remove the depletion layer and produce tunnelling out of extended states within the semiconductor through the oxide alone. The current density at this bias is very high suggesting that the oxide is relatively thin and that the bulk of the junction resistance near zero bias is contributed by the depletion layer, favouring a low value for d/a_0. The total conductance is given by the sum of the surface-state contribution (15) and the usual bulk tunnelling term, the derivative of (3) with respect to bias.

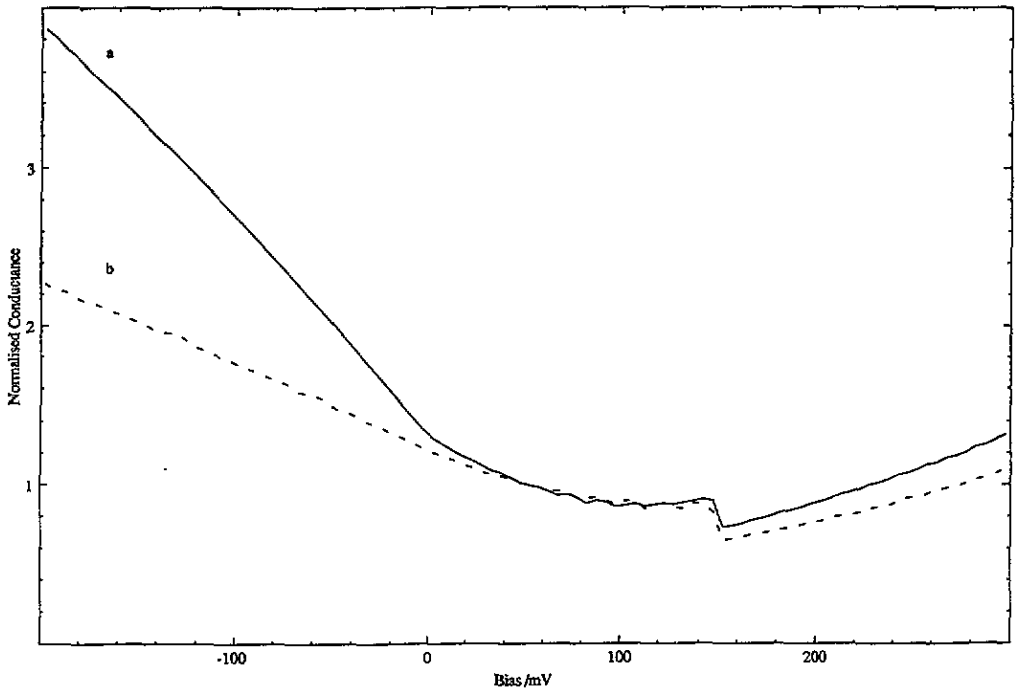

Figure 12. The effect of including an image charge from the metal electrode for a structure with a 2 nm oxide and an interfacial charge of $-0.02\,C\,m^{-2}$. The silicon doping density is $5 \times 10^{25}\,m^{-3}$. Curves a and b are without and with the image charge, respectively.

The approach adopted in this section to estimate the tunnelling contribution from surface states is easily criticized, but a more rigorous treatment of the problem is not justified in view of the uncertainties in the nature of the localized state and the tunnelling barrier.

5. Discussion

A comparison of the theoretical and experimental curves suggests that there is a depleted region at the silicon surface for all the samples examined, and the variation in shape with bias of this layer dominates the conductance characteristic. The main parameters governing the size of this layer are the metal workfunction, the silicon/oxide barrier height, the doping density and the surface-state density. The electron density is much higher in aluminium ($10^{29}\,m^{-3}$) than in the denegerately doped silicon used in this work ($10^{25}\,m^{-3}$), so that there is no depleted region for an equivalent aluminium junction. This accounts for the differences highlighted in figure 2. The main results of the model are summarized in figure 13. Curve b corresponds to a structure with a metal workfunction greater than the silicon workfunction by 2 eV, and a surface-state density of $3 \times 10^{15}\,m^{-2}\,eV^{-1}$, neglecting the possibility of tunnelling from surface states. Curve a shows the contribution to the conductance from tunnelling via surface states and curve c is the total conductance, obtained by adding (a) and (b).

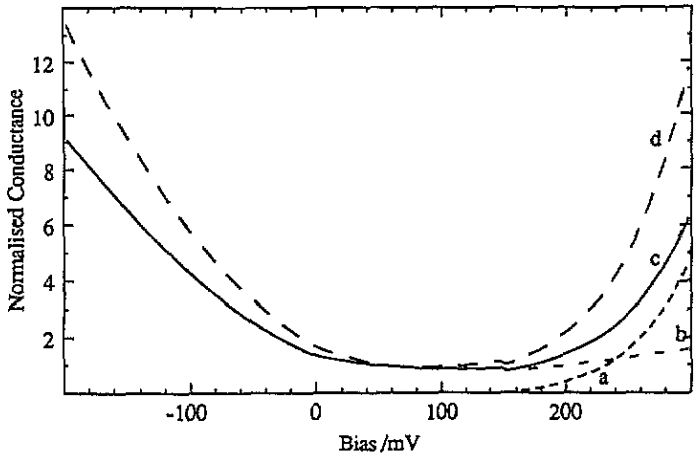

Figure 13. Summary of the model calculations for a structure with a 2 nm oxide, a metal workfunction 2 eV higher than the semiconductor workfunction and an interfacial charge of -6×10^{-4} C m^{-2}. The silicon doping density is 5×10^{25} m^{-3}. Curve a is the contribution to the conductance from surface-state tunnelling. Curve b is the contribution to the conductance from tunnelling out of extended states, including the band-bending effect of the workfunction difference and the charge. Curve c is the total tunnelling conductance, the sum of a and b. Curve d is the equivalent of c but with an interfacial charge density of -6×10^{-3} C m^{-2}.

The observed effect of a hydrogen plasma shown in figure 3 is entirely consistent with the model presented above. Reducing the surface-state density reduces the conductance change with negative bias by reducing the size of the depleted region as demonstrated in figure 8. The conductance at positive bias is also reduced directly, by reducing the density of surface states available to participate in any tunnelling process. Figure 13 curves (c) and (d) compare the computed conductances for structures with surface-state densities of 3×10^{15} m^{-2} eV^{-1} and 3×10^{16} m^{-2} eV^{-1}, respectively. The metal workfunction has been chosen as 4 eV (appropriate for a lead top electrode) and the silicon electron affinity as 2 eV. The latter value, rather less than the bulk value, is taken from Horiguchi and Yoshino 1985. For this choice of parameters the agreement with the experimental results in figure 3 is good. In our model, the difference between the shape of the computed and experimental curves at positive bias reflects the actual energy dependences of the surface-state density, the transmission factor and the recombination time found in the real experimental system (Sze 1981). Hence, the conductance change at *positive* bias is partly determined by the surface-state density as a function of energy below the conduction-band edge, whilst the conductance change at *negative* bias is controlled by the integrated density of surface states. (Possible applications of this observation to measuring interface-state densities at interfaces prepared by commercially important oxidation techniques, in energy regions inaccessible to conventional MOS measurements, are being investigated.) The experimental curves a and b in figure 2 suggest that room temperature oxidation gives a much lower surface-state density near to the conduction-band edge than oxidation at 150 °C.

Decreasing the doping density (figure 9) increases the width of the depleted region, decreasing the zero-bias conductance and increasing the conductance variation with bias. These effects are observed experimentally, as shown in figure 4. Image force effects (figure 12) probably account for the theoretical change being larger than the

experimental change. The basic effects of top electrode material variation can be understood on the basis of the changing workfunction. Figure 5(*b*), in which room temperature oxidation has been used, giving a low surface-state density, should be compared with figure 7, and figure 5(*a*) (oxidation at 150 °C for a higher surface-state density) with figure 10. In both the experimental and theoretical curves, increasing the metal workfunction produces an increasing change in conductance with bias. The presence of an interfacial charge amplifies the effect of the change in workfunction. Due to the presence of a higher surface-state density in figure 5(*a*), a 0.1 eV change in metal workfunction gives a similar change to that observed in figure 5(*b*) for a workfunction change of 0.6 eV. The observed conductance characteristic for the magnesium junction indicates that even for a metal with such a low workfunction the surface is still depleted, again suggesting that the silicon/oxide barrier height is less than its bulk oxide value; a finding in agreement with other workers (Horiguchi and Yoshino 1985).

The agreement between theory and experiment is good in view of the many uncertainties in the model parameters and the approximations made in developing the theory. The conductance derivatives, as a function of bias, for the samples decribed here have also been measured and the results will be reported elsewhere.

6. Conclusions

In this paper we have developed a detailed numerical model of the transport properties of metal–insulator–semiconductor tunnel junctions. The theoretical work has been developed alongside a rigorous experimental exploration of the effect of varying the important barrier parameters, using oxidation of n-type degenerate silicon as a model system. In particular the model accounts semi-quantitatively for the effects of changing workfunction difference, semiconductor doping density, surface-state density and insulator thickness. The model highlights the new features introduced by using a semiconducting rather than a metallic electrode, and also provides a way of understanding and characterizing semiconductor–insulator–metal junctions.

Acknowledgments

Technical support by MEMC was invaluable. Financial support for one of the authors (PMO) was given by the SERC and by MEMC, via a CASE award.

References

Adkins C J and Phillips W A 1985 *J. Phys. C: Solid State Phys.* **18** 1313
Archer R J 1957 *J. Electrochem. Soc.* **104** 620
Brinkman W F, Dynes R C and Rowell J M 1970 *J. Appl. Phys.* **41** 1915
Busmann H-G, Ewert S, Sander W and Seibert K 1985 *Z. Phys. B* **59** 439
Duke C B 1969 *Tunnelling in Solids* (New York: Academic)
Floyd R B and Walmsley D G 1978 *J. Phys. C: Solid State Phys.* **11** 4601
Froman N and Froman P O 1965 *JWKB Approximation: Contributions to the Theory* (Amsterdam: North-Holland)
Gamow G 1928 *Z. Phys.* **51** 204
Green M A and Shewchun J 1974 *Solid State Electron.* **17** 349
Grundlach H K 1966 *Solid State Electron.* **9** 949

Horiguchi S and Yoshino H 1985 *J. Appl. Phys.* **58** 1597
Merzbacher E 1970 *Quantum Mechanics* 2nd edn (New York: Wiley)
Sze S M 1981 *Physics of Semiconductor Devices* 2nd edn (New York: Wiley)
Tredwell T J and Viswanathan C R 1980 *Appl. Phys. Lett.* **36** 462
Weast R C (ed) 1988 *Handbook of Chemistry and Physics (1988–89)* 68th edn (Boca Raton, FL: Chemical Rubber Company) E89–90
Weinberg Z A 1982 *J. Appl. Phys.* **53** 5052