Fermi-level position at a semiconductor-metal interface

A. Zur, T. C. McGill, and D. L. Smith
The Thomas J. Watson, Sr. Laboratory of Applied Physics, California Institute of Technology, Pasadena, California 91125
(Received 28 February 1983)

We have investigated the phenomenon of Fermi-level pinning by charged defects at the semiconductor-metal interface. Two limiting cases were investigated. In the first case we modeled an infinitely thick metallic coverage. In the second case we modeled a submonolayer coverage by using a free semiconductor surface containing defects. In both cases we assumed that most of the defect-induced interface states are localized inside the semiconductor, not more than a few angstroms away from the metal. Under these conditions we have estimated the difference in Fermi-level position between n- and p-type semiconductors to be less than 0.05 eV in the case of a thick metallic coverage. This difference was shown to be the maximum possible one, and it occurs only when there is no pinning. When there is pinning, this difference is even smaller. No such upper bound on the difference in Fermi-level position exists in the case of submonolayer coverage. We have also found that the defect density required to pin the Fermi level is $\sim 10^{14}$ cm^{-2} in the case of a thick metallic coverage, but only $\sim 10^{12}$ cm^{-2} in the case of a submonolayer coverage.

I. INTRODUCTION

Recent experiments involving submonolayer coverage of metal atoms on semiconductor surfaces suggest that the Fermi-level position for n- and p-type semiconductors may differ by as much as 0.2 eV.[1–4] However, in measurements of Schottky barriers consisting of a bulk metal against a bulk semiconductor, the Fermi-level position at the metal-semiconductor interface is found to be the same to less than 0.1 eV.[5] Yet the measurements of Fermi-level position for submonolayer coverage seem to show the same behavior as those for thick metallic layers in that the Fermi level pins and is approximately the same for a number of different metals. Hence one might enquire as to what the connection should be between the experimental results for submonolayer coverages and those for macroscopic Schottky barrier.

Daw and Smith[6] reported very briefly in a theoretical study on the difference in the Fermi-level position on n- and p-type semiconductors. They concluded that the doping type cannot significantly influence the position of the Fermi level in a macroscopic Schottky barrier, although it can for submonolayer coverages. In this paper we report on a much more detailed study of the variation of the Fermi-level position at the metal-semiconductor interface for a macroscopic Schottky barrier and contrast it with a simple model for a submonolayer coverage. Our discussion is based on the recently proposed unified defect model of Spicer *et al.*[7] that assumes pinning of the Fermi level by charged defects at the interface. Various models of spatial arrangements of defects are considered. While in detail the models may not precisely conform to the experimental situation (i.e., in distribution of defects, etc.), we can draw a model-independent conclusion that it is difficult to have the Fermi level for p- and n-type semiconductors be substantially different (~ 0.1 eV) for a macroscopic Schottky barrier, while it is possible in the case of a submonolayer coverage. Hence we must conclude that if the experimental results for submonolayer coverage were extended to include thicker overlayers, the Fermi-level position on n- and p-type semiconductors would eventually merge, to within 0.1 eV.

This paper is organized in the following way. In Sec. II we discuss the physics of the spatial arrangement of the charge in the case of a submonolayer coverage and coverage with bulk metals. Section III describes the calculational method. Sections IV and V present the results for a submonolayer coverage model and the results for a macroscopic metal coverage. We describe the behavior of the Fermi level for the surface with defects (a model for a submonolayer coverage) and three cases of bulk metal coverage: One in which there are no defects, one in which we have a simple defect that has two charge states for the Fermi level within the gap, and one in which the defect has three charge states.[8] In Sec. VI we present our conclusions based on this study.

II. THE SPATIAL ARRANGEMENT OF CHARGE

According to the unified defect model,[7] the Fermi-level position at the interface is influenced by the presence of chargeable defects at the semiconductor surface. Since the whole system is electrically neutral, there must be charges that compensate the defects somewhere in the system. For submonolayer coverage, these charges are the shallow impurities in the depletion region of the semiconductor. When there is a thick metallic coverage on the semiconductor, there could be another major source of charge, namely, the screening charge in the metal. In each case, the equal and opposite charges form a dipole layer analogous to a parallel-plate capacitor. In both cases, the voltage across this capacitor should be of the order of 1 V. However, the distance between the plates is very different in the two cases, resulting in very different charge densities that are required to pin the Fermi level. For the submonolayer case, this distance is typically of the order of a few thousand angstroms. In that case, a 1-V potential difference is achieved with a depletion charge density of 10^{11}–10^{12} cm^{-2}. In the thick metallic coverage the separation of charge is essentially the distance between the

defects and the metal. This distance is of the order of a few angstroms. Therefore, the density of chargeable defects required to pin the Fermi level should be at least of the order of 10^{14} cm^{-2}.

The sign of the charge on the interface defects is also determined by different physical mechanisms in the two cases. In the submonolayer coverage, the depletion region can charge only one way (positive for n type, negative for p type). Hence only defect acceptors can charge on n-type, and only defect donors can charge on p-type semiconductors. Therefore, the sign of the charge on the interface defects is determined by the bulk doping type. The position of the Fermi level is determined by the energy level of that defect that can charge with the right sign, assuming both surface donors and acceptors are present. Thus the Fermi-level position can depend on the doping type for submonolayer coverages. In the thick metallic coverage case, the metal can supply charges of either sign and the magnitude of the metal screening charge can greatly exceed that of the semiconductor depletion charge. The sign of the charges on the defects will be determined by the metal work function relative to the defect energies prior to any charge transfer. Therefore, the charge on the defects will be determined mainly by the metal work function, with semiconductor doping playing only a minor role, owing to the much larger charging capacity of the metal. The Fermi-level position in the interface is determined by the energy level of the defect that is being charged. This defect level is determined by the metal work function, but not by the semiconductor doping.

The spatial arrangements of the charge model the sources of charge mentioned above. For the case of the submonolayer coverage, we have assumed defects that are very close to the surface. The only sources of charge are these defects and the bulk doping of the semiconductor. We assume here that the adatoms do not contribute any significant amount of charge. In the case of the bulk metal, we have included three sources of charge: screening charge in the metal, the defects, and the doping of the semiconductor. The schematic of our metal-semiconductor interface is given in Fig. 1. We assume that the

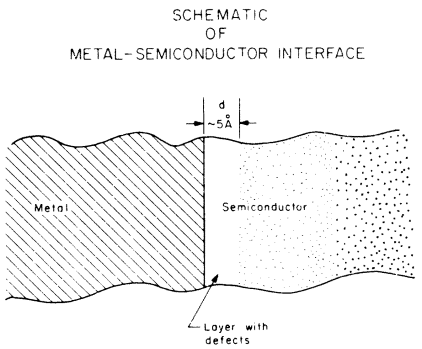

FIG. 1. Geometry of our model. We assume that the metal terminates sharply at $x=0$ and that all the interfacial transition occurs at $0<x<d$, so that the semiconductor from $x=d$ and on has a regular crystalline structure with a few imperfections and the bulk band structure is meaningful for $x>d$. All the defects are localized in the intermediate region $0<x<d$, and we assume that they are all on $x=d$. The plane $x=d$ is also the place where we calculate the Fermi-level position.

metal-semiconductor junction consists of three regions: semi-infinite bulk metal, semi-infinite bulk semiconductor, and interface region. The schematic of our submonolayer coverage model consists of the right (dotted) part of the figure, with no metal present, i.e., only two out of three regions. We now discuss those three regions in more detail.

A. The interface region in the semiconductor

The interface region in the semiconductor is the narrow region where the charged defects are located. We will take this region to be at $0<x<d$. The region is sufficiently narrow that transport of carriers through this region is uninhibited. Hence the dipole layer that exists between the charged defects and the metal is included in the abrupt barrier that is usually assumed and is not measurable separately.

We will assume that there is no charge inside this layer and that the defects are all on one side of this layer ($x=d$). Later we will show that our model is insensitive to that assumption, that is the effect of distributing the defects throughout the interface region is equivalent to putting them all on a single plane at some effective distance from the metal. The band bending in this layer will be equal to the electric work required to cross it,

$$\Delta V_i=\int_0^d q\mathscr{E}_x dx\ , \tag{2.1}$$

where $\mathscr{E}_x$ is the electric field in this interface layer, and q is the electron charge (taken positive).

The most important parameters of this layer are its effective width (d), its dielectric constant (ϵ), the defect density (σ), and their ionization energies. We will assume that almost all the defects are localized within few Å from the metal surface, and therefore in this model we take this width to be 5 Å, i.e., two to three atomic layers. We will also assume that the dielectric constant in this layer equals that of the bulk semiconductor.

B. The metal

We use a jellium model[9] to describe the metal. The electronic energies near the metal surface in this jellium approximation are shown schematically in Fig. 2. In this figure Φ_m is the work function; μ is the chemical potential which is constant throughout space for equilibrium conditions. In general, the work function is written as an internal part, plus a surface dipole contribution.[10] In this paper, we denote by Φ_m only the internal part.

Near the interface, the electrostatic potential varies due to noncancelling charge densities. The variation in the potential results in an offset between the electrostatic potential at the interface and far inside the metal. The parameter of interest to us is this difference ΔV_m, which might change somewhat in order to screen charges in the semiconductor. In this paper we are interested in variations between metal interfaces to n- and p-type semiconductors, and this ΔV_m will have to screen the difference in depletion charge between the two cases. We do not expect this difference in ΔV_m to be large, but it might play some role, and we want to estimate its magnitude. In general, this ΔV_m will be a function of the net excess charge inside the metal Q_m, where

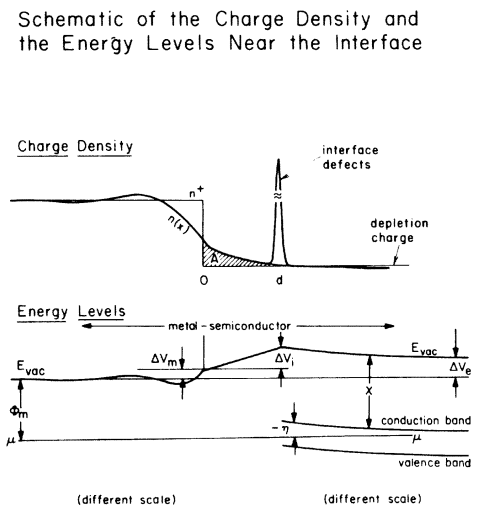

FIG. 2. Charge density and energy levels at a metal surface using our model. In the upper part we show the charge-density profile. We used the jellium model for the metal side, having a uniform background of n^+ extending up to $x=0$. The charged defects at $x=d$ are shown here as a very sharp Gaussian. In our calculations we took it to be infinitely sharp. The finite charge distribution outside the metal surface (the hatched region denoted by A) which in a free metal surface would balance the charge inside the metal is neglected in our calculation of the electrostatic potential. In the lower part we show the energy levels as functions of position. μ is the chemical potential which is constant throughout space at equilibrium. Φ_m is the metal work functions; ΔV_m is the potential difference between the jellium surface and bulk; ΔV_i is the potential difference across the dipole layer; d is the dipole layer width, assumed here to be 5 Å; $\eta(x)$ is the Fermi-level position relative to the conduction-band edge; and χ is the semiconductor electron affinity. The total potential difference across the interface is ΔV_e. Note the change in scale between the metal and the semiconductor side. We stretched the metal side for clarity.

$$Q_m=\int_{-\infty}^{d}[n^+-n(x)]dx\ ,$$

where $n^+,n(x)$ are the jellium positive background and electrons densities. In the linear approximation a change in ΔV_m can be related to a change in ΔQ_m by

$$\Delta V_m(\Delta Q_m)\approx\Delta V_m^0+\left[\frac{\partial\Delta V_m}{\partial Q_m}\right]_0\Delta Q_m\ , \tag{2.2}$$

where the zero point refers to a contact to intrinsic semiconductor, and ΔQ_m is the difference in Q_m between this case and either n- or p-type semiconductor. The parameter $(\partial\Delta V_m/\partial Q_m)_0$ could play some role in our calculation. In order to estimate the size of this parameter, we replace the metal-semiconductor interface, by a jellium-vacuum interface, and use the Mahan-Schaich theorem.[11] According to this theorem, modified to non-neutral jellium surface,

$$\Delta V_m(Q^*)=n_0\frac{dE(n_0)}{dn_0}+\frac{q^2Q^{*2}}{2n_0\epsilon_0}\ . \tag{2.3}$$

This formula was derived from (19), (2), (3), and (A5) in Mahan and Schaich's paper, but using our notation. Here n_0 is the jellium positive background density, and $E(n)$ is the electron energy functional. Q^* is the excess charge near the jellium surface, i.e.,

$$Q^*=\int_{-\infty}^{\infty}[n^+-n(x)]dx\ ,$$

and it is somewhat different from our Q_m. The difference is due to the tail of the charge distribution in the region $x>d$, which is missing in the definition of Q_m.

Using (2.3), we obtain

$$\left[\frac{\partial\Delta V_m}{\partial Q^*}\right]=\frac{e^2}{n_0\epsilon_0}Q^*\ .$$

In our results section, we are going to find that as many as 10^{14} defects per cm^2 are required to pin the Fermi level. Upon charging, these defects may deplete the metal by a similar amount of charge. Therefore, we should use $|Q^*|<10^{14}$ cm^{-2}. For $Q^*=10^{14}$ cm^{-2} and n_0 corresponding to $r_s=5$, we get $(\partial\Delta V_m/\partial Q^*)=1.1\times10^{-14}$ V cm^2. For most metals, $r_s<5$, and this number will be even smaller. Since jellium is only a rough approximation to real metals, and since Q^* and Q_m are not exactly the same, we consider this number to be an order of magnitude estimate, and in this paper we take ΔV_m^0 and $(\partial\Delta V_m/\partial Q_m)_0$ to be parameters of the model and allow $(\partial\Delta V_m/\partial Q_m)_0$ to change between 0.1 and 1.0 eV/10^{14} e cm^{-2}. Our results will show that these two cases are very similar, suggesting that the effect of $(\partial\Delta V_m/\partial Q_m)_0$ is not critical.

C. The bulk semiconductor

A number of parameters characterize the bulk semiconductor. The position of the chemical potential with respect to the conduction band edge as a function of position $\eta(x)=\mu-E_c(x)$, and the electron affinity χ, the position of the conduction-band edge relative to the vacuum for an electron with zero kinetic energy in the absence of band bending. We will take the semiconductor to occupy the space along the positive x axis, with $d<x<\infty$. We know $\eta(\infty)=[\mu-E_c(\infty)]$, the relative position of the Fermi level at $x=\infty$, for a given host semiconductor, doping and temperature, but $\eta(x)=\mu-E_c(x)$ for $x<\infty$ is yet to be determined. The important property of this semi-infinite semiconductor is that knowing $\eta(x)$ at $x=d$, we can find $\eta(x)$ for any $x>d$ and calculate the total number of charges (Q_s) per unit area of the semiconductor by solving Poisson's equation:

$$\begin{aligned}&\frac{d\eta}{dx}=\frac{q^2}{\epsilon\epsilon_0}Q_s(x),\quad \eta=(\mu-E_c)\ ,\\ &\frac{dQ_s}{dx}=n(\eta)+\begin{bmatrix}-N_D^+(\eta)\\ +N_A^-(\eta)\end{bmatrix}-p(\eta)\ ,\end{aligned} \tag{2.4}$$

where expressions for $n(\eta)$, $p(\eta)$, $N_D^+(\eta)$ and $N_A^-(\eta)$ are well known.[12] Here, n,p,N_D^+,N_A^- are the densities of electrons in the conduction band, holes in the valence band, ionized donors (in n-type semiconductor) and ionized acceptors (in p-type semiconductor). Poisson's equation (2.4) can be reduced to an integral

$$Q_s(\eta(d))=\left[\frac{2\epsilon\epsilon_0}{q^2}\int_{\eta(\infty)}^{\eta(d)}[n(\eta)-N_D^+(\eta)-p(\eta)]d\eta\right]^{1/2} \tag{2.5}$$

for n-type semiconductor, with a similar formula for p type, in which there is a minus sign in front of the square root, and $+N_A^-(\eta)$ replaces $-N_D^+(\eta)$.

In this work it was assumed that the semiconductor is not degenerately doped and that the Fermi level at the interface is inside the gap and at least $2k_BT$ from the gap edges. In this case, we could assume that the carriers in the semiconductor obey Maxwell Boltzmann statistics, and the integral in (2.5) was calculated analytically.

III. CALCULATION OF THE FERMI-LEVEL POSITION

A. Semiconductor-vacuum interface

The calculation of the Fermi-level position at the semiconductor surface is very simple. For a given Fermi-level position at the surface, the total charge density in the defects is given by

$$\sigma^+(\eta)=\frac{\sigma}{g_s\exp\left[\frac{\mu-E_s}{k_BT}\right]+1}=\frac{\sigma}{g_s\exp\left[\frac{\eta+E_c-E_s}{k_BT}\right]+1}, \tag{3.1}$$

where σ is the total density of defects, and g_s is the state degeneracy.[12] The compensating charge density in the bulk depletion layer is given by (2.5). Since these two sources of charge must balance, a solution is obtained by equating the two charges and hence obtaining the Fermi level at the surface.

B. Bulk metal–semiconductor interface

We may calculate the Fermi-level position using the model described above. The electrostatic potential variation from deep in the metal to deep in the semiconductor is given by ΔV_e

$$\Delta V_e=\int_{-\infty}^{+\infty} q\mathscr{E}(x)dx\ , \tag{3.2}$$

where $\mathscr{E}(x)$ is the electric field at x. By comparison with Fig. 2 we see that ΔV_e must be given by[13]

$$\Delta V_e=\chi-\eta(\infty)-\Phi_m\ .$$

Breaking the integral in Eq. (3.2) into three parts, $(-\infty,0)$, $(0,d)$, and (d,∞) we have

$$\eta(d)+\Phi_m+\Delta V_m-\chi=-\int_0^d q\mathscr{E}(x)dx\ . \tag{3.3}$$

We can write the electric field $\mathscr{E}(x)$ in the interface layer as a sum of two contributions, one ($\mathscr{E}_x^s$) due to dipoles formed by charged defects, and the other ($\mathscr{E}_x^b$) due to dipoles formed by space charges in the semi-infinite semiconductor. In general, ($\mathscr{E}_x^s$) is much larger than ($\mathscr{E}_x^b$). However, ($\mathscr{E}_x^b$) is responsible for any difference between n- and p-type semiconductors. We neglect any charge due to doping in the interface layer since their number in such a thin slab of a semiconductor is very small.

If we know the statistics of charged defects in the interface layer as a function of $\eta(d)$, we can find their contribution to the electric field inside that layer. The bulk contribution is also known and equals

$$\mathscr{E}_x^b=\frac{-q}{\epsilon\epsilon_0}Q_s(\eta(d))\ ,$$

where $Q_s(\eta(d))$ is given by Eq. (2.5), and ϵ is the dielectric constant in this layer, which is assumed to be the same as in the semiconductor bulk. Now, given Φ_m, ΔV_m^0, $(\partial\Delta V_m/\partial Q_m)_0$, and χ we solved Eq. (3.3) to obtain $\eta(d)$ as a function of $\Phi_m+\Delta V_m^0-\chi$. The plot of $\eta(d)$ vs $\Phi_m+\Delta_m^0-\chi$ for different cases will give the variation of the Fermi level with metal work function for a fixed semiconductor.

Throughout our calculations, we arbitrarily chose all the semiconductor parameters that enter implicitly into Eq. (2.5), (effective masses, dielectric constant, band gap, impurity levels) to be those of GaAs. ($E_{\text{gap}}=1.43$ eV, $m_e^*/m_0=0.068$, $m_h^*/m_0=0.5$, $\epsilon=13.1$, and shallow impurity levels are taken to be 5 and 25 meV for donors and acceptors, respectively.) Actually, the only relevant parameters are dielectric constant and band gap, so it is very easy to correct the results to any other semiconductor.

IV. RESULTS FOR SUBMONOLAYER COVERAGES

The results of our calculation for submonolayer coverages are presented in Fig. 3. In this figure we show results for the case of two separate defects, one a donor and the other an acceptor, in equal concentration. The donor level is taken to be 0.5 eV from the valence-band edge, and the acceptor level is taken to be 0.75 eV from the valence-band edge. These values were picked to be consistent with those obtained by Spicer *et al.*[7]; however, the qualitative behavior of the results does not depend on the specific energy levels as long as the defect acceptor level is higher than the defect donor level. If not, compensation occurs, and the n- and p-type lines would be closer. The behavior of the Fermi level with increasing density of surface defects is qualitatively like that obtained by Spicer *et al.*[7] and the pinning position, i.e., the position of the Fermi level for n- and p-type semiconductors differs by a substantial amount. For the case considered in Fig. 3 only the acceptor defects are being charged, and the donors are

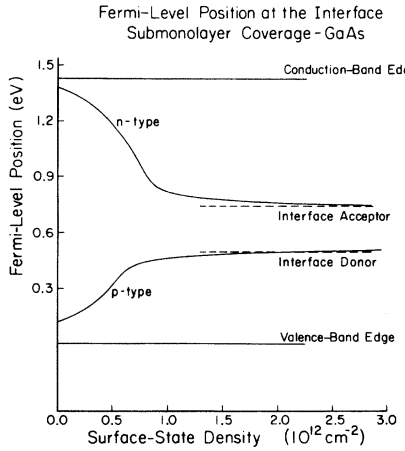

FIG. 3. Fermi-level position vs defect density for the metal submonolayer coverage model (clean surface with defects). The concentration of surface acceptors and donors was assumed equal, and their ionization energies were picked from results of Spicer *et al.* (Ref. 7).

neutral for the n-type semiconductor. For the p-type semiconductor the situation is reversed. Thus for n-type semiconductor the Fermi-level position is determined by the defect acceptor level, and for p-type semiconductor, it is determined by the defect donors.[14] Since these energy levels will, in general, be different, the Fermi-level pins at different energies for n- and p-type semiconductors. At defect densities of $\sim 10^{12}$ cm^{-2}, the Fermi-level position is seen to be largely stabilized.[14]

V. RESULTS FOR BULK METAL–SEMICONDUCTOR INTERFACE

To indicate the role of different types of defects, we have treated a number of different spatial arrangements and possible charge states for the defects.

A. Defects with two charge states

The simplest case is an interface state which has only two charge states, neutral or charged. We have examined the role of spatial distribution by considering all the defects to be located in a single plane at $x=d$ and all the defects to be located on two separate planes. The defect density is varied from 0 to 10^{14} cm^{-2}. In the case where all the defects are located in a single plane at $x=d$, Eq. (3.3) becomes

$$\Phi_m+\Delta V_m-\chi=-\eta(d)+\frac{q^2d}{\epsilon\epsilon_0}(\sigma^+(\eta(d))+Q_s(\eta(d)))\ ,$$

where q is the electron charge (positive), d is the interface layer width, ϵ is the dielectric constant in this layer, Q_s is the number of charges per unit area in the semiconductor bulk, and σ^+ is the density of (positively) ionized defects. σ^+ will be positive if the defects are all donors and negative if these defects are acceptors. In our calculations we assume all the two charge states to be donors, and a similar calculation for acceptors is straightforward and will not be presented here. $\sigma^+(\eta)$ can be calculated if we know the ionization energy of the interface state. Suppose

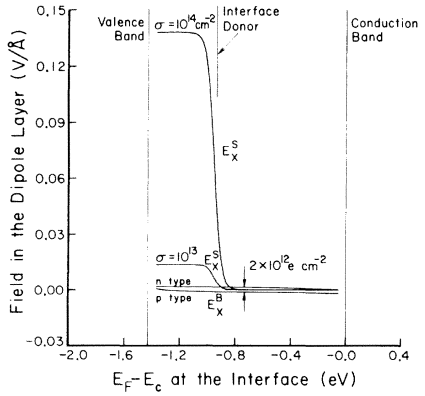

FIG. 4. Electric field in the dipole layer vs Fermi-level position in the gap at $x=d$, for defect denisties of 10^{13}–10^{14} cm^{-2}. The lines n type and p type correspond to the bulk contributions of n- or p-type GaAs having impurity concentration of 10^{17} cm^{-3}. The E_x^s lines correspond to the defects contribution, for two defect densities (σ). The interface defects in this case were chosen to be donors.

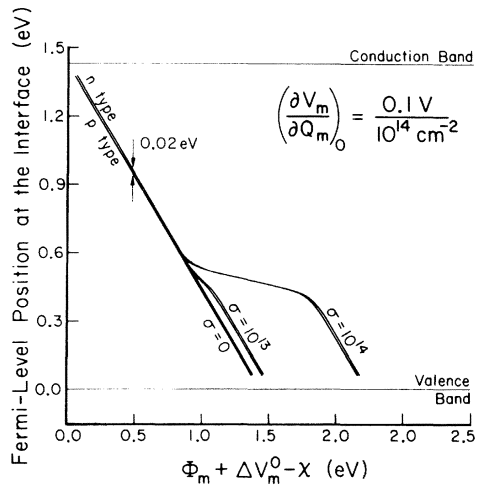

FIG. 5. Fermi-level position in the gap at $x=d$ vs $\Phi_m+\Delta V_m^0-\chi$, for three values of the defect density. The metal response $(\partial\Delta V_m/\partial Q_m)_0$ is (0.1 eV)/(10^{14} e cm^{-2}).

$(\mu-E_s)$ is the energy required to remove an electron from a neutral interface state and put it at the Fermi level, then the density of (positively) charged defects is given by (3.1).

Figure 4 shows $\mathscr{E}_x^s$ and $\mathscr{E}_x^b$, the interface and bulk contributions to the field in the interface layer, (or σ^+ and Q_s, the corresponding number of charges) as a function of η, the relative Fermi-level position, for two different values of defect density: $\sigma=10^{13}$ and 10^{14} cm^{-2}. The interface energy level appearing in this figure was chosen to agree with Spicer's value for GaAs defect donor,[7] and g_s was arbitrarily chosen to be equal to 2. The two lines denoted by n type and p type are the corresponding $\mathscr{E}_x^b$ for n- and p-type semiconductors. The doping concentration in each case is $N_D=N_A=10^{17}$ cm^{-3}. It should be noted in Fig. 4 that the electric field in the interface region is very large, on the order of 0.1 V/Å, when $\sigma=10^{14}$ cm^{-2}. It is clearly seen that at defect density of 10^{14} cm^{-2}, the defects contribution to the field dominates the depletion charge contribution. At a defect density of 10^{13} cm^{-2}, the two contributions are comparable.

In Fig. 5 we show η vs $\Phi_m+\Delta V_m^0-\chi$ for the same σ's and Q_s's that we used in Fig. 4. ΔV_m was calculated from (2.2) using

$$\left[\frac{\partial\Delta V_m}{\partial Q_m}\right]_0=\frac{0.1\text{ eV}}{10^{14}e\text{ cm}^{-2}}\ .$$

In all three plots the interface layer was assumed to be 5 Å wide. In Fig. 5, one can see three pairs of lines. The upper one in each pair corresponds to n-type and the lower one to p-type bulk semiconductor. The defect density σ varies from $\sigma=0$, to $\sigma=10^{13}$ cm^{-2} and $\sigma=10^{14}$ cm^{-2}.

For $\sigma=0$, the two lines are almost exactly straight lines with slope equal to -1 corresponding to the equation

$$\eta(d)=\chi-\Phi_m-\Delta V_m^0\ .$$

If $\Delta V_m=0$, this would be the curve predicted by the classical Schottky model assuming no interface layer between the metal and the semiconductor. The slopes of the two lines in Fig. 5(a) are approximately -1, since the change in ΔV_m is negligible in this case. We can estimate the

variation in ΔV_m for each pair of lines using Eq. (2.2). ΔQ_m is seen in Fig. 4 to change by not more than 2×10^{12} $e\,\mathrm{cm}^{-2}$ and so ΔV_m changes by not more than 2 meV throughout the whole range of the $\sigma=0$ pair. Knowing that the maximal slope of each pair of lines in Fig. 5 is -1, we will be able to estimate $\Delta\eta$, the difference in Fermi-level position at the interface, by estimating $\Delta(\Phi_m+\Delta V_m^0-\chi)$. (They will be approximately equal to each other.)

Using the abrupt (full depletion) model, we can estimate this difference to be

$$\Delta\eta\approx\Delta(\Phi_m+\Delta V_m^0-\chi)=\frac{q^2d\Delta Q_s}{\epsilon\epsilon_0}<2qd\left[\frac{N_D(Ec-Ev)}{\epsilon\epsilon_0}\right]^{1/2}. \quad (5.1)$$

The reason that $\Delta\eta\neq0$ even for $\sigma=0$ is partially an artifact of our model, even though we do not have any dipoles, we nevertheless assume an interface layer with a finite width. $\Delta\eta(0)\approx0$ but at $x=d$, $\Delta\eta(d)\neq0$ due to different band bending in the 5-Å layer.

The maximal difference in Fermi-level position occurs when there is no pinning, that is when the slopes of the n- and p-type lines are approximately -1. To see this note in the $\sigma=10^{14}$ lines that when E_j approaches the defect level, the defects charge rapidly, and both n- and p-type lines are shifted in parallel to the right. The low slope of the two lines around E_s reduces $\Delta\eta$ between the n and p type. Using (5.1) we can estimate the maximum difference to be ~3 meV/Å times the effective interface layer width for GaAs, when doping concentration is 10^{17} cm^{-3}, and $(\partial\Delta V_m/\partial Q_m)=0$. When $(\partial\Delta V_m/\partial Q_m)=(0.1\text{ eV})/(10^{14}\,e\,\mathrm{cm}^{-2})$, there will be an extra contribution of 2 meV for GaAs to this difference. As one can see in this figure, the range of $\Phi_m+\Delta V_m^0-\chi$ in which η is approximately constant is quite small for $\sigma=10^{13}$ cm^{-2}. In other words, defect density of $\sigma=10^{13}$ cm^{-2} is not quite enough to pin the Fermi level. When $\sigma=10^{14}$ cm^{-2}, however, the range of $\Phi_m+\Delta V_m^0-\chi$ in which $\Delta\eta$ is constant is quite large, on the order of 1 eV.

Figure 6 is the same as Fig. 5 but with what is likely to be a very large value:

$$\left[\frac{\partial\Delta V_m}{\partial Q_m}\right]_0=\frac{1\text{ eV}}{10^{14}\,e\,\mathrm{cm}^{-2}}.$$

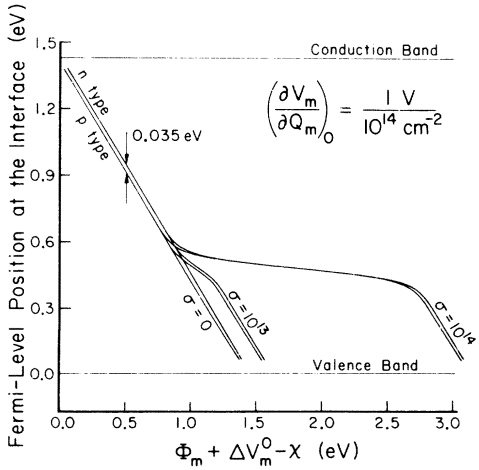

FIG. 6. Fermi-level position in the gap at $x=d$ vs $\Phi_m+\Delta V_m^0-\chi$ for three values of the defect density. The metal response $(\partial\Delta V_m/\partial Q_m)_0$ is $(1.0\text{ eV})/(10^{14}\,e\,\mathrm{cm}^{-2})$.

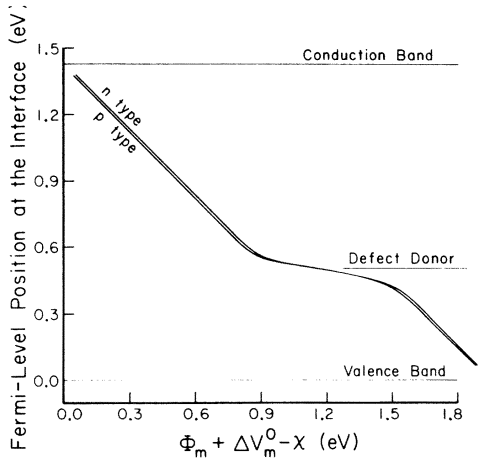

FIG. 7. Fermi-level position in the gap at $x=d$ vs $\Phi_m+\Delta V_m^0-\chi$ when the 10^{14} defects per cm^2 are divided between two planes: half on $x=d$ and half on $x=d/2$. $(\partial\Delta V_m/\partial Q_m)_0=0$.

Here exactly the same conclusions can be drawn regarding the fact that 10^{13} cm^{-2} are not enough defects to pin the Fermi level, while 10^{14} are more than enough. When there is no pinning, $\Delta\eta$, the difference in Fermi-level position between n and p type is still small, though it is bigger than before and is now ~0.035 eV.

It is important to understand what causes the pinning. Depending on the relative Fermi-level position, the defects will charge from 0 to $\pm q\sigma$ ($+$ for donor type state, $-$ for acceptor) when we change Φ_m. The source of this charge will be mainly the metal, and since the metal is thick enough, it can supply the necessary charges with relatively small change in the boundary conditions due to loss of charge. The field in the interface layer due to the charged defects will tend to diminish the effect of changes in Φ_m, but effective pinning will happen only when there are enough defects around, that is when (in eV)

$$\frac{q^2\sigma}{\epsilon\epsilon_0}d\approx1.$$

For d approximately equal to a few Å and $\epsilon\approx10$, that implies $\sigma\approx10^{14}$ cm^{-2}.

To explore the role of spatial distribution we have treated the case when the defects lie on two parallel planes. We assume that half the defects are at a distance d, and the other half are $d/2$ away from the metal. d was taken as before to be 5 Å. In that case one can integrate (2.1) from $d/2$ to d and find $\eta(d/2)$. Then find out how many charges will be on that plane; substitute back into (3.3) and find $\Phi_m+\Delta V_m^0-\chi$. In Fig. 7 we display the Fermi-level position at the interface for this case. The doping is 10^{17} cm^{-3}, and we have 10^{14} cm^{-2} defects, half of them on the plane $x=d$ and half on the plane $x=d/2$. The defects on the plane closer to the metal charge more rapidly when we decrease $\eta(d)$, so the width in terms of $\eta(d)$ of the pinning region will not increase. However, the width in terms of

$\Phi_m+\Delta V_m^0-\chi$ will decrease since the integration (3.3) is done for the same amount of charged defects only to half the distance, while in the other half, we have contributions to the field from only half the defects. This effect introduces an effective interface layer width, as can be seen in Fig. 7. For simplicitiy, we have only examined the case of $(\partial\Delta V_m/\partial Q)_0=0$. It is apparent that the effect of putting all the defects on a single plane is not a crucial assumption and was made only for convenience.

B. Defects with three charge states

Defects that have three charge states, for example, negative, neutral, and positive, were suggested as a possible mechanism for pinning the Fermi level.[6] The statistics of such defects is different than that of two independent defects, one a donor and the other an acceptor, localized on different centers. The difference comes from the fact that the three-charge-state defect has only one neutral state. The independent donor and acceptor can be neutral in two distinct ways, namely, when both are neutral, or when both are charged. In the case of a single type of defect, with three charge states, we can write down the equilibria equations,

$$S^-\rightarrow S^0+e^-,\quad \Delta E=\mu-E_s^D,$$
$$S^0\rightarrow S^++e^-,\quad \Delta E=\mu-E_s^A,$$

From these equations we can calculate the ratios of positive and negative to neutral defects.

$$\frac{\sigma^+}{\sigma^0}=\frac{1}{g_s^D}\exp\left[\frac{E_s^D-\mu}{k_BT}\right],$$
$$\frac{\sigma^-}{\sigma^0}=\frac{1}{g_s^A}\exp\left[\frac{\mu-E_s^A}{k_BT}\right].$$

Here $\sigma^+,\sigma^0,\sigma^-$ are the surface densities of positive, neutral, and negative defects. g_s^A and g_s^D are the corresponding degeneracy factors, respectively. Since the total defect density σ must satisfy

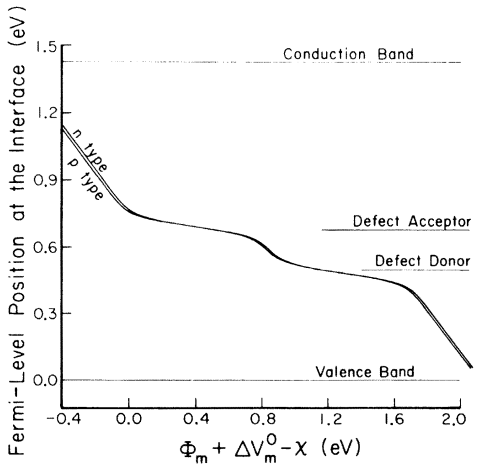

FIG. 8. Fermi-level position in the gap at $x=d$ vs $\Phi_m+\Delta V_m^0-\chi$ for the case when each defect acts both as a donor and an acceptor, and the donor level is *lower* than the acceptor level.

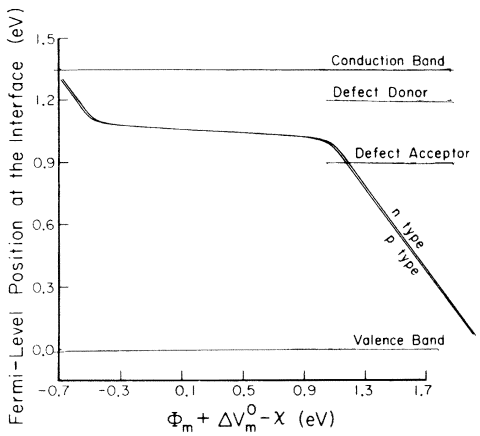

FIG. 9. Fermi-level position in the gap at $x=d$ vs $\Phi_m+\Delta V_m^0-\chi$ for the case when each defect acts both as a donor and an acceptor and the donor level is *higher* than the acceptor level.

$$\sigma^-+\sigma^0+\sigma^+=\sigma,$$

we have

$$\sigma^0=\frac{\sigma}{\frac{1}{g_s^A}\exp\left[\frac{\mu-E_s^A}{k_BT}\right]+1+\frac{1}{g_s^D}\exp\left[\frac{E_s^D-\mu}{k_BT}\right]} \tag{5.2}$$

with similar expressions for σ^+ and σ^-.

We see that there are two different cases, one when $E_s^D>E_s^A$ and the other when $E_s^D<E_s^A$. If $E_s^D<E_s^A$, than for values of μ such that $E_s^D<\mu<E_s^A$, most defects are neutral. (This case would correspond to GaAs according to Spicer *et al.*[7] and E_s^A and E_s^D were picked to be their values of 0.75 and 0.5 eV from the valence-band edge.) This region of μ separates two rapid charging regions around E_s^D and E_s^A as can be seen in Fig. 8, and the whole system behaves as if there are two types of two-charge states (one neutral or negative, the other neutral or positive) which act independently, rather than a single three-charge state. There will be, therefore, two different pinning positions depending on the metal work function, but independent of the semiconductor type.

If $E_s^A<E_s^D$, the denominator of (5.2) will be large for every possible value of η ($\sigma^0\ll\sigma$ for every η). S^0 will be unstable,[15] i.e., S^+ will capture and S^- will release electrons in pairs, and there will be only one value of $\eta(d)$ around which rapid charging will occur, as can be seen in Fig. 9. E_s^A and E_s^D were picked to be 0.9 and 1.2 eV, respectively, from the valence-band edge, and a band gap of 1.35 eV corresponding to InP was used. Since the defects change their charges in quantities of $2e$ instead e, as well as reversing the sign, it is a very efficient way to pin the Fermi level, and if there are enough defects, $\eta(d)$ can be pinned for any possible value of $\Phi_m+\Delta V_m^0-\chi$ at the same place.

V. SUMMARY AND DISCUSSION

We have investigated the phenomenon of Fermi-level pinning at the semiconductor-metal interface by charging defects. We estimated the defect density required to pin

the Fermi level to be $\sim 10^{14}$ cm^{-2}, assuming that most of these defects are localized not more than few Å from the metal and that the electrostatic potential at the metal surface responds to excess charges by not more than (1 eV)/(10^{14} e cm^{-2}). Under these conditions, we also estimated the difference in Fermi-level position between n-type and p-type semiconductor to be less than 0.05 eV. This difference is the sum of two terms, one due to different metal response, the other due to different band bending in the interface region. Since the maximal difference in bulk semiconductor charge between n- and p-type GaAs for a given Fermi-level position is about 2×10^{12} e cm^{-2} for doping concentration 10^{17} cm^{-3}, we can estimate the metal response term to be at most 0.02 eV and the "interface dipole layer" term to be not more than 3 meV/Å. This difference was shown to be the maximum possible one and to occur only when there is no pinning. When there is pinning, this difference is smaller. For the submonolayer coverage we have found that defect density of only 10^{12} cm^{-2} will pin the Fermi level and that the Fermi-level position for n- and p-type semiconductors can differ by a substantial amount.

The difference between the results of submonolayer coverage and thick metallic coverage, regarding both the difference in Fermi-level position and the defect density required to pin the Fermi level could be attributed to the fact that these are *two different pinning mechanisms,*[6] originating from the main source of charge that balances the charge on the defects. With 10^{11}–10^{12} defects per cm^2 but *without* substantial metal coverage, these charges can come almost entirely from the semiconductor bulk. The requirement of total charge neutrality will determine the Fermi-level position, and the Fermi-level position can differ substantially between n- and p-type semiconductors.

For the case of a bulk metal, the charge in the interface layer is balanced mainly by charge in the metal, setting up a very thin dipole layer. The charge in the metal can easily respond in such a way as to balance the depletion charge, and, hence the n- and p-type pinning positions are very similar.

ACKNOWLEDGMENTS

We would like to acknowledge A. Nedoluha and H. Weider for useful discussion. This work was supported in part by the U. S. Office of Naval Research under Naval Contract No. N00014-79-C-0797.

[1]W. E. Spicer, P. W. Chye, P. R. Skeath, C. Y. Su, and I. Lindau, J. Vac. Sci. Technol. 16, 1422 (1979).

[2]P. Skeath, I. Lindau, P. W. Chye, C. Y. Su, and W. E. Spicer, J. Vac. Sci. Technol. 16, 1143 (1979).

[3]W. E. Spicer, P. Skeath, C. Y. Su, and P. Chye, Phys. Rev. Lett. 44, 420 (1980).

[4]R. W. Grant, J. R. Waldrop, S. P. Kowalczyk, and E. A. Kraut, J. Vac. Sci Technol. 19, 477 (1981).

[5]C. A. Mead, Solid State Electron, 9, 1023 (1966).

[6]M. S. Daw and D. L. Smith, J. Vac. Sci. Technol. 17, 1028 (1980).

[7]W. E. Spicer, I. Lindau, P. Skeath, and C. Y. Su, J. Vac. Sci. Technol. 17, 1019 (1980).

[8]M. S. Daw and D. L. Smith, Phys. Rev. B 20, 5150 (1979).

[9]J. Bardeen, Phys. Rev. 49, 653 (1936).

[10]M. Schlüter, J. Vac. Sci. Technol. 15, 1374 (1978).

[11]G. D. Mahan and W. L. Schaich, Phys. Rev. B 10, 2647 (1974).

[12]S. M. Sze, *Physics of Semiconductor Devices* (Wiley-Interscience, New York, 1969), p. 25.

[13]The real work function is given by $q\,\Delta\phi-\mu$. Here $\Delta\phi$ is the electrostatic potential difference between a point far outside, and a point deep inside the metal. This quantity depends on the details of the charge rearrangement near the surface. μ is the bulk chemical potential, that involves the average electron kinetic energy, as well as exchange and correlation effects. This quantity depends only on the bulk properties, and does not contain any surface effects. Unfortunately, μ does not yield itself to a direct measurement, which is probably the reason why the difference in work functions is generally used instead of the difference in chemical potentials. Despite the difficulties in its measurement, μ is a well-defined thermodynamical quantity. In this paper we adhere to the more common $\Phi_m-\chi$ notation, which should mean the difference in chemical potentials, rather than difference in work functions. The conclusions of our paper are independent of the exact values, and the only condition we use, is that the integral (3.2) is a function of the bulk properties of the metal and the semiconductor, and that its value will change by not more than few electron volts upon changing metals.

[14]One should keep in mind that due to thermal effects, some small fraction of the defect donors are charged even when the Fermi-level position is above the donor level, and some defect acceptors are charged when the Fermi level is below the defect acceptor level. This will result in partial compensation, and therefore when the defect density increases, eventually the n- and p-type lines will merge. The required defect density for this merger is much higher and is temperature dependent.

[15]G. A. Baraff, E. O. Kane, and M. Schlüter, Phys. Rev. Lett. 43, 956 (1979).