PHYSICAL REVIEW B VOLUME 43, NUMBER 5 15 FEBRUARY 1991-I

Defect-state occupation, Fermi-level pinning, and illumination effects on free semiconductor surfaces

Robert B. Darling
Solid-State Laboratory, Department of Electrical Engineering, FT-10, University of Washington, Seattle, Washington 98195

(Received 24 September 1990)

A comprehensive physical model for equilibrium and illuminated free semiconductor surfaces interacting with an arbitrary number of monoenergetic surface and bulk defect states is presented. The position of the surface Fermi energy is shown to be determined by the overall balance of surface charge and surface electric field, and is not merely set by alignment to the energy of the highest-density surface state. Boundary conditions are derived for nonequilibrium device analysis, which includes the full electrostatic and recombinative details of the surface states. An approximate analytical solution is given for the semiconductor transport equations with these boundary conditions and this model is used to describe illumination effects on the free surface. The solution can be represented in a graphical form, which allows the most significant contributions to be easily visualized. Fermi-level pinning on free surfaces is then classified according to the primary contributions to the charge balance and also according to the sensitivity of the surface potential to charge perturbations. The occupation factors for the surface states under illumination conditions are derived from the analytic model and are shown to both shift and saturate with increasing illumination intensity. This allows the origin of free-surface photovoltage, depletion-edge contraction, and band flattening to be reinterpreted as arising from a change in surface-state occupation rather than from a quenching of the depletion-region field by free electrons and holes.

I. INTRODUCTION

Free semiconductor surfaces may be defined as those in which the potential and carrier densities are not constrained to a specific value by the external environment of the media. From a one-dimensional circuit viewpoint, the potential of the surface may be considered as "floating" due to the lack of any externally applied fields or contact metallizations. Free semiconductor surfaces consist ideally of a semiconductor-vacuum interface, or at minimum a semiconductor-gas interface, the most common case obviously being air. In practice, the surface usually consists of an oxide layer that is sufficiently thin to assume the same potential as that of the semiconductor surface.

In the analysis of semiconductor devices, free surfaces pose a particularly subtle problem because the true boundary conditions associated with a free surface do not fall into the usual Dirichlet or Neumann type. The boundary conditions most commonly used to model free surfaces are simplified minority-carrier surface recombination velocities that merely set the excess carrier densities proportional to the surface recombinative currents. These do not require the electron and hole surface current densities to sum to zero and thus violate charge conservation. In addition, any charge that is bound to the surface defect levels is generally ignored, thus violating Gauss' law. A consequence of this is that these simple models cannot account for the well-established phenomenon of surface depletion.

Free semiconductor surfaces are found in numerous electronic devices, such as in the interelectrode gaps of field-effect transistors, in between the electrodes of photoconductive elements, and around the periphery of mesa-isolated structures. Despite their common occurrence, free surfaces have not been studied in sufficient detail to produce a general model for either the boundary conditions or the overall behavior that includes the many contributions from bulk and surface defects. Surface states lying within the energy-band gap play a dominant role in the electrical behavior of the surface and arise from a wide range of sources. These include (i) Shockley-Tamm surface states that are associated with the simple termination of the periodic crystal lattice,[1,2] (ii) surface defect states that, as proposed by Brattain and Bardeen,[3] may act as localized donor or acceptor levels, and (iii) immobile adsorbed ions that can contribute a fixed charge density to the surface.

A very complete analysis of semiconductor surfaces was presented by Garrett and Brattain[4] almost 35 years ago, which has served well as a general reference for the spatial shape of the depletion region, surface conductivity, field effects, and photovoltage effects. The principal limitations to this analysis have been the restriction to only shallow and thus fully ionized impurities, the assumption that the surface states lie sufficiently close to the band edges that they may be described in terms of quasi-Fermi-potentials, and a unipolar nonequilibrium analysis that only includes the flow of minority carriers. Bardeen[5] originally noted that partial filling of the surface states is necessary to account for Fermi-level pinning of the surface, such as that experimentally found by Meyerhof.[6] The Brattain and Bardeen model[3] was then proposed whereby the surface defect states must be of either donor or acceptor type to ascribe the proper surface charge to the partial filling of the states. Despite the long-standing acceptance of this model, few if any analyses of the semiconductor surface have included the full Fermi-Dirac equilibrium statistics or nonequilibrium rate terms and occupation factors that this model implies. While a continuous distribution of surface states within the energy band gap is clearly the more general case, the Brattain-Bardeen model utilizes monoenergetic surface states, which may be combined in any multiplicity. A uniform continuous distribution of surface states versus energy has been used with good success in various models of the metal-semiconductor interface,[7] but the majority of experimental measurements have indicated that free semiconductor surface states are localized to a rather narrow band of energies.[8–12] Hence, the monoenergetic surface-state assumption of Brattain and Bardeen will be continued here. Van Roosbroeck[13] has analyzed a semi-infinite semiconductor domain using Green's functions for a point-source excitation of electron-hole pair generation, but this did not include any detailed structure of the surface states or bulk impurities. Kingston and Neustadter[14] have also given calculations of surface potential and carrier density, but only for the equilibrium case of fully ionized donor and acceptor states.

This paper presents a generalized model of the free surface that includes the full electronic detail of both surface and bulk defect states. Particular attention is paid in the development to practical calculation procedures that can be utilized in the construction of numerical or analytical device models. Section II of this paper presents the general model for the free surface under both equilibrium and illumination conditions that is assembled from simple electrostatics and current flow considerations of the contributing defect states. Some general features of this model are examined in Sec. III, particularly in regard to electrostatic limitations to Fermi-level pinning, surface-state repopulation, and band flattening effects with illumination. Final conclusions are given in Sec. IV.

II. GENERAL MODEL OF THE FREE SEMICONDUCTOR SURFACE

A. Equilibrium

Consider a semi-infinite semiconductor along the positive z axis with its surface at $z=0$ as shown in Fig. 1. In the following, a subscript of o will indicate equilibrium values, a subscript of s will indicate a surface quantity, a d for a donor quantity, an a for an acceptor quantity, and a b for a bulk quantity existing far enough away from the surface so as to be unaffected by it ($z \rightarrow \infty$). Within the bulk regions of the semiconductor where no band bending from the surface exists, the equilibrium ionization of donor and acceptor levels is given by standard Fermi-Dirac occupation factors,

$$N_d^+ = \frac{N_d}{1+g_d e^{(E_{Fb}-E_d)/kT}} = \frac{N_d}{1+F_d} \ , \tag{1a}$$

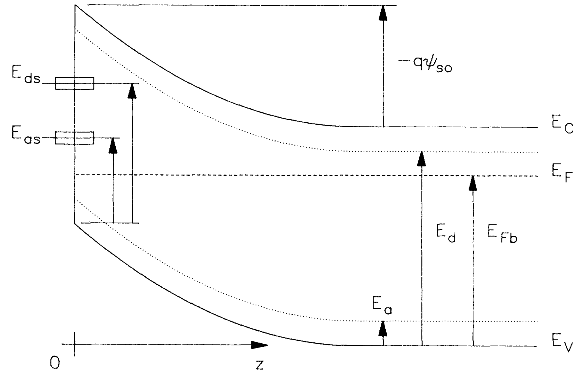

FIG. 1. Energy-band diagram of the free semiconductor surface.

$$N_a^- = \frac{N_a}{1+g_a e^{(E_a-E_{Fb})/kT}} = \frac{N_a}{1+F_a} \ , \tag{1b}$$

where g_d and g_a are the ratios of the filled-state–to–empty-state degeneracies of the levels and F_d and F_a are defined above for convenience. For nondegenerate carrier concentrations, the electron and hole densities are similarly

$$n = N_c e^{(E_{Fb}-E_c)/kT} = N_c e^{(E_{Fb}-E_v-E_g)/kT} \ , \tag{2a}$$

$$p = N_v e^{(E_v-E_{Fb})/kT} \ , \tag{2b}$$

where all energies unless stated otherwise are measured upwards from the valence-band maximum.

The position of the Fermi energy in the neutral bulk substrate can be obtained by setting the charge density

$$\rho = q(p-n+N_d^+-N_a^-) \tag{3}$$

to zero, which yields a transcendental equation of the form

$$\rho(E_{Fb})=0 \ . \tag{4}$$

Gaylord and Linxwiler[15] have shown that a numerical bisection routine[16] provides an efficient method for solving this system that allows for all ranges of temperature, both deep and shallow impurity levels, and any range of semiconductor band gap. Numerical bisection routines are used for finding all roots to transcendental equations in this work because of their guaranteed convergence when the search range brackets the root. Multiple donor and acceptor levels may be included in this description, but only a single donor and acceptor are carried through the equations for simplicity. It must be stressed that it is necessary to allow for only partial ionization of the impurity levels, particularly when the level is near the middle of the energy band gap. Such deep-lying donor and/or acceptor levels are necessary for the proper description of compensated semi-insulating materials such as GaAs and InP.

A central idea to the physics of free surfaces is that any band bending within the semiconductor must be supported by an appropriate surface charge in order to satisfy Gauss' law. The most interesting source of such a surface charge is the presence of surface donor or acceptor levels, which may arise from vacancies, interstitials, dislocations, surface oxides, surface reconstructions, or adsorbed ions. For compound semiconductors, the presence of antisite defects is another origin of surface donor and acceptor levels that is presently thought to influence the degree and position of Fermi-level pinning.[11,12] Regardless of their physical origin, the equilibrium ionization of surface donors and acceptors is again given by the Fermi-Dirac occupation factors of

$$N_{ds}^{+}=\frac{N_{ds}}{1+g_{ds}e^{(E_{Fb}-E_{ds}+q\psi_{so})/kT}}\,, \tag{5a}$$

$$N_{as}^{-}=\frac{N_{as}}{1+g_{as}e^{(E_{as}-E_{Fb}-q\psi_{so})/kT}}\,, \tag{5b}$$

where ψ_{so} is the electrostatic potential evaluated at the surface. The electrostatic potential is referenced to a bulk substrate value of $\psi(\infty)=\psi_b=0$; thus, ψ_{so} gives a direct measure of the amount and polarity of band bending at the surface. In addition to surface donor and acceptor states, the surface charge may also be supported by adsorbed ions whose charge remains fixed. While less interesting physically, these adatoms may be quite abundant and influential in practice. The total surface-charge density is then

$$\rho_{so}=qN_{ds}^{+}-qN_{as}^{-}+qZN_{fs}\,, \tag{6}$$

where Z is the ionization number of the fixed surface ions. Gauss' law requires that the above surface charge terminate the fluxlines of a surface field intensity of

$$E_{zo}(0)=E_{zso}=\frac{\rho_{so}}{\epsilon}\,, \tag{7}$$

which is the same electrostatic field as that which establishes the band bending in the semiconductor. Thus, the equilibria of the free surface may be described as a balancing of the surface electric field due to band bending with that due to the occupation of the surface states and their associated charge. This assumes that there is no normal electric field external to the semiconductor.

Solving the equilibrium free surface problem amounts to finding the surface electric field due to the band bending and equating it to Eq. (7). Within the region of band bending, the changing position of the Fermi energy alters the occupation of the donor and acceptor levels so that Eqs. (1a) and (1b) become

$$N_d^{+}=\frac{1}{1+F_d e^{\beta\psi_o}}\,, \tag{8a}$$

$$N_a^{-}=\frac{N_a}{1+F_a e^{-\beta\psi_o}}\,, \tag{8b}$$

where $\beta=q/kT$. The electron and hole densities may similarly be referenced to their bulk substrate values to put Poisson's equation into the form of

$$\frac{d^2\psi_o}{dz^2}=-\frac{q}{\epsilon}\left[p_{bo}e^{-\beta\psi_o}-n_{bo}e^{\beta\psi_o}+\frac{N_d}{1+F_d e^{\beta\psi_o}}-\frac{N_a}{1+F_a e^{-\beta\psi_o}}\right], \tag{9}$$

where $n_{bo}=n_o(\infty)$ and $p_{bo}=p_o(\infty)$. This may be integrated across the surface band-bending region from $z=\infty$ to $z=0$ by manipulating it into the form of

$$\int_0^{\partial\psi_o/\partial z}\left[\frac{\partial\psi_o}{\partial z}\right]d\left[\frac{\partial\psi_o}{\partial z}\right]=\int_0^{\psi_o}-\frac{\rho}{\epsilon}\,d\psi_o\,. \tag{10}$$

If the bulk impurities are uniformly distributed, the integration may be performed exactly to yield

$$E_{zo}^{2}(z)=\frac{2q}{\epsilon\beta}\left[p_{bo}(e^{-\beta\psi_o}-1)+n_{bo}(e^{\beta\psi_o}-1)+N_d\ln\left[\frac{e^{-\beta\psi_o}+F_d}{1+F_d}\right]+N_a\ln\left[\frac{e^{\beta\psi_o}+F_a}{1+F_a}\right]\right], \tag{11}$$

a result originally obtained by Garrett and Brattain,[4] but extended here to include the case of partially ionized impurities. The above result may be used to find the potential profile, since it specifies the electric field (the derivative of the potential) as a closed-form function of the potential. For the present case of the semiconductor residing on the positive z axis, the sign of the electric field should be chosen to be the same as that of the potential. The above result may be used to find the equilibrium point for a free surface via

$$E_{zso}(\psi_{so})=\frac{\rho_{so}(\psi_{so})}{\epsilon}\,, \tag{12}$$

where the left-hand side is obtained from Eq. (11) and the right-hand side from Eq. (6). This yields a transcendental equation for the equilibrium surface potential ψ_{so} that may be solved using a numerical bisection routine. Once ψ_{so} has been found, all of the other equilibrium variables may be directly calculated in turn. It is worth mentioning that this solution procedure is exact for the case of steady state to the accuracy of the numerical bisection routine, and in its most general form allows for any number of surface and bulk impurity levels, each at possibly different densities, energies, and degeneracies.

B. Nonequilibrium boundary conditions for device analysis

When a free semiconductor surface is illuminated with ionizing radiation or otherwise driven away from equilibrium, the analytical tool of a constant Fermi energy can no longer be used to describe the state of the system and rate and transport equations must be solved instead. Within the bulk region of the semiconductor, it becomes necessary to solve the five basic semiconductor device equations, shown below in steady-state, one-dimensional form:

$$J_{nz} = q\mu_n n E_z + qD_n \frac{dn}{dz} , \tag{13a}$$

$$J_{pz} = q\mu_p p E_z - qD_p \frac{dp}{dz} , \tag{13b}$$

$$0 = G(z) - U(n,p) + \frac{1}{q}\frac{dJ_{nz}}{dz} , \tag{13c}$$

$$0 = G(z) - U(n,p) - \frac{1}{q}\frac{dJ_{pz}}{dz} , \tag{13d}$$

$$\frac{d^2\psi}{dz^2} = -\frac{dE_z}{dz} = -\frac{\rho}{\epsilon} = \frac{q}{\epsilon}(n - p - N_d^+ + N_a^-) . \tag{13e}$$

The electron and hole current densities are J_{nz} and J_{pz}, the electron and hole mobilities are μ_n and μ_p, the electron and hole diffusivities are D_n and D_p, $G(z)$ is the generation rate of electron-hole pairs arising from the incident illumination, and $U(n,p)$ is the recombination rate of electron-hole pairs within the bulk regions of the semiconductor.

The essential problem for one-dimensional semiconductor device analysis is to solve the above system for the variables of $\{n,p,J_{nz},J_{pz},E_z\}$ over a fixed domain of $[0,L]$. As such, the system is one of five coupled ordinary first-order nonlinear differential equations requiring exactly five boundary conditions for a unique solution. Poisson's equation (13e) is normally solved as an ordinary second-order differential equation by specifying two boundary conditions on the electrostatic potential, $\psi(0)$ and $\psi(L)$, instead of a single condition on the electric field E_z. Obviously, since the potential is referenced to an arbitrary level, the two boundary conditions on ψ merely set the applied bias across the device $[\psi(0)-\psi(L)]$ and the arbitrary reference point of the potential. The remaining four boundary conditions on the variables of $\{n,p,J_{nz},J_{pz}\}$ normally specify $\{n(0),p(0),n(L),p(L)\}$ for a device with electrical contacts at each end of the domain. This is the traditional case for a voltage-controlled device for which the current is to be determined as a function of the applied bias. Current-controlled devices alternatively specify the current densities at both contacts and the electric field at one point within the domain.

The situation of a free semiconductor surface introduces special complications into the analysis of a semiconductor domain, whether in one or multiple dimensions. These complications are mathematically that neither Dirichlet nor Neumann boundary conditions are satisfactory in describing the free-surface values of $\{n,p,J_{nz},J_{pz}\}$. Even Poisson's equation cannot be solved as a two-point boundary value problem because the surface value of ψ is unknown prior to the solution.

In solving steady-state, one-dimensional semiconductor problems, a useful transformation is to define sum and difference current densities

$$J_z = J_{nz} + J_{pz} , \tag{14a}$$

$$J_{dz} = J_{nz} - J_{pz} , \tag{14b}$$

from which the total current density J_z must be a constant with respect to position z in order for current continuity to hold. This effectively reduces Eqs. (13c) and (13d) into a single first-order equation and mathematically reduces the overall system to fourth order. Instead of needing to find both J_{nz} and J_{pz} as a function of z, one needs to find only J_{dz} as a function of z and J_z, which is a constant. Because of the unknown constant J_z, the number of required boundary conditions is still five.

This transformation is particularly useful for the analysis of free semiconductor surfaces, because the obvious boundary condition that the total normal current density vanishes at the surface sets the unknown constant $J_z(0)=0$. For a one-dimensional system, the total current density must therefore vanish at all parts of the domain as well, and $J_{dz} = 2J_{nz} = -2J_{pz}$. Assuming that the Einstein relations hold, the differential equations describing a free semiconductor surface may thus be phrased as the fourth-order system of

$$\frac{dn}{dz} + \beta E_z n = \frac{J_{dz}}{2qD_n} , \tag{15a}$$

$$\frac{dp}{dz} - \beta E_z p = \frac{J_{dz}}{2qD_p} , \tag{15b}$$

$$\frac{dJ_{dz}}{dz} = 2qU(n,p) - 2qG(z) , \tag{15c}$$

$$\frac{d^2\psi}{dz^2} = -\frac{dE_z}{dz} = -\frac{\rho}{\epsilon} = \frac{q}{\epsilon}(n - p - N_d^+ + N_a^-) , \tag{15d}$$

which requires exactly four boundary conditions for a unique solution. Assuming that the only free surface is located at $z=0$, two boundary conditions may immediately be imposed upon the neutral bulk substrate as $n(\infty) = n_{bo}$ and $p(\infty) = p_{b0}$. Equilibrium conditions deep into the substrate in turn require that $E_z(\infty)=0$ and $J_{dz}(\infty)=0$ as a consequence. The two remaining boundary conditions must be obtained from the surface at $z=0$ itself.

In order to derive the appropriate boundary conditions for the free surface, it is first necessary to note that recombination centers and charge traps are two different effects arising from any one given impurity level, regardless of whether it lies within a bulk region or at a surface. The kinetics and rate equations for an arbitrary midgap state have been described in detail in the important works by Shockley and Read[17] and Hall.[18] Only the results will be summarized here.

For a localized state of density N_t located at an energy of E_t within the band gap, the steady-state rate of electron-hole-pair recombination through this state is

$$U(n,p) = \frac{v_{\rm th}\sigma_n\sigma_p g_e g_f N_t (np - n_i^2)}{\sigma_n\left(g_e n + g_f n_i e^{(E_t - E_i)/kT}\right) + \sigma_p\left(g_f p + g_e n_i e^{(E_i - E_t)/kT}\right)} , \tag{16}$$

where σ_n and σ_p are the electron and hole capture cross sections, n_i is the intrinsic carrier density, E_i is the intrinsic energy, and the mean thermal carrier velocity is given by $\frac{1}{2}m^*v_{\text{th}}^2=\frac{3}{2}kT$. The degeneracy of the localized state when empty of electrons is g_e and is g_f when filled. In addition, the nonequilibrium occupation factor for electrons in the level is

$$f=\frac{\sigma_n g_e n+\sigma_p g_e n_i e^{(E_i-E_t)/kT}}{\sigma_n(g_e n+g_f n_i e^{(E_t-E_i)/kT})+\sigma_p(g_f p+g_e n_i e^{(E_i-E_t)/kT})} . \tag{17}$$

If equilibrium expressions for the carrier densities are substituted, $n=n_i e^{(E_F-E_i)/kT}$ and $p=n_i e^{(E_i-E_F)/kT}$, then Eq. (17) reduces to the usual Fermi-Dirac occupation factor with the degeneracy ratio given by the usual $g=g_f/g_e$. It is important to note that the occupation factor for these localized states changes from being a function of the Fermi energy in equilibrium to being a function of both carrier densities in nonequilibrium. For bulk recombinative centers (N_t in cm^{-3}, U in cm^{-3}/sec), the relative fluency of electron and hole capture is expressed by the low-level minority-carrier lifetimes of

$$\frac{1}{\tau_n}=v_{\text{th}}\sigma_n N_t , \tag{18a}$$

$$\frac{1}{\tau_p}=v_{\text{th}}\sigma_p N_t . \tag{18b}$$

The analogous case for surface recombinative centers is obtained by replacing N_t with N_{ts} (in cm^{-2}) and replacing n and p with their surface values n_s and p_s to yield a pair recombination rate of U_s (in cm^{-2}/sec). The relative fluency of electron and hole capture is then usually expressed in terms of low-level minority-carrier surface recombination velocities of

$$s_n=v_{\text{th}}\sigma_n N_{ts} , \tag{19a}$$

$$s_p=v_{\text{th}}\sigma_p N_{ts} . \tag{19b}$$

It should be noted that while minority-carrier lifetimes and surface recombination velocities can be used appropriately for characterization of a given level, the use of linearized minority-carrier recombinative rates invariably violates conservation of particle number when used analytically. In steady state the rate of electron recombination must equal the rate for holes; thus, it is necessary to use the general form of U and U_s above rather than a linearized rate that depends upon only one carrier species.

One feature that is commonly neglected in numerical and analytical solutions of the semiconductor device equations is that the charge on a given type of localized state changes along with the electron and hole densities that determine its occupancy. If the density of the localized states is small in comparison to n and p, then this is certainly justified. However, for the case of a free surface and other situations where deep-level charges play a dominant role, such as in semi-insulating wide-band-gap materials, the change in the localized state occupancy must be included in any nonequilibrium model. For the case of a free semiconductor surface it is therefore necessary to formulate the surface-charge density as

$$\rho(n_s,p_s)=qN_{ds}[1-f_{ds}(n_s,p_s)]-qN_{as}f_{as}(n_s,p_s)+qZN_{fs} , \tag{20}$$

where each donor and acceptor occupancy factor becomes a function of the surface electron and hole density.

As in the case of equilibrium, this surface charge must terminate the fluxlines of the surface electric field in order to satisfy Gauss' law. This leads to the nonequilibrium boundary condition of

$$E_z(0)=\frac{\rho_s(n_s,p_s)}{\epsilon} . \tag{21}$$

The remaining boundary condition on the surface is that the electron and hole flux densities incident upon the surface equal the rate of surface recombination,

$$\frac{J_{dz}(0)}{2q}=U_s(n_s,p_s) . \tag{22}$$

Equations (21) and (22) are the general nonequilibrium boundary conditions for modeling any free semiconductor surface. In equilibrium, both sides of Eq. (22) vanish while the right-hand side of Eq. (21) becomes dependent upon the single variable of the Fermi energy. These boundary conditions do not lend themselves well to either analytical or numerical procedures, because they are of neither Dirichlet nor Neumann type. However, their use is necessary if the full physics of the surface is to be represented.

C. Approximate analytical model

The solution of the free semiconductor surface problem consists of solving Eqs. (15a)–(15d) with the boundary conditions of $n(\infty)=n_{bo}$, $p(\infty)=p_{b0}$, Eq. (21), and Eq. (22). This system is too nonlinear and too closely coupled to provide any simple analytical solutions. However, some accurate approximate solutions can be obtained by assuming spatial profiles for several of the variables. Some exact analytical manipulations make these assumptions and restrictions more obvious.

Integral forms of Eqs. (15a)–(15c) may be readily obtained as

$$n(z)=n_{bo}e^{\beta\psi(z)}-e^{\beta\psi(z)}\frac{1}{D_n}\int_z^\infty \frac{J_{dz}(z')}{2q}e^{-\beta\psi(z')}dz' , \tag{23a}$$

$$p(z)=p_{bo}e^{-\beta\psi(z)}-e^{-\beta\psi(z)}\frac{1}{D_p}\int_z^\infty \frac{J_{dz}(z')}{2q}e^{\beta\psi(z')}dz' , \tag{23b}$$

$$\frac{J_{dz}(z)}{2q}=\int_z^{\infty} G(z')dz'-\int_z^{\infty} U(n,p)dz' \ . \tag{23c}$$

Equations (23a) and (23b) are obtained through the use of an integrating factor $\exp\pm\beta\int_0^z E_z(z')dz' = \exp\mp\beta[\psi(z)-\psi(0)]$ and referencing the electrostatic potential to the bulk substrate, $\psi_b=\psi(\infty)=0$.

Evaluating Eq. (23c) at $z=0$ allows the conservation of charge requirement to be easily interpreted,

$$\int_0^{\infty} G(z')dz'=\frac{J_{dz}(0)}{2q}+\int_0^{\infty} U(n,p)dz' \ . \tag{24}$$

The left-hand side gives the electron-hole-pair generation per unit surface area resulting from the incident photon flux, while the two right-hand-side terms balance this against electron-hole-pair recombination at the surface and within the bulk, respectively. From this it is clear that the surface recombination $J_{dz}(0)/2q$ and the total bulk recombination must both have minimum values of 0 and maximum values equal to the total generation within the semiconductor, the left-hand side of Eq. (24). The total generation within the semiconductor can be readily evaluated from the absorptive properties of the semiconductor. When the semiconductor region occupies the positive z axis, $J_{dz}(0)/2q$ must be a non-negative quantity with a net flux of electrons and holes being absorbed at the surface but never created. Deeper into the semiconductor, $z>0$, $J_{dz}(z)/2q$ may take on either sign, since both carriers may flow toward the surface to recombine there, or flow further into the substrate as a result of diffusion. Several different cases are illustrated in Fig. 2.

The integrals of Eqs. (23a)–(23c) may be immediately evaluated if the spatial dependence of the potential and the bulk recombinative rate are known. The spatial dependence of the generation rate is presumed to be a known functional input. While these spatial profiles may not be well known for the general case, a large number of interesting surface problems involve the situation of surface depletion. Here, a well-defined layer adjacent to the surface becomes evacuated of electrons and holes in comparison to the carrier densities deeper into the substrate. When the carrier densities of this surface layer are small in comparison to the density of ionized impurities, the bulk space-charge density is approximately constant over the extent of the surface layer. This allows the electric field to be approximated as asymptotically linear with position z and the potential profile as asymptotically parabolic.

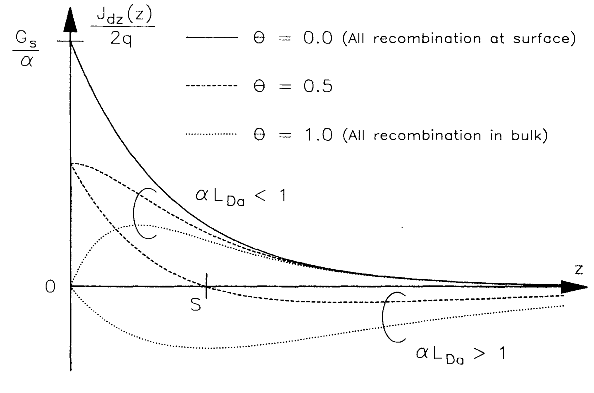

FIG. 2. Difference current density as a function of position into the semiconductor. Positive values correspond to a flow of both electrons and holes toward the surface; negative values correspond to a flow of both carriers further into the substrate. Only when the ambipolar diffusion length is greater than the absorption depth, $L_{Da}>1/\alpha$, do carriers diffuse further into the substrate. When the surface recombination is present along with $L_{Da}>1/\alpha$, a stagnation point S occurs as shown.

Following this line of thought, the surface layer is postulated to exist over a range of $[0,d]$, where d is termed the depletion depth. The potential profile is then approximated as

$$\psi(z)=\begin{cases} k(z-d)^2 & \text{for } 0\le z\le d \ , \\ 0 & \text{for } d\le z<\infty \ , \end{cases} \tag{25}$$

and the associated electric field profile is

$$E_z(z)=\begin{cases} -2k(z-d) & \text{for } 0\le z\le d \ , \\ 0 & \text{for } d\le z<\infty \ . \end{cases} \tag{26}$$

The surface values of the potential and field are

$$\psi_s=\psi(0)=kd^2 \ , \tag{27}$$

$$E_{zs}=E_z(0)=2kd \ , \tag{28}$$

and the value of k is set by the space-charge density of the surface layer,

$$k=\begin{cases} -\dfrac{qN_d}{2\epsilon} & \text{for the } n\text{-type case, } E_{zs}<0, \psi_s<0 \ , \\ +\dfrac{qN_a}{2\epsilon} & \text{for the } p\text{-type case, } E_{zs}>0, \psi_s>0 \ . \end{cases} \tag{29}$$

Thus, the surface electric field and surface potential can be represented by two parameters: d, which gives the extent of the surface depletion, and k, which gives the polarity and spatial rate of the band bending. It can be shown that the above approximations to the electrostatic potential are correct to within one kT/q when the impurity densities of the surface layer are constant with position.[4] This potential approximation is valid for the case of simple depletion and weak inversion of the surface.

Strong inversion or accumulation at the surface additionally introduce a significant free-carrier contribution to the space-charge density. Band bending at the surface makes this contribution exponentially largest at $z=0$; therefore, surface inversion or accumulation charges are usually described as a sheet charge density (cm^{-2}). Surface electron and hole sheet densities may be defined as

$$n_{ss}=\int_0^d n(z)dz \ , \tag{30a}$$

$$p_{ss}=\int_0^d p(z)dz \ , \tag{30b}$$

where the upper integration limit is chosen somewhat arbitrarily to be the depth of the depletion region in order to keep the integrals convergent for small to zero amounts of surface band bending. For the parabolic potential profile of Eq. (25), these sheet densities may be integrated approximately to yield

$$n_{ss}=n_s\frac{1-e^{-2\beta k d^2}}{2\beta k d} \ , \tag{31a}$$

which applies only for $k>0$, $E_{zs}>0$, $\psi_s>0$, and

$$p_{ss}=p_s\frac{1-e^{2\beta k d^2}}{-2\beta k d} \ , \tag{31b}$$

which applies only for $k<0$, $E_{zs}<0$, and $\psi_s<0$. Because these accumulation or inversion charge densities exist mainly at the surface itself, they cause little change in the parabolic potential profile assumed deeper into the band-bending region. Thus, they may be either considered as a direct contribution to the surface-charge density or as a correction term on the surface electric field.

The known characteristics of a uniform space-charge region may also be used to approximate the spatial dependence of the bulk recombinative rate. Sah, Noyce, and Shockley[19] have shown that the bulk recombinative rate is very nearly constant over the full extent of such a depletion region. Additionally, standard junction theory gives the recombinative rate as decaying exponentially within the quasineutral portion of the substrate, $z<d$, with a decay length equal to the ambipolar diffusion length of the carriers. In this context, the ambipolar diffusion length is defined via

$$\frac{1}{L_{Da}^2}=\frac{1}{D_n\tau_n}+\frac{1}{D_p\tau_p} \ . \tag{32}$$

From these features, the bulk recombinative rate may be approximated as of the form

$$U(z)=\begin{cases} U_1 & \text{for } 0\le z\le d \ , \\ U_1 e^{-(z-d)/L_{Da}} & \text{for } d\le z<\infty \ , \end{cases} \tag{33}$$

where U_1 is the unknown constant. The various approximated profiles are illustrated in Fig. 3.

The optical generation rate of electron-hole pairs is taken to be a simple exponential decay into the semiconductor, set by the absorption coefficient α,

$$G(z)=G_s e^{-\alpha z} \ . \tag{34}$$

The difference current density may now be found by simply integrating Eqs. (33) and (34) as per Eq. (23c). Evaluating this at the surface, which thereby yields the surface recombination rate, gives

$$\frac{J_{dz}(0)}{2q}=\frac{G_s}{\alpha}-(L_{Da}+d)U_1 \ . \tag{35}$$

Thus, the unknown constant of U_1 may be expressed in terms of G_s and $J_{dz}(0)/2q$. This is once more a simple statement that generated carriers must either recombine at the surface or in the bulk. It is convenient to define a parameter

$$\Theta=\frac{\dfrac{G_s}{\alpha}-\dfrac{J_{dz}(0)}{2q}}{\dfrac{G_s}{\alpha}} \ , \tag{36}$$

which gives the fraction of the total recombination that occurs within the bulk regions of the semiconductor. Clearly, $0\le\Theta\le 1$. The limiting case of zero bulk recombination corresponds to $\Theta=0$, and the case of zero surface recombination corresponds to $\Theta=1$. The difference current density may then be written as

$$\frac{J_{dz}(z)}{2q}=\left[\frac{G_s}{\alpha}\right][e^{-\alpha z}-\Theta F_1(z)] \ , \tag{37}$$

where

$$F_1(z)=\begin{cases} \dfrac{L_{Da}+d-z}{L_{Da}+d} & \text{for } 0\le z\le d \ , \\ \dfrac{L_{Da}}{L_{Da}+d}e^{-(z-d)/L_{Da}} & \text{for } d\le z\le\infty \ . \end{cases} \tag{38}$$

This difference current density may then be substituted into Eqs. (23a) and (23b) along with the potential profile of Eq. (25). The integrations are straightforward and lead to surface electron and hole densities of

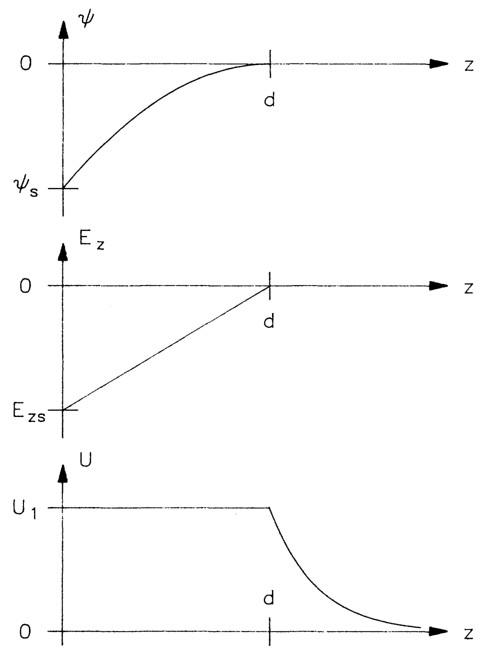

FIG. 3. Assumed profiles for the electrostatic potential, electric field, and bulk recombinative rate. The electrostatic potential is parabolic about its vertex at d when the space-charge density is uniform with position.

$$n_s = n(0) = e^{\beta k d^2}\left[n_{bo} - \frac{1}{D_n}\left[\frac{G_s}{\alpha}\right][F_2(-k,d) - \Theta F_3(-k,d)]\right] , \tag{39a}$$

$$p_s = p(0) = e^{-\beta k d^2}\left[p_{bo} - \frac{1}{D_p}\left[\frac{G_s}{\alpha}\right][F_2(k,d) - \Theta F_3(k,d)]\right] , \tag{39b}$$

where

$$F_2(k,d) = \int_0^d e^{-\alpha z'} e^{\beta k (z'-d)^2} dz' + \int_d^\infty e^{-\alpha z'} dz'$$

$$= \frac{1}{\alpha} e^{-\alpha d} + \begin{cases} \dfrac{e^{\beta k d^2}}{\alpha + 2\beta k d}(1 - e^{-\alpha d} e^{-2\beta k d^2}) & \text{for } k > 0 , \\ O\left[\dfrac{1}{2}\left[\dfrac{\pi}{-\beta k}\right]^{1/2} e^{-\alpha d}\right] & \text{for } k < 0 , \end{cases} \tag{40}$$

$$F_3(k,d) = \int_0^d F_1(z') e^{\beta k (z'-d)^2} dz' + \int_d^\infty F_1(z') dz'$$

$$= \left[\frac{L_{Da}}{L_{Da} + d}\right]\left[L_{Da} + \frac{e^{\beta k d^2 - 1}}{2\beta k L_{Da}} + \begin{cases} \dfrac{e^{\beta k d^2}}{2\beta k d}(1 - e^{-2\beta k d^2}) & \text{for } k > 0 \\ O\left[\dfrac{1}{2}\left[\dfrac{\pi}{-\beta k}\right]^{1/2}\right] & \text{for } k < 0 \end{cases}\right] . \tag{41}$$

Using the above approximations for the spatial dependence of the potential and the bulk recombinative rate allows the surface electron and hole densities to be expressed as explicit functions of the surface electric field and the surface recombination rate. These are the same two variables that are involved in the general surface boundary conditions. Thus, the illuminated free-surface problem can be reduced to two simultaneous transcendental equations in the unknowns of $\{E_z(0), J_{dz}(0)/2q\}$. But since the surface recombination rate is upper bounded by the total generation within the semiconductor, it is more convenient (and numerically stable) to consider the two unknown variables as $\{E_z(0), \Theta\}$:

$$E_z(0) = \frac{1}{\epsilon}\rho_s[n_s(E_z(0),\Theta), p_s(E_z(0),\Theta)] , \tag{42a}$$

$$\frac{G_s}{\alpha}(1-\Theta) = U_s[n_s(E_z(0),\Theta), p_s(E_z(0),\Theta)] . \tag{42b}$$

The functional dependence of ρ_s and U_s upon n_s and p_s is given by Eqs. (20) and (16), respectively, and n_s and p_s themselves are calculated according to Eqs. (39a) and (39b). With $E_z(0)$ and Θ determined from the above system of equations, all of the remaining variables may be found in turn as well as the spatial profiles for $\{n(z), p(z), J_{dz}(z)/2q\}$.

If the bulk defect capture cross sections for electrons and holes are significantly greater than those for the surface defects, then bulk recombination will dominate and produce the limiting case of $\Theta = 1$. The solution then reduces to merely solving Eq. (42a) alone with this choice of Θ. Similarly, any other particular choice of Θ simplifies the problem to a single transcendental equation for $E_z(0)$, which can again be efficiently found using a numerical bisection routine. Some care must be exercised when approaching the limiting case of dominant surface recombination where $\Theta \to 0$. If bulk recombination is specifically precluded, then $U(n,p) = 0$, which requires $J_{dz}(z)/2q$ to be strictly positive for all z, i.e., both carrier species must be moving toward the surface in order to recombine there and there can be no diffusion of the carriers deeper into the substrate than the point to which they were originally generated. Mathematically, the carrier bulk lifetime and diffusion length become infinite. From Eqs. (23a) and (23b) the surface electron and hole densities must then be reduced below their values associated with band bending alone in order to produce a sufficient gradient that will cause all of the generated carriers to eventually reach the surface. Choosing $\Theta = 0$ often results in the unphysical result of negative surface carrier densities, a consequence of the fact that the complete absence of bulk recombination is in itself a nonrealizable physical situation.

Because Θ may become very small in comparison to unity (but never zero) for certain physical values of bulk and surface capture cross sections, it is generally best to solve both Eqs. (42a) and (42b) simultaneously. Numerically, this can be accomplished by embedding a bisection routine for $E_z(0)$ inside another bisection routine for Θ. While this will not guarantee convergence to the root for an arbitrary mathematical problem, it has nevertheless been found to produce convergent and physically meaningful results for all cases that have been tried in the above surface problem.

III. PROPERTIES OF THE FREE SURFACE

A. Equilibrium

1. Charge balance

The balance of charge at the semiconductor surface is expressed by the solution to Eq. (12) under the conditions of equilibrium. The model of Brattain and Bardeen uses the behavior of surface states as donor and acceptor levels to allow the degree of ionization to vary as a function of surface band bending, which then provides the needed degree of freedom to account for pinning of the Fermi level.[3] In practice, a large number of surface defects may be present with different energy levels, degeneracies, and surface concentrations, and the solution to Eq. (12) becomes rather involved. Shockley originally introduced a graphical method to determine the Fermi-level position in an arbitrary semiconductor, which also involves the problem of partial ionization of many impurity levels.[20] This graphical method, sometimes referred to as a "Shockley plot," provides an easy means of visualizing the contributions of each charge source to the ultimate position of the Fermi level and is an alternative way of solving Eq. (4). The same idea used in the Shockley plot is extended here to visualize the contributions to the equilibrium band bending of a free surface.

Because the ionization of surface donor or acceptor levels follows simple Fermi-Dirac occupation factors, they are most easily plotted on a semilogarithmic scale versus the position of the Fermi energy. The value of the Fermi energy at the surface is set by the position of the Fermi energy in the bulk substrate and the amount of band bending, ψ_s. Hence, the solution of Eq. (12) involves finding the value of the surface potential ψ_s, which balances the surface charge against the field caused by the bulk impurities within the depletion region. The original idea of the Shockley plot was to sum all of the positive contributions to the charge density and equate them to the sum of the negative contributions, with each sum taking place on a logarithmic scale. The crossing of the two lines as a function of Fermi energy then determines the charge neutrality point. The same method may be applied to the case of Eq. (12) except that the contribution of the electric field may be either positive or negative, depending upon the direction of the band bending.

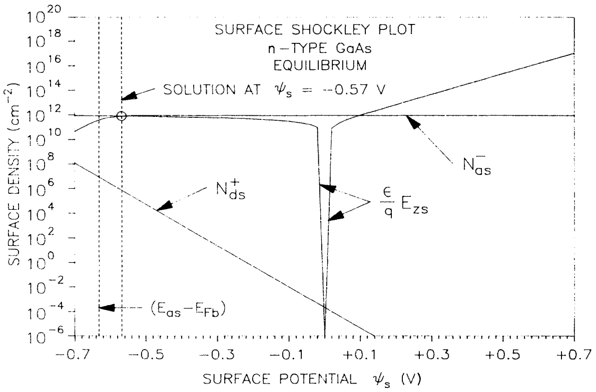

FIG. 4. Surface Shockley plot for *n*-type GaAs at equilibrium. Fermi level and surface potential are set by surface acceptor-field pinning. This pinning is relatively weak because the N_{as}^- and $\epsilon E_{zs}/q$ curves cross at a very shallow angle, making the solution point very susceptible to changes in the bulk space-charge density.

To account for this, Eq. (12) may be rewritten as

$$N_{ds}^+ - \frac{\epsilon}{q} E_{zs} \;\; (\text{if } E_{zs} < 0) = N_{as}^- + \frac{\epsilon}{q} \; E_{zs} \;\; (\text{if } E_{zs} > 0) \; , \tag{43}$$

where both sides contain the sum of strictly positive quantities and an implicit sum over a possible multiplicity of surface donors and acceptors is assumed. For the case of equilibrium, the surface electric field E_{zs} may be obtained exactly from Eq. (11), whose first two terms include the possible contributions from surface inversion or accumulation charges. A representative graphical solution of Eq. (43) is shown in Fig. 4 for the case of an *n*-type GaAs substrate whose surface is Fermi level pinned by a dominant midgap surface acceptor level. The surface states are taken to be the two proposed by Spicer, *et al.*:[11,12] a dominant acceptor state located at $E_{as} = 0.75$ eV above the valance-band edge with an assumed density of $N_{as} = 10^{12}$ cm^{-2} and a secondary donor state located at $E_{ds} = 0.50$ eV above the valence-band edge with an assumed density of $N_{ds} = 10^{11}$ cm^{-2}.

2. *Fermi-level pinning*

The previous graphical solution provides an improved means for visualizing Fermi-level pinning on free semiconductor surfaces. It must be clearly pointed out that this precludes any electrostatic effects introduced by the presence of a metallization in contact with the surface, and is therefore not applicable to the case of Fermi-level pinning in metal-semiconductor junctions.

First it should be noted that the knee of the ionized surface donor curve, i.e., the point of 50% ionization, is centered over the potential value of $\psi_s = E_{ds} - E_{Fb}$, and the knee of the ionized surface acceptor curve is centered over the potential value of $\psi_s = E_{as} - E_{Fb}$. The intersection of the positive and negative contributions to the surface Shockley plot will generally occur in between the knees of the ionized surface donor and acceptor curves when $E_{as} > E_{ds}$. When the surface impurity levels are additionally near to midgap, an *n*-type substrate will produce a large value of E_{Fb} that will in turn force the solution toward a negative value of ψ_s and deplete the surface of free electrons. Similarly, a *p*-type substrate will produce a comparatively small value of E_{Fb} that will force the solution toward a positive value of ψ_s and deplete the surface of free holes.

From this one important feature of Fermi-level pinning can be observed. If the surface donor energy is smaller than that of the surface acceptor, then the Fermi energy will be pinned in between the two levels, regardless of the relative densities of the two surface impurity levels. If the surface donor energy is larger than that of the surface acceptor, then the Fermi energy will tend to be pinned to the energy level of the impurity with the higher surface density. Since the surface impurities tend to be partially ionized, the position of the pinned Fermi level will be slightly above a dominant surface donor energy, or slightly below a dominant surface acceptor energy. Regardless of the energy ordering of the surface impurity levels, the position of the pinned Fermi energy is set by overall charge balance; not merely alignment to the impurity level with the greatest surface density.

For *n*- and *p*-type substrates where the bulk Fermi energy is significantly different from the energy of the surface impurity levels, the solution point on the surface Shockley plot involves either an intersection of the ionized surface donor curve with the negative surface electric field curve or an intersection of the ionized surface acceptor curve with the positive surface electric-field curve. These two cases can be termed "surface donor-field pinning" for the *p*-type case and "surface acceptor-field pinning" for the *n*-type case. Semi-insulating, compensated, or intrinsic substrates offer a much different situation, since the position of the bulk Fermi level corresponds closely to the position of midgap surface impurity levels. The surface Shockley plot for semi-insulating GaAs with the same set of surface impurity parameters is shown in Fig. 5. Since the bulk Fermi-level position lies close to the level of midgap surface impurity states, very little equilibrium band bending is exhibited. In the case of Fig. 5, the position of the Fermi level may again be classified as surface acceptor-field pinning. However, the much closer proximity of the ionized surface donor curve allows the possibility of "surface donor-acceptor pinning" if the surface densities of the two impurities both exceed the corresponding value of the surface electric-field curve. Surface donor-acceptor pinning represents a much stronger binding of the Fermi level to a particular surface energy, because the equilibrium position is set essentially by charge neutrality at the surface, with virtually no contributing influence from the surface electric field of the bulk. Surface donor and acceptor densities need to only exceed the level of $\epsilon E_{zs}/q$, which is on the order of 10^{12} cm^{-2}, and which corresponds to an atomic surface coverage of only 0.0007, to achieve surface donor-acceptor pinning.

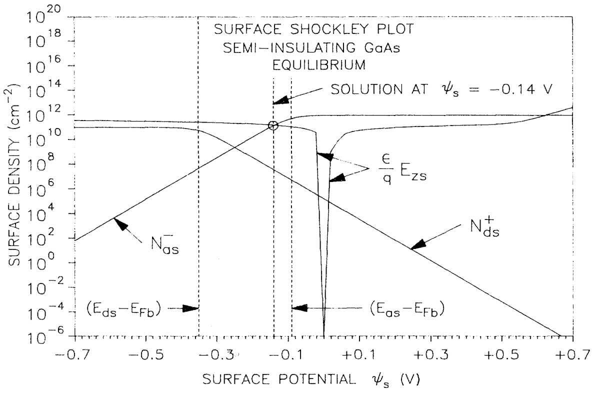

FIG. 5. Surface Shockley plot for semi-insulating GaAs at equilibrium. Surface potential is set by surface acceptor-field pinning and is again relatively weak. Reduced ionized impurity densities in semi-insulating GaAs produce smaller $\epsilon E_{zs}/q$ values and the intersection of the N_{ds}^+ and N_{as}^- curves can easily exceed this, producing surface donor-acceptor pinning, which is then independent of the bulk space charge and classified as strong pinning.

Because any modulation of the depletion region charge will alter the position of surface-donor-field-pinned or surface-acceptor-field-pinned Fermi levels, these may be referred to as "weak" Fermi-level pinning. The only mechanism whereby surface-donor-acceptor-pinned Fermi levels may be altered is through a change in the ionization fractions, which entails a severe change in the surface-state occupation factors. Thus, surface donor-acceptor pinning may be regarded as "strong" Fermi-level pinning. Large changes in the electron and hole surface densities may be obtained by illumination of the free surface, and sufficiently strong illumination may cause repositioning of the Fermi level in both weak and strongly pinned systems, as will be discussed next.

B. Illumination

Under conditions of illumination, the steady-state alignment of the free surface is obtained as a solution to Eqs. (42a) and (42b). For a prescribed fraction of the generated carriers recombining within the bulk region of the semiconductor Θ, the solution is obtained from Eq. (42a) alone, which manifests the same general features as the equilibrium case of Eq. (12) when a multiplicity of defect levels is involved. A graphical solution can again be obtained, but the critical occupation factors for the surface states must be obtained from the surface carrier densities, rather than from an equilibrium Fermi level.

1. Nonequilibrium surface-state occupation

In the general nonequilibrium case, Eqs. (39a) and (39b) can be compacted into the form of

$$n_s = e^{\beta\psi_s}(n_{bo} + n_i \gamma_n G_s) , \tag{44a}$$

$$p_s = e^{-\beta\psi_s}(p_{bo} + n_i \gamma_p G_s) , \tag{44b}$$

where

$$\gamma_n = \frac{1}{\alpha n_i D_n}[\Theta F_3(-k,d) - F_2(-k,d)] , \tag{45a}$$

$$\gamma_p = \frac{1}{\alpha n_i D_p}[\Theta F_3(k,d) - F_2(k,d)] . \tag{45b}$$

The Fermi potential in the equilibrium substrate is given by $\phi_b = (E_{Fb} - E_i)/q$ such that

$$n_{bo} = n_i e^{\beta\phi_b} , \tag{46a}$$

$$p_{bo} = n_i e^{-\beta\phi_b} . \tag{46b}$$

Substituting these expressions for the surface electron and hole densities into Eq. (17) yields a general electron occupation factor of

$$f = \frac{1}{1 + \frac{g_f}{g_e} M e^{(E_t - E_i)/kT} e^{-\beta(\psi_s + \phi_b)}} , \tag{47}$$

where

$$M=\frac{\sigma_p e^{-\beta(\psi_s+\phi_b)}e^{(E_i-E_t)/kT}+\sigma_p\gamma_p G_s e^{(E_i-E_t)/kT}e^{-\beta\psi_s}+\sigma_n}{\sigma_p e^{-\beta(\psi_s+\phi_b)}e^{(E_i-E_t)/kT}+\sigma_n\gamma_n G_s e^{-\beta\phi_b}+\sigma_n}\;. \tag{48}$$

When there is no illumination present, $G_s=0$, the factor M becomes unity and the occupation factor f reduces to the usual Fermi-Dirac function.

It is important to note that the factor M saturates with increasing illumination intensity,

$$\lim_{G_s\to\infty} M=\frac{\sigma_p\gamma_p}{\sigma_n\gamma_n}e^{(E_i-E_t)/kT}e^{\beta(\phi_b-\psi_s)}\;, \tag{49}$$

which is then independent of G_s. In the limit of strong illumination, the occupation factor thus becomes independent of the energy of the defect state relative to the band gap

$$f=\frac{1}{1+\dfrac{\sigma_p g_f D_n[\Theta F_3(k,d)-F_2(k,d)]e^{-\beta\psi_s}}{\sigma_n g_e D_p[\Theta F_3(-k,d)-F_2(-k,d)]e^{\beta\psi_s}}}\;. \tag{50}$$

Furthermore, $F_2(k,d)\approx F_2(-k,d)$ when $d\ll 1/\alpha$ and $F_3(k,d)\approx F_3(-k,d)$ when $d\ll L_{Da}$, both of which are made more accurate when the depletion region contracts as a result of illumination. An important result is that the occupation factor under conditions of saturating illumination depends only upon the surface potential and the ratio of the capture cross sections, the degeneracy ratio, and the ratio of the electron and hole diffusivities,

$$f=\frac{1}{1+\dfrac{\sigma_p g_f D_n}{\sigma_n g_e D_p}e^{-2\beta\psi_s}}\;. \tag{51}$$

This is equivalent to the knee of the ionized surface donor and acceptor curves being shifted by an amount $\ln(M)/\beta$ in potential to a point at

$$\psi_s=\frac{1}{2\beta}\ln\left[\frac{\sigma_p g_f D_n}{\sigma_n g_e D_p}\right]\;. \tag{52}$$

2. *Depletion-region contraction and photovoltage*

In terms of application towards semiconductor device modeling, the most influential features of an illuminated free surface are the photoinduced changes in surface potential and surface depletion depth. These can be calculated directly by the procedures outlined in Sec. II, but it is again useful to examine the behavior by means of a Shockley plot. This can be constructed by modifying the equilibrium case. The first change is to replace the equilibrium occupation factors by the nonequilibrium result of Eq. (47). This effectively shifts the ionized donor and acceptor curves horizontally by a potential of $\ln(M)/\beta$, saturating with increasing illumination at the point given by Eq. (52).

For the nonequilibrium case, Eq. (11) can no longer be used to evaluate the surface electric field, and the approximate analytical solution must be used instead. Without corrections for possible surface inversion or accumulation charges, the surface electric field is simply

$$E_{zs}=2\sqrt{k\psi_s}\;. \tag{53}$$

The surface inversion or accumulation charges may instead be added directly to the surface-charge density via

$$N_{ds}^{+}+p_{ss}-\frac{\epsilon}{q}E_{zs}\ (\text{if } E_{zs}<0)=N_{as}^{-}+n_{ss}+\frac{\epsilon}{q}E_{zs}\ (\text{if } E_{zs}>0)\;, \tag{54}$$

where

$$n_{ss}=n_{bo}\frac{\sinh(\beta\psi_s)}{\beta\sqrt{k\psi_s}} \tag{55a}$$

only for $k>0$, and

$$p_{ss}=p_{bo}\frac{\sinh(-\beta\psi_s)}{\beta\sqrt{k\psi_s}} \tag{55b}$$

only for $k<0$. The resulting Shockley plots are shown in Figs. 6 and 7 for n-type and semi-insulating GaAs, which are exposed to a saturating level of illumination. As is commonly noted in experiments, the surface potential and the surface depletion depth are both observed to decrease in magnitude with the application of illumination. The shift in the surface potential from the dark to the illuminated case is the developed photovoltage of the free surface.

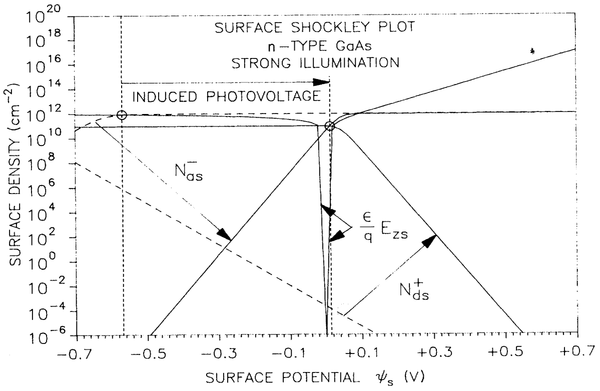

FIG. 6. Surface Shockley plot for n-type GaAs under strong illumination. The shift in the $\epsilon E_{zs}/q$ curve from equilibrium is negligible, but the ionized surface donors and acceptors are significantly altered, saturating at the locations shown. The shift in the solution point changes the surface potential by the indicated amount of induced photovoltage.

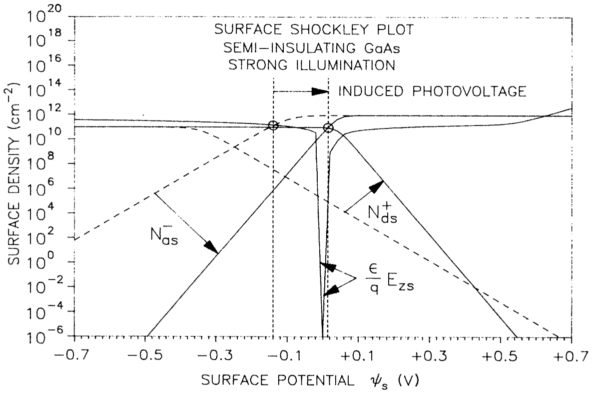

FIG. 7. Surface Shockley plot for semi-insulating GaAs under strong illumination. The induced photovoltage is less than in the *n*-type case, but the final solution point is nearly identical, since in both cases the surface potential under strong illumination is set by surface donor-accepting pinning, which is independent of the bulk charge density.

3. Band flattening origins

Gärtner[21] was the first to include the effects of optical generation within the depletion region of a metal-semiconductor junction. In this analysis, the developed photovoltage is derived as the amount of forward bias needed to produce a dark forward current that just cancels the photocurrent under open-circuit conditions. Here, the metal contact establishes the necessary surface charge on the semiconductor to support (or reduce) the band bending. In fact it is the presence of the metal contact that produces the exponential current-voltage characteristics that ultimately yield the logarithmic dependence of the photovoltage upon the incident light flux. The essential point here is that it is completely invalid to model the free semiconductor surface as an open-circuited Schottky barrier because of the lack of the metal contact, without which the barrier does not exist. To impose a solution based upon the boundary conditions of a metal-semiconductor junction fully ignores the actual (and different) boundary conditions of the free surface. It is similarly improper to consider the band flattening and induced photovoltage as arising from photodiffusion or Dember[22–24] effects, as these are also based upon the boundary conditions of a metal-semiconductor junction.

Under sufficiently intense illumination, the photogenerated electron and hole density will become large enough to screen the space charge of the ionized bulk impurities. This means is often cited as the cause of depletion-edge contraction and induced photovoltage on free surfaces. However, the radiation intensity required to achieve this quenching of the depletion region field is between 2 and 3 decades higher than the radiation levels which substantially reduce the depth of the depletion region, as determined by rf surface conductivity measurements.[25] The present derivation and model easily show that the cause of band flattening, depletion-edge contraction, and induced photovoltage all lie with a change in the occupation factors of the surface states. In nonequilibrium, the occupation factors are a function of the surface electron and hole concentrations, which are very susceptible to alteration by incident light. In fact, the change in the surface carrier concentrations is the greatest for any point within the domain of the semiconductor, and this is also where the surface states themselves are located. The level of incident illumination required to alter the surface electron and hole densities and, subsequently, the occupation factors of the surface states is 2–4 decades smaller than the level required to quench the depletion region field via bulk photogenerated carriers. Hence, the origin of these band flattening effects should be reinterpreted as arising from changes in the nonequilibrium occupancy of the surface states rather than from merely a high density of photogenerated carriers in general.

IV. CONCLUSIONS

The physical alignment of the Fermi energy in equilibrium free semiconductor surfaces can be easily obtained through numerical bisection routines, which admit the full range of temperature, semiconductor band gap, Fermi-Dirac statistics, impurity location and density, and surface-state density, energy, and degeneracy. These procedures are particularly well suited to when a multiplicity of defect states is involved, either in the bulk or at the surface. The consideration of partial ionization of the defect states is essential to allow for charge balance at the surface, as well as to accurately represent the case of wide-band-gap materials that are dominated by deep-level bulk states.

Free semiconductor surfaces cannot be represented by boundary conditions that mimic metal-semiconductor junctions or other open-circuit device conditions. Instead, the two key boundary conditions of charge balance and surface recombination must replace the more usual specification of carrier densities or terminal currents. While these boundary conditions are difficult to treat both numerically and analytically, their use is required if the full physics of the free surface is to be represented.

A simple approximate analytical model has been presented that allows the equilibrium solution techniques to be extended to the case of incident illumination on the free surface. This model reduces identically to the equilibrium case when illumination is absent, and allows the photoinduced changes in the surface potential, surface field, carrier densities, current densities, and surface-state occupation factors to all be readily obtained from a two-variable bisection routine. The influence of surface versus bulk recombination is also developed within this model, and even though the case of zero bulk recomination becomes unphysical, the actual fraction of recombination occurring within the bulk becomes only a second-order effect in the overall model.

A graphical approach has been developed to show the relative contributions to the charge balance of the surface. Fermi-level pinning can be classified according to the origin of the dominant contributions. Weak Fermi-level pinning occurs when the electric field of the depletion region balances against a single dominant surface donor or surface acceptor. The pinning of the surface is weak in this case because the solution point is very susceptible to small changes in the surface electric field, which could be induced by any small change in the depletion-region space charge. Alternatively, strong pinning of the Fermi level occurs when the solution point is set by the balance between a surface donor and a surface acceptor. The pinning in this case is strong because the solution is set essentially by neutrality of the surface and is independent of the surface electric field produced by the ionized bulk impurity states. This case of strong Fermi-level pinning can easily occur when the surface donor and acceptor densities exceed levels of around 10^{12} cm^{-2}, which represents a relatively small atomic surface coverage.

Finally, an important result is that the occupation factors of the surface states are strongly influenced by incident illumination, and the curves of occupation factor versus surface potential shift and saturate with increasing light intensity. Under strong illumination, the occupation factors become dependent only upon the surface potential and the ratios of carrier capture cross sections, degeneracies, and carrier diffusivities. The shift in the occupation factors is such as to force the surface potential more toward zero, i.e., to flatten the bands, as the illumination level is increased. A consequence of this is that the flattening of the bands on free semiconductor surfaces should be reinterpreted in terms of changes in the occupancy of the surface states, rather than from an overall increase in electron and hole densities, which would quench the depletion-region electric field. The required change in the surface-state occupancy occurs several decades lower in illumination intensity than the field-quenching mechanism.

ACKNOWLEDGMENTS

The author would like to thank Dr. K. J. Kuhn, Dr. J. E. Samaras, and Dr. H. J. Youn for many engaging discussions. This work was supported by the National Science Foundation under Contract No. MIP-8822121, the Washington Technology Center under Contract No. 09-1091, and Boeing Aerospace and Electronics under Contract No. 65-6293.

---

[1] W. Shockley, Phys. Rev. **56**, 317 (1939).

[2] I. Tamm, Phys. Z. Sowjetunion **1**, 733 (1932).

[3] W. H. Brattain and J. Bardeen, Bell. Syst. Technol. J. **32**, 1 (1953).

[4] C. G. B. Garrett and W. H. Brattain, Phys. Rev. **99**, 376 (1955).

[5] J. Bardeen, Phys. Rev. **71**, 717 (1947).

[6] W. E. Meyerhof, Phys. Rev. **71**, 727 (1947).

[7] A. M. Cowley and S. M. Sze, J. Appl. Phys. **36**, 3212 (1965).

[8] H. Statz, G. A. deMars, L. Davis, Jr., and A. Adams, Jr., Phys. Rev. **101**, 1272 (1956).

[9] F. G. Allen and G. W. Gobeli, Phys. Rev. **127**, 150 (1962).

[10] J. E. Thomas, Jr. and R. H. Rediker, Phys. Rev. **101**, 984 (1956).

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

[12] W. E. Spicer, Z. Lilienthal-Weber, E. R. Weber, N. Newman, T. Kendelewicz, R. Cao, C. McCants, P. Mahowald, K. Miyano, and I. Lindau, J. Vac. Sci. Technol. B **6**, 1245 (1988).

[13] W. van Roosbroeck, J. Appl. Phys. **26**, 380 (1955).

[14] R. H. Kingston and S. F. Neustadter, J. Appl. Phys. **26**, 718 (1955).

[15] T. K. Gaylord and J. N. Linxwiler, Jr., Am. J. Phys. **44**, 353 (1976).

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Pascal: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1989).

[17] W. Shockley and W. T. Read, Jr., Phys. Rev. **87**, 835 (1952).

[18] R. N. Hall, Phys. Rev. **87**, 387 (1952).

[19] C. T. Sah, R. N. Noyce, and W. Shockley, Proc. IRE **45**, 1228 (1957).

[20] W. Shockley, *Electrons and Holes in Semiconductors* (Van Nostrand, Princeton, 1950).

[21] W. W. Gärtner, Phys. Rev. **116**, 84 (1959).

[22] H. Dember, Phys. Z. **32**, 554 (1931).

[23] H. Dember, Phys. Z. **32**, 856 (1931).

[24] H. Dember, Phys. Z. **33**, 207 (1932).

[25] E. Yablonovitch, B. J. Skromme, R. Bhat, J. P. Harbisor J. T. Gmitter, Appl. Phys. Lett. **54**, 555 (1989).