DIODES

The word *diode* originates from the Greek word Διοδος meaning passage or way through. In electronics terminology, in fact, diode refers to a two-terminal device that allows current to flow in one direction while it blocks the flow of current in the opposite direction. Such devices usually employ semiconductor junctions or metal-semiconductor junctions. There are also diodes made of vacuum tubes or metal—purely ionic crystal contacts. This article deals with semiconductor p–n junction diodes because they are the most widely used in practice due to their versatility, reproducibility, stability, and compatibility with integrated circuit technology. Additionally, an insight into the operation of the p–n junction diode is the basis for understanding the device physics of other semiconductor devices, the majority of which use the p–n junction as the building block. The semiconductor of our choice will be silicon because almost all diodes, discrete or integrated, are made of this element. Extensions to other semiconductors will be made to generalize theoretical results or to set limits to the validity of certain equations.

Figure 1(a) shows the electrical symbol of a diode. The arrow-type symbol indicates the conduction direction. For a diode to conduct an appreciable electric current, the voltage on the left side of the symbol must be a little higher than the voltage on the right side. If this polarity is reversed, the current drops to negligible values even for a large bias. The two previous polarity modes are known as forward and reverse bias, respectively. In Fig. 1(b) the very basic material structure of a diode is shown. The starting material is a high-purity silicon crystal, the properties of which are properly modi-

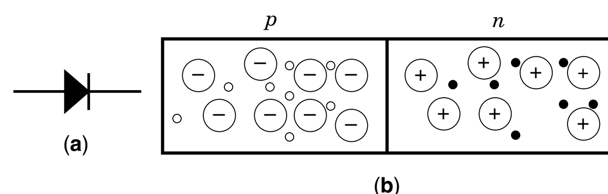

Figure 1. Electrical symbol of a diode (a) and illustration of a semiconductor p–n junction (b). In (b) the large circles with the minus and the plus signs are the acceptor and the donor ions, whereas the small circles are the holes (empty) and the electrons (dark).

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*. Copyright © 1999 John Wiley & Sons, Inc.

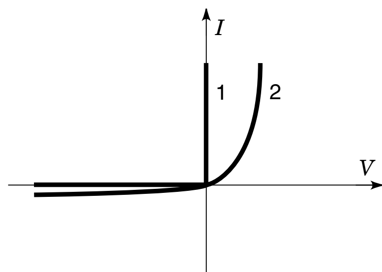

Figure 2. The current-voltage characteristics of an ideal diode, curve 1, and a realistic one, curve 2.

fied by selectively introducing dopants (elements) from either the third or the fifth column of the periodic table. The third-column elements, when introduced into the silicon lattice, behave as acceptors: they trap electrons from the valence band, thereby creating positively charged holes in the valence band and negatively charged immobile acceptor ions. The acceptor-doped part of the diode is called the *p*-side. On the other hand, fifth-column elements behave as donors: they give up their fifth electron, creating a population of conduction band electrons and positively charged immobile donor ions. The donor-doped part of the device is the *n*-side. The introduction of acceptor and donor dopants into silicon creates the two polarity sides of the diode, as shown in Fig. 1(a). Schematically speaking, when applying a forward bias, the higher voltage on the *p*-side makes the electron and hole gases move into each other. Thus, an electric current is created through electron-hole pair recombination. On the contrary, a lower voltage on the *p*-side moves the charge carriers away, thus preventing recombination and eliminating the current. In terms of dopants, the previous account of how the *p–n* diode is formed also holds for germanium diodes, which also is a fourth-column elemental semiconductor. For compound semiconductors (e.g., GaAs, InP, CdTe), the chemical origin of donor and acceptor dopants is more complex in relation to the semiconductor elements themselves.

In Fig. 2 the current-voltage (*I–V*) characteristic of an ideal diode as well as of a realistic one is shown. The ideal diode would behave as a perfect switch when forward biased: unlimited current flows without any voltage drop across the device. The same ideal diode would allow no current in the reverse direction, no matter what the magnitude of the reverse bias is. Now, a realistic semiconductor diode would exhibit a resistance to current flow in the forward direction, whereas in reverse bias a small current would always be present due to leakage mechanisms. The disagreement between the ideal and the actual electrical behavior is not restricted only to the static *I–V* characteristics shown in Fig. 2. It extends to the transient response obtained when applying a time-dependent terminal excitation. The response of a realistic diode cannot follow at exactly the same speed as the terminal excitation of an ideal diode would. When designing a diode to be used as a switching device, care is taken to bring the device electrical characteristics as close to the ideal ones as possible. This is done by choosing both the geometrical features and the fabrication process steps in a way to suppress the parasitic components of the diode. As a result of the semiconductor electronic band structure as well as technological constraints, material limitations impose certain basic restrictions on the device performance and create the subsequent deviation from the ideal performance. In the following sections, these restrictions will be investigated, and the deviation from the ideal performance will be analyzed in terms of the basic device physics, material constants, and geometry considerations.

Before considering the device physics of the diode, we will briefly discuss the steps in the basic fabrication process employed when making a silicon diode. These steps determine its basic geometrical and technological characteristics, which in turn determine the device electrical behavior. Today, almost all silicon diodes are made through the standard planar process of the silicon integrated circuit technology. A silicon wafer is first oxidized at temperatures in the vicinity of 1000°C. Such oxidation creates a silicon dioxide (SiO_2) cover layer with a thickness on the order of a micron. This layer is used as a mask for the subsequent technological steps. The SiO_2 film is then patterned by lithographic techniques and through etching, which allows windows of exposed silicon to be opened. Then, either by diffusion or by ion implantation, dopants are introduced into the exposed areas. The dopants are of a type opposite to the one already existing in the original wafer. In this way, *p–n* junctions are created in the exposed areas. In the rest of the wafer, the SiO_2 layer stops the ions and prevents diffusion into the silicon bulk. On the back surface, another diffusion or implantation of the same dopants as in the bulk is usually applied for reasons that will become clear in the next sections. At the end, metal contacts are evaporated on the front and the back. Lithography, again, on the front side defines the contacts of the individual diodes. The metal contacts are required for the diodes to interact with the external world in terms of terminal excitation (voltage or current) and terminal response (current or voltage, respectively). Similar methods are used for germanium diodes, whereas the compound semiconductor devices are usually made by epitaxial growth on proper substrates and by in situ doping.

FUNDAMENTALS OF *P–N* JUNCTIONS

The basic *p–n* junction device physics was proposed by Shockley (1). He derived the current-voltage characteristics, considering the electron and hole current continuity equations and the relationship between the carrier quasi-Fermi levels and the externally applied potentials. Here, we rederive the general current-voltage relation of a *p–n* junction based on Shockley's classic work (1) and its later extension (2).

Basic Equations and Assumptions

To formulate the electron and hole transport in a semiconductor device mathematically, we can always start by expressing the carrier densities and currents in terms of the carrier quasi-Fermi potentials under uniform temperature conditions:

$$J_n = -e\mu_n n \nabla F_n \tag{1a}$$

$$J_n = e\mu_n n E + eD_n \nabla n \tag{1b}$$

$$n = n_i \exp\left(\frac{-eF_n - E_i}{kT}\right) \tag{1c}$$

$$J_p = -e\mu_p p \nabla F_p \tag{2a}$$

$$J_p = e\mu_p p E - eD_p \nabla p \tag{2b}$$

$$p = n_i \exp\left[\frac{-(-eF_p) + E_i}{kT}\right] \tag{2c}$$

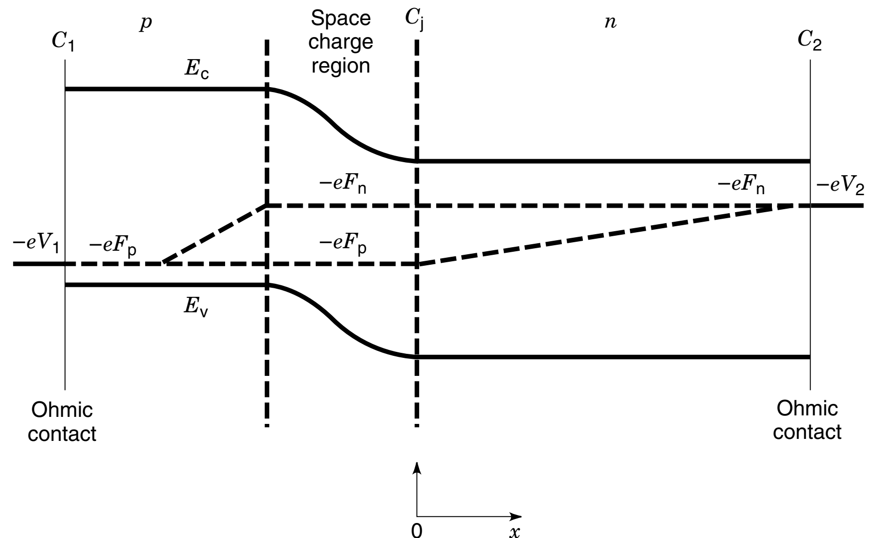

Figure 3. The band diagram of a forward-biased *p–n* junction. Boundaries C_1 and C_2 are the ohmic contacts, whereas C_j is the base-injecting boundary.

In Eqs. (1a) and (2a), J_n and J_p are the electric current densities of electrons and holes, respectively. Equations (1b) and (2b) express the currents in terms of drift and diffusion, where μ_n, D_n, n and μ_p, D_p, p are the mobilities, diffusivities, and volume densities of electrons and holes, respectively. Finally, F_n and F_p are the electron and hole quasi-Fermi potentials, E_i is the intrinsic energy level, and E the electric field density. Figure 3 shows the energy band diagram of a *p–n* junction under forward bias and illustrates the space dependence of the quasi-Fermi potentials, of the bottom of the conduction band E_c, and of the top of the valence band E_v. Equations (1c) and (2c) hold provided that the differences $E_c - (eF_n)$ and $(eF_p) - E_v$ are positive and at least several times the thermal energy kT. Equations (1a) and (2a) are borrowed from thermodynamics and hold provided that the bias is such that perturbations from equilibrium are small. Small, here, implies that the energy distribution of electrons and holes in the conduction and the valence band, respectively, continue (within a good approximation) to follow the Boltzmann statistics. Additionally, we assume that the mean free paths of the carriers are negligible compared to the physical dimensions of the device. Finally, Eqs. (1) and (2) hold provided quantum mechanical tunneling of carriers across potential barriers is not important. Such a constraint is relaxed in the last section of nonconventional transport diodes.

The second set of equations to be considered is the electron and hole continuity equations:

$$\frac{\partial n}{\partial t} = \frac{1}{e}\nabla \cdot J_n - U(n,\ p) + G \tag{3}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{e}\nabla \cdot J_p - U(n,\ p) + G \tag{4}$$

where U is the electron-hole net recombination rate either by band-to-band transitions or through traps, whereas G is the band-to-band generation rate resulting from ionizing radiation or impact ionization processes. For the sake of simplicity, we assume that U and G are the same for both carriers. Equations (3) and (4) are more general than the previous ones because they do the "bookkeeping" by equating the increase in the rate of carrier density to minus the carrier losses resulting from carrier out-fluxing ($\nabla J_p/e$ and $-\nabla J_n/e$) and recombination.

Next is the Poisson equation, which relates the electric field to the charge density caused by both mobile and immobile charges:

$$\nabla \cdot E = \frac{e}{\epsilon}[-n + p + N_{\mathrm{D}} - N_{\mathrm{A}}] \tag{5}$$

where ϵ is the semiconductor dielectric constant and N_{D} and N_{A} are the donor and acceptor densities, respectively. The charge density resulting from donors and acceptors is not carrier density dependent, unless the temperature drops to the cryogenic region.

The final equation to be considered is the one that equates the electric field to the gradient of the electrostatic potential:

$$E = \frac{1}{e}\nabla E_{\mathrm{c}} = \frac{1}{e}\nabla E_{\mathrm{v}} = \frac{1}{e}\nabla E_{\mathrm{i}} \tag{6}$$

Equation (6) implies that the electrostatic potential is determined by conduction and the valence band edges because the carriers there have only potential energy. The last equation assumes that the separation in the energy scale of the three levels (E_c, E_v, and E_i) is space independent. So, in Eq. (6), as well as in Eqs. (1b, c) and (2b, c), we neglect band distortion resulting from heavy doping or other effects (e.g., mechanical stain). This is discussed in a later section on heavy doping effects.

Boundary Conditions. Equations (1)–(6) form a system of six relations with six unknown variables: F_n, F_p, E_i, J_n, J_p, and E. They apply, within the range of their validity, to any semiconductor device. In this sense, any semiconductor device understanding, design, operation, and performance is based on this set of six equations. What distinguishes a device of a particular kind is its boundary conditions, as well as the doping and trap density and type.

Ohmic Contacts. For a diode, a two-terminal device, the boundary conditions necessarily include two ohmic contacts that will supply the charge to be transported through the de-

vice. The voltage across and the current through the two ohmic contacts are, interchangeably, the excitation or the response of the device. The ohmic contacts are realized by depositing metals (e.g., Ti or Al) on heavily doped regions of the semiconductor. An ideal ohmic contact should, by definition, establish thermodynamic equilibrium between the metal and the semiconductor at all the contact points. In analytical terms, this is expressed by equating the carrier quasi-Fermi potentials to the metal Fermi potential:

$$F_n(C_1) = F_p(C_1) = V_1 \tag{7a}$$

$$F_n(C_2) = F_p(C_2) = V_2 \tag{7b}$$

where V_1 and V_2 are the voltages of the metal contacts ($V_1 - V_2$ is the terminal voltage), whereas C_1 and C_2 are the contact areas of the first and the second ohmic contact, respectively. The pinning of the Fermi potentials at the externally applied voltages is illustrated in Fig. 3. If $V_1 = V_2$ and G in Eqs. (3) and (4) were zero, the device would be in equilibrium. Then the solution of the previous system of six equations would be zero currents and equal and flat Fermi-potentials throughout the diode. When an external voltage is applied, the splitting of the Fermi-potential values between the two ohmic contacts drives the device out of equilibrium. Such a boundary value split enforces a separation of the electron and hole quasi-Fermi potentials through the device, as shown in Fig. 3. The separation of the two potentials implies that the nonequilibrium conditions mainly refer to the interaction between electrons and holes. At any point, excluding ohmic contacts, electrons are out of equilibrium with respect to holes because the relaxation time of interband transitions (recombination and generation mechanisms) required to bring them into equilibrium are too slow (milliseconds or microseconds for germanium and silicon and nanoseconds for most compound semiconductors). On the contrary, the intraband transitions resulting from scattering have short relaxation times (picoseconds) so that electrons or holes are nearly in equilibrium within their band. This is required for the carrier Fermi-potentials to have a meaning, as mentioned in the discussion following Eqs. (1) and (2).

Semiconductor-Insulator Interfaces. The surface that bounds the device includes, in addition to the ohmic contact, the semiconductor-vacuum or semiconductor-insulator interface. The exposed semiconductor surface is usually covered by an insulating film (SiO_2 in silicon) to reduce recombination. If we assume that there is no injection in the insulator, then at the interface the boundary conditions for Eq. (3) and (4) are

$$\frac{\partial n_s}{\partial t} = -\frac{1}{e} J_n^n - U_s \tag{8}$$

$$\frac{\partial p_s}{\partial t} = \frac{1}{e} J_p^n - U_s \tag{9}$$

where the subscript s refers to surface densities and the superscript n refers to the normal component looking into the insulator. The boundary conditions for Eq. (5) are dictated by the lows of electrostatics. The discontinuity of the normal component of the dielectric displacement vector must be equal to the surface-charge density, whereas the tangential component of the electric field must be continuous.

Because the boundary conditions have been set, the system of six equations [Eqs. (1)–(6)] can be solved, in principle. As it turns out, the solution of such a nonlinear system of coupled equations can be found only numerically even for one-dimensional p–n junctions with uniform acceptor and donor densities. To derive analytical approximations, we need to make certain assumptions regarding the physical makeup of the device and the degree of bias. These analytical expressions help predict the device response under reasonable bias, whereas the appreciation of their validity range provides an insight into the diode device physics.

DOPING-CARRIER PROFILES IN EQUILIBRIUM AND THE QUASI-NEUTRAL APPROXIMATION

As mentioned earlier, a p–n junction diode consists of an acceptor-doped p-region in contact with a donor-doped n-region. The two-dimensional area where the donor and acceptor densities are equal is called the metallurgical junction. Let's assume, for the moment, equilibrium conditions. In such a case, the currents are zero and the quasi-Fermi potentials are equal and spatially independent, $F_n = F_p = F$. Therefore, from Eqs. (1c) and (2c), $pn = n_i^2$, where n_i is the intrinsic-carrier density. Now, the six equations reduce to the Poisson equation, which, with the help of Eqs. (1c), (2c), and (6), takes the form

$$\nabla^2[E_i - (-eE)] = \frac{e^2}{\epsilon}\left[-n_i \exp\left(\frac{(-eF) - E_i}{kT}\right) + n_i \exp\left(\frac{E_i - (-eF)}{kT}\right) + N_D - N_A\right] \tag{10}$$

The last equation is known as the Boltzmann–Poisson equation. Approximate analytical solutions are possible when the donor and acceptor densities are uniform in the n- and p-regions, respectively. In this case, the field is zero, and the bands are flat everywhere except at and near the metallurgical junction. The finite field region around the metallurgical junction is called the space-charge region, whereas the zero-field regions are called quasi-neutral regions, for reasons to be explained shortly. In the n and p quasi-neutral regions, electrons and holes are the majority carriers, respectively. The majority-carrier densities are equal to the respective doping densities. With reference to the metallurgical junction, the space-charge region extends W_A and W_D within the p- and the n-regions. At zero bias, and in one dimension, an approximate solution of Eq. (10) gives

$$W_A = \sqrt{\frac{2\epsilon}{e} V_{bi} \frac{N_D}{N_A(N_A + N_D)}} \tag{11a}$$

$$W_D = \sqrt{\frac{2\epsilon}{e} V_{bi} \frac{N_A}{N_D(N_A + N_D)}} \tag{11b}$$

$$V_{bi} = \frac{kT}{e} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{11c}$$

In Eq. (11c) V_{bi} is the zero bias electrical potential difference, or barrier, between the p- and the n-side reflected in the level differences of the flat bands of each side. Such a barrier pre-

vents majority carriers from diffusing into the other side. These approximations result by assuming that the electron and hole densities are zero in the space-charge region. This is the depletion approximation, which reduces Eq. (10) to a linear second-order differential equation with constant terms and coefficients.

The zero-field condition for the rest of the *n*- and *p*-sides, outside the space-charge region, apparently justifies the term quasi-neutral regions. This term also applies when the *n*- and *p*-regions have gradually changing dopant profiles in the sense that the net space charge is much less than the majority-carrier charge. Here, by gradually changing we mean that the doping profile $N(x)$ in the quasi-neutral region must such that (3)

$$\frac{\epsilon kT}{e^2}\left|\nabla^2\left[\ln\frac{N(x)}{n_i}\right]\right| \ll N(x) \tag{12}$$

In such regions, the zero-bias majority-carrier density continues to be nearly the same as the net dopant density, but the electric field is not zero as in the uniform doping case.

FORWARD AND REVERSE BIAS CONDITIONS

The quasi-neutrality condition of the *n*- and *p*-regions is preserved even under bias, but now the boundaries with the space-charge region move appropriately to accommodate the new boundary conditions. This neutrality condition can be expressed as

$$n \approx p + N_{\mathrm{D}} - N_{\mathrm{A}} \tag{13}$$

Under a small forward bias, the applied voltage changes the electric field preferentially at the space-charge region, because it is the region with the fewest carriers, has the highest resistance, and is in series with more conductive *n*- and *p*-regions. The equilibrium barrier height V_{bi} lowers under forward bias, and the majority electrons overcoming the repulsive field diffuse from the *n*-side to the *p*-side, whereas the holes are doing the opposite. The carrier quasi-Fermi potentials are no longer equal, as shown in Fig. 3. The diffusion process, through the space-charge region and inside the quasi-neutral regions, increases dramatically the minority-carrier population on either side and gives rise to an appreciable electric current.

For forward voltages, the degree of bias defines three injection-level regimes distinguished by how the minority-carrier density compares to the majority one in the quasi-neutral regions. These regimes are the low-level, the moderate-level, and the high-level injection condition. In the low-level injection regime, the minority-carrier density is well below the majority-carrier density, and the electric field in the quasi-neutral regions is practically unaffected by the bias. As a result, the applied voltage drops across the space-charge region and reduces the barrier height from V_{bi} to $V_{\mathrm{b}} = V_{\mathrm{bi}} - V$. Provided the depletion approximation still holds, Eq. (11) still applies with V_{bi} being replaced by V_{b}. In low-level injection, the majority-carrier density is the same as at zero bias, as Eq. (13) points out, and is nearly equal to the net doping density. In the high-level injection regime, the minority-carrier injection is so intense that the injected carriers have densities far exceeding the dopant densities. Now, both carrier densities are about the same, $n = p$, to preserve neutrality in the quasi-neutral region. In other words, there is no real distinction between minority and majority carriers in terms of concentrations, but we obtain an electron-hole plasma having densities well above that of the dopant densities instead. In the moderate injection, the minority-carrier density approaches the order of magnitude of the majority-carrier density causing the majority-carrier density to start to increase, as Eq. (13) implies.

Under reverse bias, the built-in barrier increases in the space-charge region, the repulsive forces on the majority carriers coming from the quasi-neutral regions increase, and injection of minority carriers is not possible. The space-charge region is now totally depleted from both carriers, and a small leakage current exists as a result of thermal generation of electron and hole pairs in the depletion region.

Recombination Currents in the Steady State

Here, we will introduce the base and emitter terms as well as a general expression for the terminal current as the sum of recombination components. Between the two quasi-neutral regions, emitter is the one that is heavily doped, usually by diffusion or implantation, whereas the base is more lightly doped and occupies most of the substrate on which the device is made, at least in silicon. The heavy doping of the emitter excludes the possibility of moderate- or high-level injection conditions in this region. At forward bias, majority carriers from the emitter diffuse as minority carriers to the base where they recombine. Simultaneously, recombination occurs in the emitter, because minority carriers are back-injected from the base, as well as in the space-charge region. At steady state, $\partial n/\partial t = \partial p/\partial t = 0$, and in the dark $G = 0$. Now, the continuity equations [Eqs. (3)–(4)] become after volume integration:

$$I_{\mathrm{D}} = I_{\mathrm{e}} + I_{\mathrm{b}} + I_{\mathrm{SCR}} \tag{14}$$

where I_{D} is the terminal current and I_{e}, I_{b}, and I_{SCR} are the net recombination currents in the emitter, the base, and the space-charge region, respectively. Equation (14) expresses the total current as the sum of the recombination currents in the three regions of the device. Therefore, excess carrier recombination along with diffusion are the two basic transport mechanisms that determine the diode current at a given bias. The carrier recombination occurs either at the ohmic contacts, at the surface, or in the bulk. The minority carriers that arrive at the ohmic contact are supposed to recombine simultaneously there, to preserve the boundary condition, Eq. (7). The bulk recombination occurs either through traps or through band-to-band transitions. In terms of trap-mediated recombination, the Shockley–Read–Hall mechanism (4,5) is the most common:

$$U_{\mathrm{SRH}} = \frac{(pn - n_{\mathrm{i}}^2)N_{\mathrm{t}}}{\frac{1}{\sigma_p v_{\mathrm{th}}}\left[n + n_i \exp\left(\frac{E_{\mathrm{t}} - E_{\mathrm{i}}}{kT}\right)\right] + \frac{1}{\sigma_n v_{\mathrm{th}}}\left[p + n_i \exp\left(-\frac{E_{\mathrm{t}} - E_{\mathrm{i}}}{kT}\right)\right]} \tag{15}$$

In Eq. (15), N_{t} is the trap density, σ_{n} and σ_{p} are the electron and hole-capture cross sections, respectively, E_{t} is the trap

energy level in the gap and ν_{th} is the carrier thermal velocity. The band-to-band recombination is discussed in the heavy-doping effects section.

APPROXIMATE ANALYTICAL EXPRESSIONS IN THE STEADY STATE

As previously mentioned, the set of Eqs. (1)–(6) has no analytical solution in the general case. Approximate closed form expressions, though, are possible when low-level injection conditions dominate in the quasi-neutral base region. Without loss of generality, we assume that we are dealing with a p–n diode with a heavily doped p-emitter and an n-base. The steady state situation results when a terminal bias, say a terminal voltage V, is steadily applied on the terminals, and we wait long enough for the initial transient to disappear. The steady state version of the continuity equations [Eqs. (3)–(4)] is simplified because the time derivatives are set equal to zero. First, we will derive the expressions for the base current, and then extensions will be made for the recombination current in the emitter and the space-charge regions. If low-level injection conditions prevail in the base, then, to a good approximation, the original system of equations [Eqs. (1)–(6)] reduces to the minority-carrier equations [Eqs. (2) and (4)], which are now decoupled from Eq. (5) (the Poisson equation). This decoupling results because, as previously mentioned, at low-level injection the electric field is practically bias independent. Any small field variations would affect only the drift current of the majority carriers because of their high density; the minority carriers would not be influenced. That is why we focus on the minority-carrier transport to exploit the Poisson equation decoupling. Another reason for focusing on the minority carriers is the fact that the recombination in low-level injection, where $p \ll n$, can always be written as a linear function of their density:

$$U = (p - p_0)/\tau \tag{16}$$

In Eq. (16), p_0 is the equilibrium carrier density, whereas the variable τ, called minority-carrier lifetime, is the inverse of the derivative of the recombination rate with respect to the minority-carrier density. In the case of Shockley–Read–Hall recombination, $\tau = 1/\sigma_p \nu_{th} N_t$. Therefore, from Eqs. (2b), (4), and (16), we end up at

$$-\nabla \cdot (E\mu_p p' - D_p \nabla p') - p'/\tau + G = 0 \tag{17}$$

where, $p' = p - p_0$ is the excess minority-carrier density. Because of the field independence on p', Eq. (17) is linear and becomes homogenous if $G = 0$. In the later case, the solution is proportional to $p'(C_j)$, the excess minority-carrier density at the injecting boundary (Fig. 3).

Forward Bias and Low-Level Injection

Under forward bias, a basic assumption will be made. This assumption allows the coupling of the minority-carrier density to the externally applied terminal voltage: the Fermi levels are flat in the regions where the carriers are a majority and also in the space-charge region. Under this condition and from Eqs. (1c), (2c), and (7),

$$p(C_j)n(C_j) = n_i^2 \exp(eV/kT) \tag{18}$$

$$p'(C_j) = \frac{n_i^2[eV/kT) - 1]}{N_D(C_j)} \tag{19}$$

Equation (18) holds under any injection level, provided that the flat Fermi potential assumption holds, whereas Eq. (19) for the excess minority-carrier density holds only in low-level injection. The proportionality of the solution with respect to $p'(C_j)$ forces all carrier densities and currents to become proportional to the term $\exp(eV/kT) - 1$. Here, we note that the surface recombination is also a linear function of the excess minority-carrier density when $p \ll n$. More analytically, Eq. (9) becomes

$$J_p^n = eS_p p' \tag{20}$$

where S_p is called surface recombination velocity. Therefore, the total base recombination current in Eq. (14) is proportional to the term $\exp(eV/kT) - 1$. The same is true for the quasi-neutral emitter recombination. Thus, Eq. (14) becomes

$$I_D = (I_{oe} + I_{ob})[\exp(eV/kT) - 1] + I_{SCR} \tag{21}$$

where the pre-exponential factors I_{ob} and I_{oe} are called base and emitter saturation currents, respectively. Equation (19) points out that the saturation currents are proportional to n_i^2.

The space-charge region recombination is a current component that is hard to express in analytical terms. This difficulty arises because in this region the field depends on the bias and there is no such entity as a minority carrier. Consequently, the linearity conditions that allowed us to derive Eq. (21) no longer hold. To derive an approximate expression for the bias dependence of I_{SCR}, certain simplifications must be made throughout the space-charge region regarding the integral of Eq. (15). These simplifications result in a bias dependence of the form $\exp(eV/nkT) - 1$, where n, the ideality or slope factor, takes values from 1 to 2 (2). The specific value depends on the trap position in the energy gap, the doping profiles, and the cross section for hole capture relative to the cross section for electron capture. This range for n holds provided that the capture coefficients do not depend on the electric field. Now the pre-exponential factor I'_{SCR} is proportional to n_i. Finally, the expression for the forward current of a diode in the base of which low-level injection conditions prevail becomes

$$I_D = (I_{oe} + I_{ob})[\exp(eV/kT) - 1] + I'_{SCR}[\exp(eV/nkT) - 1] \tag{22}$$

For voltages higher than $3nkT/e$, the unity can be dropped from Eqs. (21) and (22). Because of a better slope factor, the emitter and base recombination will dominate the diode current for voltages above a certain level. Below this level, the space-charge region recombination must be considered too.

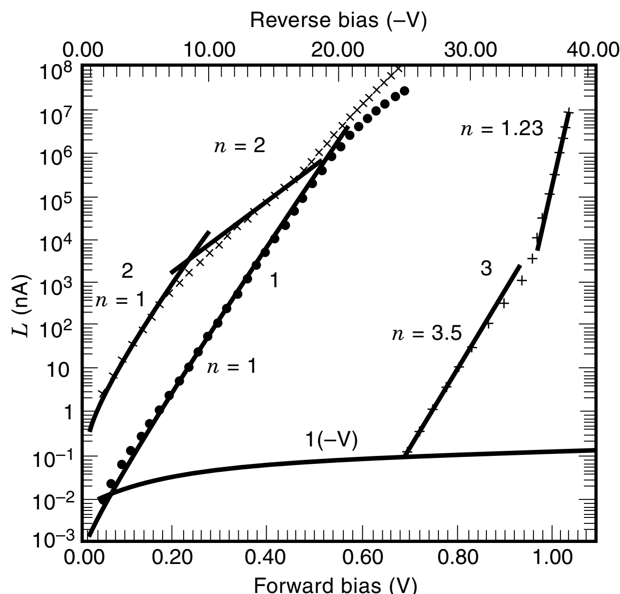

Figure 4. The experimental I–V characteristics of two different diodes. Diode 1 has a base thickness of 250 μm, a base-doping density of 5.5×10^{14} cm^{-3}, and an area of 2.9×2.9 mm^2. Diode 2 has a base thickness of 300 μm, a base-doping density of 4.5×10^{11} cm^{-3}, and an area of 5×5 mm^2. The base in both devices is of n-type. Plot 3 is the I–V characteristic of diode 2 at 78 K. The other plots are at 300 K. Curve 1(−V) is the reverse bias characteristic, with reversed sign, of diode 1 (top axis). The straight lines in curves 1, 2, and 3 are the exponential $\exp(eV/nkT) - 1$ fits to the experimental points. The slope factor n is also shown.

Such trends are seen in Fig. 4. Curve 1 is the I–V characteristic of a diode with a base doping of 5.5×10^{14} cm^{-3} and has an ideality factor of 1 in the bias range from 0.2 up to 0.4 V. For lower voltages, the space-charge region recombination slightly increases the ideality factor and makes the measured current deviate from the $\exp(eV/kT) - 1$ dependence. The ideality factor also increases for voltages above 0.4 V because of high-injection effects, which are discussed in the next subsection. The device of curve 2 has a very light doping density in the base, 4.5×10^{11} cm $^{-3}$ and is driven in high injection at even smaller bias. As discussed in the next subsection, curve 2 exhibits unity slope factor even at very low voltages.

Curve 3 shows what happens if the temperature is reduced to 78 K. The sharp reduction of the intrinsic-carrier density due to its $\cong \exp(-E_g/2kT)$ dependence requires much higher voltages to reach the same current as at 300 K. In fact, to reach a current density of 10 mA/cm^2, a voltage in excess of 1 V is required. The reduction of n_i reduces the recombination in the base and the emitter is much faster than in the space-charge region because the proportionality constants are n_i^2 and n_i, respectively. Therefore, at low temperatures, the bias regions with higher than 1 ideality factor are expected to be wider. This is evidenced in curve 3 of Fig. 4, where the ideality factor is 3.5 for voltages below 950 mV. The increase of the ideality factor above 2 is a result of the Poole–Frenkel effect, which reduces the effective energy separation of the traps from the bands (6). The influence of small values of n_i on the ideality factor is evident not only when the temperature drops but also when the band gap increases. In several compound semiconductor devices, their large band gap, compared to 1.1 eV of silicon, results in an intrinsic-carrier density, which is several orders of magnitude smaller than the 10^{10} cm^{-3} value for silicon at 300 K (7,8). Consequently, their I–V characteristics show slope factors substantially larger than 1 for the entire range of bias. On the contrary, germanium diodes have slope factors of 1 even at reduced temperatures because of the smaller gap, 0.66 eV, of the semiconductor.

The One-Dimensional Case. Equation (22) holds for any three-dimensional geometry and doping profiles because no assumption, except for low-level injection, was made so far regarding doping profiles and device topology. If, however, we want to express in close form the saturation values of the emitter and base recombination currents, then one-dimensional devices with uniform doping profiles must be considered. In such a case, the one-dimensional, homogenous, and constant-coefficient version of Eq. (17) becomes

$$\frac{d^2p'}{dx^2} = \frac{p'}{L_p^2} \tag{23}$$

where $L_p = \sqrt{D_p\tau}$ is the minority-carrier diffusion length. The first boundary condition for Eq. (23) is Eq. (19) applied at the injecting boundary. The other one refers to the ohmic contact. If it is an ideal ohmic contact deposited directly on the uniformly doped base, then the second boundary condition becomes, from Eq. (7), $p'(l) = 0$. Here, l is the base length and the coordinate origin is at the injecting boundary, as shown in Fig. 3. In many cases, between the ohmic contact and the uniformly doped base, a thin and heavily doped region intervenes.

This region has of thickness on the order of a micrometer and a doping of the same type as the rest of the base. The purpose of such a layer, called back-surface field, is to provide a better ohmic contact and to isolate the contact from the lightly doped base so that carrier recombination-generation is reduced (9). Such a back-surface field terminates the lightly doped base of diode 2 in Fig. 4 making it a p–i–n diode, where i stands for intrinsic. Therefore, in the presence of this contact layer, the base ends at a “low/high” n–n^+ junction. In terms of minority-carrier recombination, this interface is characterized by an effective recombination velocity S_{pe}, experienced by the minority carriers at the low side of the junction. The expression for S_{pe} is

$$S_{pe} = \frac{I_{oc}N_D}{en_i^2S} \tag{24}$$

where I_{oc} is the saturation value of the recombination current in the back-surface field and S is the device cross section. Equation (24) can be derived from Eq. (20), by applying Eq. (19) at the n–n^+ junction and by equating the minority current at the low/high junction to the recombination in the heavily doped region.

Under the previous boundary conditions, the solution of Eq. (23) yields for the base saturation current:

$$I_{ob} = S\frac{en_i^2}{N_D}\frac{D_p}{L_p}\,\frac{1 + \dfrac{D_p}{S_{pe}L_p}\tanh(\ell/L_p)}{\tanh(\ell/L_p) + \dfrac{D_p}{S_{pe}L_p}} \tag{25}$$

Equation (25) shows that, in terms of the one-dimensional geometry, the quantity that matters is the ratio l/L_p. Values of this ratio much less than one define the short base, whereas values above 3 define the long base. In the long base case, Eq. (25) becomes $I_{ob} = Sen_i^2D_p/(N_DL_p)$. Similar equations hold for a uniform emitter, too, but now the heavy doping effects could modify the value of n_i^2, as will be discussed in the section on heavy doping effects. If the base doping is very light, as in a p–i–n diode, then the increased value of I_{ob} will make the base recombination dominate the current components in Eq. (22). Accordingly, the influence of space-charge region recombination current on the slope factor will be suppressed even for voltages as low as a few kT/e, as shown in Fig. 4, curve 2. Also, by extrapolating the $\exp(eV/kT)$ fit of curve 2 at zero voltage, a base recombination current of 0.7 nA is obtained. This corresponds to a 300 K saturation current density of 2.8 nA/cm^2 compared to emitter saturation current densities on the order of pA/cm^2. On the other hand, diode 1, with a base-doping density three orders of magnitude higher than that in diode 2, exhibits a saturation current of 24 pA/cm^2. This saturation current comes mainly from the base recombination as a result of its relatively light doping density and the absence of a back-surface field which gives S_{pe} very high values.

Equation (25) applies to uniformly doped regions. If the doping is nonuniform, close form expressions are not possible, in the general case. This is the case because the one-dimensional version of Eq. (17) is still an ordinary differential equation with nonconstant coefficients. However, analytical approximations can be derived based on iterative techniques (10).

Diffusion in Three Dimensions. Equation (23) holds provided the cross-sectional dimensions of the diode are much larger than the diffusion length. Otherwise, lateral diffusion of minority carriers in the base becomes important. In such a case, the three-dimensional version of Eq. (23) takes the form

$$\nabla^2_{p'} = \frac{p'}{L_p^2} \tag{26}$$

The last equation can be solved very accurately by semianalytical techniques based on the two-dimensional Fourier transform (11). Simulation results are as shown in Fig. 5. As illustrated, in the case of a point contact diode having emitter dimensions of 0.1 L_p, the base recombination is expected to increase by a factor of 25 as a result of the lateral carrier diffusion.

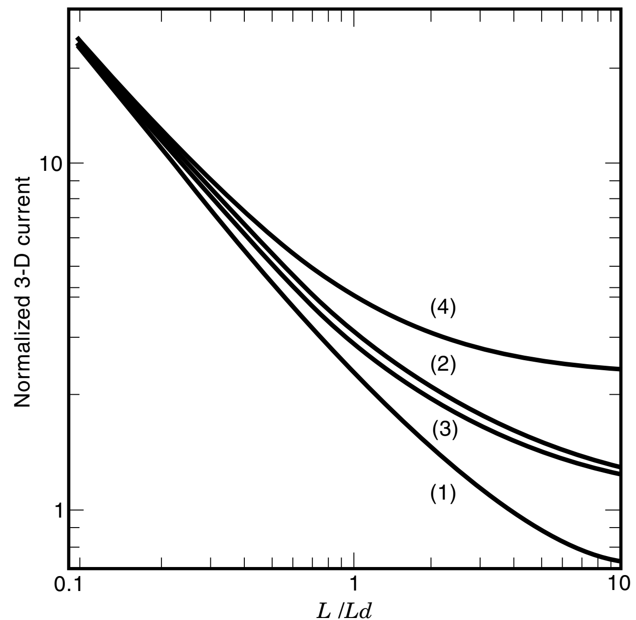

Figure 5. Three-dimensional diffusion base saturation current of a planar p–n junction with a square emitter having a side length L. The top surface of the base is supposed to have zero recombination velocity. The current is normalized with respect to the one-dimensional diffusion current $I_{ob} = en_i^2L^2D_p/N_DL_p$. In curve 2, $S_{pe} = 0$ and $l = L_p/2$. In curve 2, l is assumed to be infinite. In curve 3, $S_{pe} = D_p/L_p$ and $l = L_p/2$, whereas in curve 4, S_{pe} is assumed to be infinite and $l = L_p/2$.

High-Level Injection

So far, our analytical approaches were based on the low-level injection assumption. In high-level injection, where $n = p$, an equation similar to Eq. (23) can be derived where now the hole diffusion length is replaced by the ambipolar diffusion length (12). The boundary conditions, however, are not linear and depend on the electric field, which, now, is a function of bias. If the quasi-Fermi potentials are flat in the quasi-neutral base, then the electron-hole plasma density p is space independent and equals $n_i \exp(eV/2kT)$, as can be derived from Eq. (18). In such a case, by integrating the recombination current in the emitter the base and the back surface field, we obtain

$$I_D = (I_{oe} + I_{oc})\exp(eV/kT) + (eSl/\tau')\exp(eV/2kT)n_i \tag{27}$$

In Eq. (27), τ' is the high-injection lifetime defined as the ratio of the recombination rate divided by the plasma density. The flat Fermi level condition can easily be satisfied in p–i–n diodes where the light base doping density makes the high-level injection possible even at a bias of 0.4 V. In Fig. 4, curve 2 shows the $\exp(eV/2kT)$ dependence, or slope factor of 2, for voltages of about 0.4 V, which drive the p–i–n device to high-level injection. For even higher voltages, the emitter and back-surface field recombination in Eq. (27) starts dominating the current, and the slope factor drops again. For higher base doping densities, as in curve 1 of diode 1, the required voltage for high-level injection conditions could exceed 0.5 V at 300 K. Now, the heavily doped region recombination in Eq. (27) competes with the bulk recombination, and the slope factor of 2 does not appear. The bent of both curves 1 and 2 at voltages near 0.6 V is a result of series resistance effects, which invalidate the assumption of flat Fermi levels across the base. In such a case, the simulation is possible only by device simulators that solve the complete system of the transport equations.

Reverse Bias

Under reverse bias where $V < 0$, the assumption of flat Fermi levels across the space-charge region that led to Eq. (22) no

longer hold. On the other hand, however, the space-charge region can be considered to be fully depleted from free carriers. In such a case, Eq. (11) holds with V_{bi} replaced by $V_{bi} + |V|$. Therefore, the depleted space-charge region will expand toward the base according to the square root of the bias for $|V| > 5$ V. In this region, the Shockley–Read–Hall Eq. (15) predicts a negative recombination or generation of electron-hole pairs. This generation current is the basic component of the leakage current in reverse bias. The contribution of the diffusion components from the base and the emitter, $-I_{ob} - I_{oe}$, is usually negligible unless the base is very lightly doped. The bottom line in Fig. 4 shows the reverse-bias current for diode 1. The square-root dependence on voltage is not exactly obeyed because of the Poole–Frenkel effect, which increases the generation rate at higher fields.

TRANSIENT RESPONSE OF DIODES

If a diode is subjected to a transient terminal bias, then in addition to currents due to carrier diffusion and recombination, we also have the dielectric displacement current resulting from the time dependence of the electric field. If low-level injection is observed in the quasi-neutral regions, the displacement current is restricted in the space-charge region. At the same time, low-level injection ensures that linearity holds in the base and the emitter, and Eq. (17) still applies with $\partial p'/\partial t$ replacing zero in the right-hand side of the relation. The solution of the time-dependent edition of Eq. (17) provides the minority-carrier currents at the injecting boundaries of the base and the emitter, $I_b(t)$ and $I_e(t)$, respectively. These currents have now two components: the minority-carrier recombination and the minority-carrier storage current $\partial Q'/\partial t$, where Q' is the total excess minority-carrier charge. To calculate the total transient current, reconsider Eq. (14) in its transient version. Therefore, in addition to $I_b(t)$ and $I_e(t)$, the transient space-charge-region current is required. Unlike the base and the emitter, this current in addition to the recombination and storage component also includes the displacement current (13). In so far as the displacement current is concerned, the space-charge region behaves as a parallel plate capacitor with a plate distance $W = W_A + W_D$, Eq. (11), a dielectric constant ϵ and a capacitance $C_{SCR} = \epsilon S / W$. During transit, the dielectric displacement current is supplied by the majority carriers from either side of the junction.

To calculate the transient currents in the base and the emitter, the boundary conditions must be defined. Boundary condition Eq. (20) holds because of linearity. The other condition at the injecting boundary depends on the kind of transient to be considered (14). Here we will assume that the device is in equilibrium for $t < 0$, whereas at $t = 0$ a constant voltage V is applied. We can now assume that Eq. (19) applies with $p'(C_j)$ replaced by $p'(C_j, t)$ for $t > 0$. This assumption has a validity range depending on how fast the flat quasi-Fermi potential condition can be established across the space-charge region. As matter of fact, even in the absence of series resistance effects, it takes a short time for this condition to be established. This short time relates to the dielectric response time of the majority carriers and the minority-carrier diffusion time across the space-charge region (14). For almost all practical cases, the delay in establishing a fixed minority-carrier density at the edge of the quasi-neutral region will not exceed the limit of a few tens of a picosecond (14), in the absence of series resistance effects. Therefore, if the time granularity used in solving the time-dependent version of Eq. (17) is restricted to about a nanosecond, then the solutions will be accurate. In practical cases, however, the very first part of the transient current, following the sudden application of a voltage, will be determined by charging C_{CSR} through the series resistance of the majority carriers in the base and the emitter. The respective time constant could be on the order of a nanosecond. In such a case, the minority-carrier transport in the base will determine the transient only after several nanoseconds have elapsed since the application of the voltage. The transient base transport can be expressed in semianalytical forms using Laplace transform techniques (14), especially in the case of uniform and one-dimensional quasi-neutral regions. In a long-base diode, the transition will last for about a minority-carrier lifetime. In a short-base device with an ohmic contact at the base end, the transient will last approximately $l^2/2D_p$, which is the minority-carrier diffusion time through the base.

Small-Signal Response

In many cases, the device operates under sinusoidal small-signal excitation superimposed on a steady state excitation. In such cases, Eq. (17) still holds, but now $1/\tau$ will have to be replaced by $1/\tau + j\omega$, where j is the imaginary unit and ω is the angular frequency of the excitation. This is the case because the time derivative of the small-signal carrier density is the carrier density amplitude times $j\omega$. Having done the complex lifetime replacement, the analysis that followed Eq. (17) still holds. Now, however, the small-signal value of the excess minority-carrier density at the injecting boundary will be the steady state value in Eq. (19) times ev/kT. Here, v is the small-signal terminal voltage, which is supposed to be much less than kT/e. Under low-level injection and in view of the previous transient response discussion, the small-signal version of Eq. (22) will refer to a terminal current I_D^* having a real and an imaginary component:

$$I_D^* = (I_{oe}^* + I_{ob}^*)\,\frac{ev}{kT}\left[\exp\left(\frac{eV}{kT}\right) - 1\right] + I_{SCR}^*\frac{ev}{nkT}\left[\exp\frac{eV}{nkT} - 1\right] + j\omega C_{SCR}v = v(G + j\omega C) \tag{28}$$

The star exponents denote the complex values of the saturation currents as a result of the complex lifetime. In Eq. (28), G and C are the diode small-signal parallel conductance and capacitance, respectively. These two components are of great importance because their frequency dependence can reveal minority-carrier properties, such as diffusivity and lifetime (15) and allow the device circuit representation when the diode is part of a greater small-signal circuit. For uniformly doped quasi-neutral regions, I_{oe}^* and I_{oe}^* can be obtained from Eq. (25) by replacing the diffusion length $L = (D\tau)^{1/2}$ with the complex diffusion length $L^* = L/(1 + j\omega\tau)^{1/2}$. For frequencies sufficiently high, the magnitude of the complex diffusion length will become much shorter than the base thickness. Then, the complex version of Eq. (25) predicts that the base current would change as $1/L^*$. If the base component in Eq.

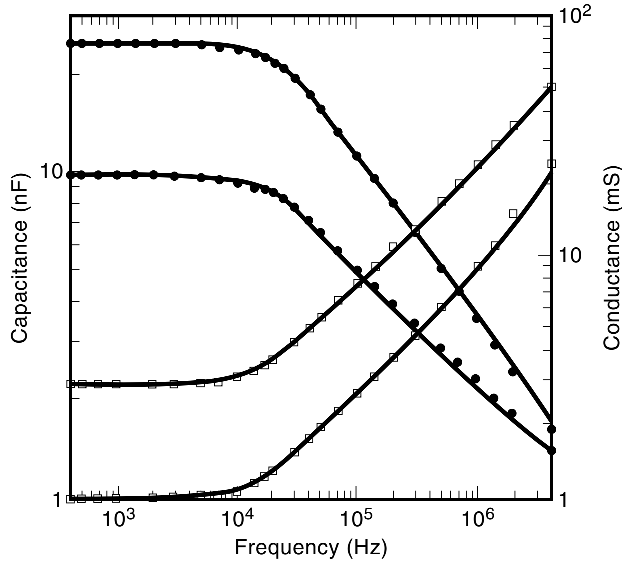

Figure 6. Experimentally measured capacitance (dots) and conductance (squares) at 300 K for diode 1. The bottom and the top curves correspond to two different bias points: 420 mV and 450 mV, respectively. The solid curves are the theoretical fits from the equivalent circuit of Fig. 7.

(28) were to dominate, then beyond a certain frequency, C would change as $\omega^{-1/2}$ while G would change as $\omega^{1/2}$.

This frequency dependence is confirmed in Fig. 6, which shows the frequency response of diode 1, from Fig. 4, at two bias points. The theoretical fit to the experimental results was obtained on the basis of the diode equivalent circuit shown in Fig. 7. This circuit includes all the components relating to carrier injection and storage in the device's three regions in accordance with Eq. (28). It also includes the base resistance R_N, which has been ignored in Eq. (28). In Fig. 6, the square root law is better obeyed at the higher bias point and for frequencies less than 1 MHz, especially for the capacitance. This is a combined result of the space-charge-region capacitance, the relative contribution of which increases at lower bias, and the series resistance, the influence of which is stronger at

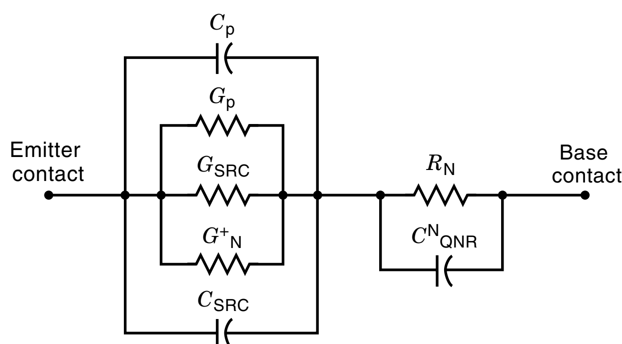

Figure 7. The equivalent circuit model of a diode. The base injection currents in Eq. (28) corresponds to C_p, imaginary part, and G_p, real part. The injection in the emitter is represented by G_N^+. The space-charge region recombination is represented by G_{SCR}, whereas C_{SRC} is the space-charge region capacitance. The rest of the components are accounted for in the text.

high frequencies. The corner frequencies of the conductance and the capacitance depend on the base thickness and the lifetime. The fit shown in Fig. 6 gave a minority-hole lifetime in the n-type base of about 30 μs. Such a lifetime and Eq. (25) imply that the saturation current density of 24 pA/cm^2 at 300 K, as shown in Fig. 4, is 90% due to base recombination. The emitter contribution of 10% is discussed in the section on heavy doping effects.

The series resistance R_N becomes the bulk majority-carrier resistance under reverse bias or even under forward bias, provided that the frequency is high ($|L^*| \ll l$). The capacitance C_{QNR}^N in parallel with R_N, as shown in Fig. 7, is the geometrical capacitance of the quasi-neutral base (16,17). For ordinary resistivity devices, it can be ignored unless the frequency is in the gigahertz range. However, for diodes made on high resistivity substrates, this capacitance must be considered especially at reverse bias and high frequencies (17). From Fig. 7 and in the limit of very high frequencies under forward bias, the parallel conductance saturates at $1/R_N$ whereas the parallel capacitance does so at C_{QNR}^N. This is because of the combination of the increasing injection conductancies and the space-charge-region capacitance. Then, the product $R_N\ C_{QNR}^N$ becomes the dielectric response time of the majority carriers in a uniform base. At high injection, the parallel conductance will saturate at the sum of the two carrier conductances (18). Under reverse bias, the circuit of Fig. 7 reduces to the space-charge region capacitance in series with the parallel combination of C_{QNR}^N and R_N. Unlike the forward bias case, where the circuit parameters depend roughly exponentially on the terminal voltage V, in reverse bias the voltage dependence would be restricted to $V^{-1/2}$. In the sense of the voltage dependence, the circuit of Fig. 7 is the circuit of a varactor.

HEAVY DOPING EFFECTS IN THE EMITTER

In the previous subsection, the emitter saturation current density was estimated to be about 2 pA/cm^2. From Eq. (25) and by assuming microsecond lifetimes, we would expect saturation currents on the order of a fA/cm^2 from an emitter doped in the range 10^{19} to 10^{20} cm^{-3}. Such a discrepancy by three orders of magnitude is due to the heavy doping effects, namely the short lifetime resulting from Auger recombination and the effective increase of n_i due to band-gap narrowing. In the Auger recombination process, a minority carrier recombines directly with a majority one, and the energy is transferred to another majority carrier. Because of such kinetics, the Auger minority-carrier lifetime is inversely proportional to the square of the majority-carrier density. The proportionality constant is $\sim 10^{-31}$ cm^6/s for minority electrons in p^+ emitters and 3×10^{-31} cm^6/s for minority holes in n^+ emitters (19). In heavily doped regions, the Auger recombination rate is by far higher than the Shockley–Read–Hall rate and determines the lifetime. Therefore, nanosecond lifetimes are expected, especially for holes, in emitters doped in the vicinity of 10^{20} cm^{-3}.

In a heavily doped region, every minority carrier interacts strongly with the majority carriers because of their high density. The minority-majority carrier attraction along with the carrier-dopant interaction and the semiconductor lattice random disruption by the dopant atoms reduces the bang gap and changes the density of states in both bands (20,21). The

net result is an effective shrinkage of the gap depending on the doping type and density (22–24). This shrinkage changes the intrinsic-carrier density n_i to a much higher effective n_{ie}. The result of the band distortion is that the original system of transport equations [Eqs. (1)–(6)] no longer holds. More specifically, Eqs. (1b,c) and (2b,c) are not valid for the majority carriers even if n_{ie} substitutes n_i because Boltzmann statistics must be replaced by Fermi–Dirac statistics. Also, Eq. (6) no longer holds in a nonuniform region because the band edges are not parallel any more and each carrier experiences a different field. However, the minority carriers still follow the Boltzmann statistics, and Eq. (17) holds for the minority carriers. Now E is the minority-carrier field ($1/e\nabla E_c$ for electrons), and the boundary condition Eq. (19) is valid with n_i replaced by n_{ie}. Therefore, Eq. (25) still applies for the minority-carrier recombination in a uniformly doped emitter. For an emitter doped at about 10^{20} cm^{-3}, a gap narrowing of about 100 meV is expected (22–24), which makes n_{ie} several tens higher than n_i. If such an n_{ie} as well as nanosecond lifetimes replace n_i and microsecond lifetimes in Eq. (25), an emitter saturation current on the order of pA/cm^2 is predicted, in accordance with the experimental results of the previous section.

DIODES OF NONCONVENTIONAL TRANSPORT

So far in this article, devices based on the drift and diffusion model of Eqs. (1) and (2) were studied. Charge carriers can be transported from one region to another by tunneling. Also, they can be temporarily trapped in energy-gap states, atom clusters, or crystallites imbedded in insulating films, thereby affecting the tunneling or the conventional transport of the free carriers.

In this respect, the first device to be examined is the p–i–n diode 2 of Fig. 4, operating at cryogenic temperatures. Around 4.2 K, the equilibrium Fermi level in the lightly doped n^- region is pinned at the donor level. These levels, now, are not ionized except for a fraction to compensate the charge of the unintentionally introduced acceptor ions. At such low temperatures, there are no free carriers in the base, and no measurable conduction is possible unless the voltage is raised enough to achieve the flat-band condition (25,26). For silicon, this voltage V_0 would be about 1.1 V. For even higher voltages, conduction is possible only if electrons and holes can be injected in the frozen substrate from the n- and p-regions, respectively. In this sense, Eq. (24) based on the assumption of flat majority-carrier Fermi levels no longer holds. For $T < 10$ K, injection is possible by carrier tunneling through the small potential barrier existing at each of the p–i and i–n interfaces (26). These barriers exist because of the band distortion in the heavily doped regions and the smaller gap there, as outlined in the previous section. For $V > V_0$, electrons tunnel in the i-layer, and the higher the forward bias, the higher the current due to a field-induced effective lowering of the barriers.

As shown in Fig. 8, for temperatures below 10 K it takes at least several volts to establish a current of few nanoamps. The injected electrons in the i-layer are trapped by the ionized donors and built a space charge and a subsequent potential barrier. For even higher voltages approaching 10 V, the barrier at the i–p interface lowers, holes now enter the i-layer

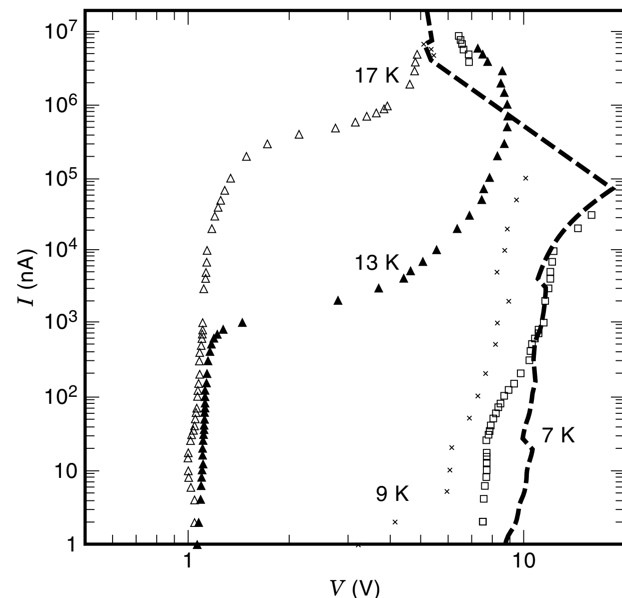

Figure 8. Measured I–V characteristics of diode 2 at cryogenic temperatures. The square points correspond to 4.2 K. The $T < 10$ K plots exhibit a distinct voltage breakdown. Reprinted from K. Misiakos, D. Tsamakis, and E. Tsoi, Measurement and modeling of the anomalous dynamic response of high resistivity diodes at cryogenic temperatures, *Solid State Electronics,* **41**: 1099–1103, 1997, with kind permission from Elsevier Science Ltd, The Boulevard, Langford Lane, Kidlington 0X5 1GB, UK.

in large numbers. Their charge neutralizes the trapped electron charge and causes the voltage breakdown and the negative differential resistance that appears in Fig. 8 for $T < 10$ K. The negative resistance persists and beyond breakdown as a result of new carrier generation by the impact ionization of occupied shallow donors by the injected carriers. The interaction of free and trapped carriers through impact ionization gives rise to a negative dynamic conductance and capacitance which for frequencies high enough change as ω^{-2} (27). For $T > 10$ K the injection mechanism changes to thermion emission over the interface potential barriers, whereas the space-charge effects are now less pronounced.

Another example of tunneling injection mechanism is the breakdown effect in zener diodes. Here, the base is quite heavily doped ($\approx 10^{18}$ cm^{-3}), and the strong electric field in the space-charge region increases even further by applying a reverse bias. For fields approaching 10^6 V/cm, a valence band electron can tunnel to a conduction band state of the same energy. This way, electron–hole pairs are created, and the reverse current sharply increases. Another diode structure based on tunneling is a new metal–insulator–semiconductor device having silicon nanocrystals imbedded in the thin insulating film (28). One way to realize such diodes is by depositing an aluminum electrode on a thin (on the order of 10 nm) SiO_2 layer containing silicon nanocrystals. The substrate is n-type crystalline silicon. The silicon nanocrystals can be created either by oxidizing deposited amorphous silicon layers (28) or by low-energy silicon-ion implantation in the SiO_2 film (29). In the absence of the nanocrystals, by applying a negative voltage of a few volts on the aluminum electrode relative to the n-type silicon substrate, only a small tunneling current would be present.

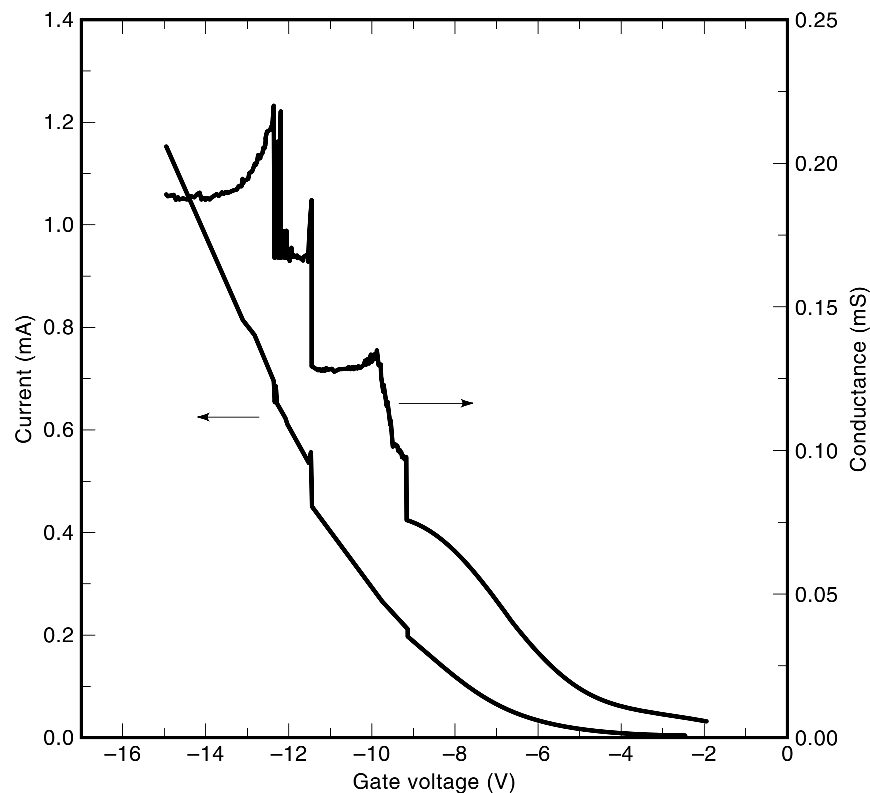

Figure 9. Current and conductance plots of a reverse biased quantum-dot diode. The conductance peaks correspond to steps in current curve. Reprinted from P. Normand et al., Silicon nanocrystal formation in thin thermal-oxide films by very low energy Si^{+} ion implantation, *Microelectronic Engineering,* **36** (1–4): 79–82, 1997, with kind permission of Elsevier Science-NL, Sara Burgerharstraat 25, 1055 KV Amsterdam, The Netherlands.

When the nanocrystals are introduced, much higher currents are observed while the conductance cuve exhibits characteristic peaks. Such peaks are shown in Fig. 9 showing the reverse current and conductance of a quantum dot diode formed by low-energy implantation of silicon in a 10 nm SiO_2 layer (29). The conductance peaks appear when the metal Fermi level is swept across the discrete energy states of the nanocrystals, thus enabling resonant tunneling from the metal to the semiconductor (28). The three-dimensional confinement of electrons in the quantum-box crystallites creates a large separation between energy states, which along with the Coulomb blockade effect of the occupied states explains the large voltage separation of the three first conductance peaks in Fig. 9 (28,29). Such quantum dot devices hold the promise of single-electron transistors (30) and silicon-based light emitting diodes (31).

Ending this article, we would like to mention the basic uses of the diode as a device. The most frequent use of the diode is the protection of complementary metal oxide semiconductor (CMOS) integrated circuits from electrostatic discharges by clamping the output pads to the power-supply voltages through reverse-biased *p–n* junctions. In analog integrated circuits, forward-biased diodes are used for voltage shifting. Such diodes usually come from properly wired bipolar transistors (e.g., emitter-base diodes with base-collector short circuited). Diodes, as discrete devices, find applications mainly as rectifying elements in power circuits. The breakdown effect of zener diodes makes these devices useful as voltage reference sources in power supplies. Photodiodes are widely used for detecting photons or charge particles. Finally, large area diodes with exposed front surface and proper design and engineering can efficiently convert solar light into electricity and are used as solar cells (32).

BIBLIOGRAPHY

1. W. Shockley, *Electrons and Holes in Semiconductors,* Princeton, NJ: Van Nostrand, 1950.
2. C. T. Sah, R. N. Noyce, and W. Shockley, Carrier generation and recombination in p-n junction and p-n junction characteristics, *Proc. IRE,* **45**: 1228–1243, 1957.
3. J. A. del Alamo, Charge neutrality in heavily doped emitters, *Appl. Phys. Lett.,* **39**: 435–436, 1981.
4. W. Shockley and W. T. Read, Statistics of the recombination of holes and electrons, *Phys. Rev.,* **87**: 835–842, 1952.
5. R. N. Hall, Electron-hole recombination in germanium, *Phys. Rev.,* **87**: 387, 1952.
6. J. C. S. Woo, J. D. Plummer, and J. M. C. Stork, Non-ideal base current in bipolar transistors at low temperatures, *IEEE Trans. Electron Devices,* **34**: 131–137, 1987.
7. A. B. Sproul and M. A. Green, Intrinsic carrier concentration and minority carrier mobility from 77 K to 300 K, *J. Appl. Phys.,* **74**: 1214–1225, 1993.
8. K. Misiakos and D. Tsamakis, Accurate measurements of the intrinsic carrier density from 78 to 340 K, *J. Appl. Phys.,* **74**: 3293–3297, 1993.
9. J. G. Fossum, Physical operation of back surface field solar cells, *IEEE Trans. Electron Devices,* **24**: 322–325, 1977.
10. J. S. Park, A. Neugroschel, and F. A. Lindholm, Systematic analytical solution for minority-carrier transport in semiconductors with position dependent composition with application to heavily doped silicon, *IEEE Trans. Electron Devices,* **33**: 240–249, 1986.
11. S. Kavadias and K. Misiakos, Three-Dimensional simulation of planar semiconductor diodes, *IEEE Trans. Electron Devices,* **40**: 1875–1878, 1993.
12. S. M. Sze, *Physics of Semiconductor Devices,* 2nd ed., New York: Wiley, 1981, p. 87.

13. F. A. Lindholm, Simple phenomenological model of transition region capacitance of forward biased p-n junction diodes or transistor diodes, *J. Appl. Phys.,* **53**: 7606–7608, 1983.

14. T. Jung, F. A. Lindholm, and A. Neugroschel, Unifying view of transient responses for determining lifetime and surface recombination velocity in silicon diodes and back-surface field solar cells with application to experimental short circuit current decay, *IEEE Trans. Electron Devices,* **31**: 588–595, 1984.

15. A. Neugroschel et al., Diffusion length and lifetime determination in p-n junction solar cells and diodes by forward biased capacitance measurements, *IEEE Trans. Electron Devices,* **25**: 485–490, 1978.

16. B. M. Vul and E. I. Zavatitskaya, The capacitance oh p/n junctions at low temperatures, *Sov. Phys.–JETP, (Engl. Transl.),* **11**: 6–11, 1960.

17. S. Kavadias et al., On the equivalent circuit model of reverse biased diodes made on high resistivity substrates, *Nucl. Instrum. Methods Phys. Res.,* **A322**: 562–565, 1992.

18. K. Misiakos and D. Tsamakis, Electron and hole mobilities in lightly doped silicon, *Appl. Phys. Lett.,* **64**: 2007–2009, 1994.

19. J. Dziewior and W. Schmid, Auger coefficients for lightly doped and highly excited silicon, *Appl. Phys. Lett.,* **31**: 346–348, 1977.

20. G. D. Mahan, Energy gap in Si and Ge: Impurity dependence, *J. Appl. Phys.,* **51**: 2634–2646, 1980.

21. P. T. Landsberg et al., A model for band-p shrinkage in semiconductors with application to silicon, *Phys. Status Solidi B,* **130**: 255–266, 1985.

22. J. W. Slotboom and H. C. de Graaff, Measurements of band gap narrowing in Si bipolar transistors, *Solid-State Electron,* **19**: 857–862, 1976.

23. A. W. Wieder, Emitter effects in shallow bipolar devices: measurements and consequences, *IEEE Trans. Electron Devices,* **27**: 1402–1408, 1980.

24. J. A. del Alamo and R. M. Swanson, Measurement of steady-state minority-carrier recombination in heavily doped n-type silicon, *IEEE Trans. Electron Devices,* **34**: 1580–1589, 1987.

25. A. K. Jonscher, p-n junctions at very low temperatures, *Br. J. Appl. Phys.,* **12**: 363–371, 1961.

26. Y. N. Yang, D. D. Coon, and P. F. Shepard, Thermionic emission in silicon at temperatures below 30 K, *Appl. Phys. Lett.,* **45**: 752–754, 1984.

27. K. Misiakos, D. Tsamakis, and E. Tsoi, Measurement and modeling of the anomalous dynamic response of high resistivity diodes at cryogenic temperatures, *Solid-State Electronics,* **41**: 1099–1103, 1997.

28. E. H. Nicollian and R. Tsu, Electrical properties of a silicon quantum dot diode, *J. Appl. Phys.,* **74**: 4020–4025, 1993.

29. P. Normand et al., Silicon nanocrystal formation in thin thermal-oxide films by very-low energy Si^+ ion implantation, *Microelectronic Engineering,* **36** (1–4): 79–82, 1997.

30. K. Yano et al., Romm-temperature single-electron memory, *IEEE Trans. Electron Devices,* **41**: 1628–1638, 1994.

31. D. J. Dimaria et al., Electroluminescence studies in silicon dioxide films containing tiny silicon islands, *J. Appl. Phys.,* **56**: 410, 1984.

32. F. Zhang, S. Wenham, and M. A. Green, Large area, concentrator buried contact solar cells, *IEEE Trans. Electron Devices,* **42**: 145–149, 1995.

KONSTANTINOS MISIAKOS
NCSR "Demokritos"

DIODES, AVALANCHE. See AVALANCHE DIODES.