2

p-n Junctions

2.1 INTRODUCTION
2.2 DEPLETION REGION
2.3 CURRENT-VOLTAGE CHARACTERISTICS
2.4 JUNCTION BREAKDOWN
2.5 TRANSIENT BEHAVIOR AND NOISE
2.6 TERMINAL FUNCTIONS
2.7 HETEROJUNCTIONS

2.1 INTRODUCTION

p-n junctions are of great importance both in modern electronic applications and in understanding other semiconductor devices. The *p-n* junction theory serves as the foundation of the physics of semiconductor devices. The basic theory of current-voltage characteristics of *p-n* junctions was established by Shockley.[1,2] This theory was then extended by Sah, Noyce, and Shockley[3], and by Moll.[4]

The basic equations presented in Chapter 1 are used to develop the ideal static and dynamic characteristics of *p-n* junctions. Departures from the ideal characteristics due to generation and recombination in the depletion layer, to high injection, and to series resistance effects are then discussed. Junction breakdown, especially that due to avalanche multiplication, is considered in detail, after which transient behavior and noise performance in *p-n* junctions are presented.

A *p-n* junction is a two-terminal device. Depending on the doping profile, device geometry, and biasing condition, a *p-n* junction can perform various terminal functions which are considered briefly in Section 2.6. The chapter closes with a discussion of an important group of devices—the heterojunctions, which are junctions formed between dissimilar semiconductors (e.g., *n*-type GaAs on *p*-type AlGaAs).

2.2 DEPLETION REGION

2.2.1 Abrupt Junction

Built-in Potential and Depletion-Layer Width. When the impurity concentration in a semiconductor changes abruptly from acceptor impurities N_A to donor impurities N_D, as shown in Fig. 1a, one obtains an abrupt junction. In particular, if $N_A \gg N_D$ (or vice versa), one obtains a one-sided abrupt p^+-n (or n^+-p) junction.
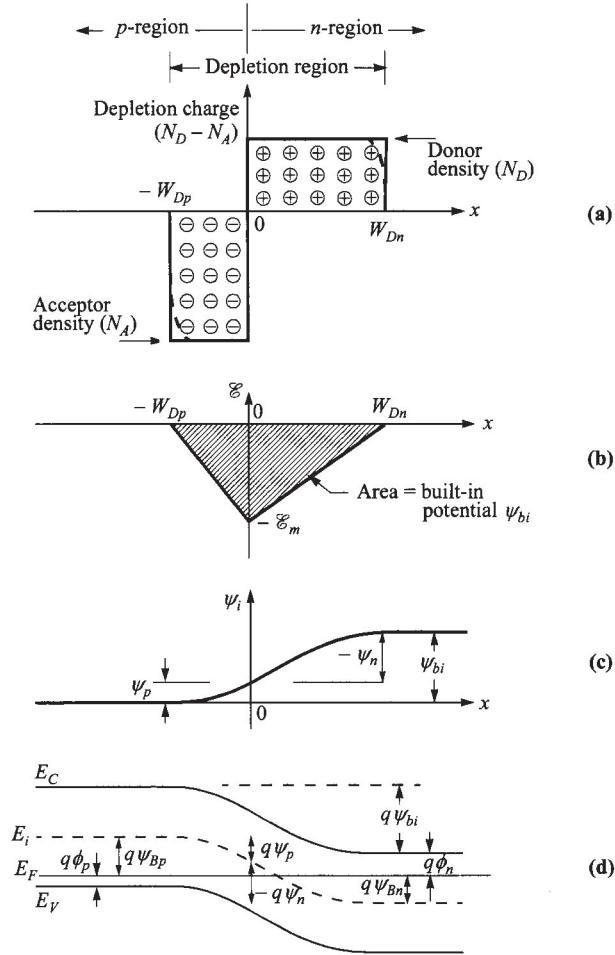

Fig. 1   Abrupt *p-n* junction in thermal equilibrium. (a) Space-charge distribution. Dashed lines indicate corrections to depletion approximation. (b) Electric-field distribution. (c) Potential distribution where ψ_{bi} is the built-in potential. (c) Energy-band diagram.

We first consider the thermal equilibrium condition, that is, one without applied voltage and current flow. From the current equation of drift and diffusion (Eq. 156a in Chapter 1),

$$J_n = 0 = q\mu_n\left(n\mathscr{E} + \frac{kT}{q}\frac{dn}{dx}\right) = \mu_n n\frac{dE_F}{dx} \tag{1}$$

or

$$\frac{dE_F}{dx} = 0. \tag{2}$$

Similarly,

$$J_p = 0 = \mu_p p\frac{dE_F}{dx}. \tag{3}$$

Thus the condition of zero net electron and hole currents requires that the Fermi level must be constant throughout the sample. The built-in potential ψ_{bi}, or diffusion potential, as shown in Fig. 1b, c, and d, is equal to

$$q\psi_{bi} = E_g - (q\phi_n + q\phi_p) = q\psi_{Bn} + q\psi_{Bp}. \tag{4}$$

For nondegenerate semiconductors,

$$\psi_{bi} = \frac{kT}{q}\ln\left(\frac{n_{no}}{n_i}\right) + \frac{kT}{q}\ln\left(\frac{p_{po}}{n_i}\right)$$

$$\approx \frac{kT}{q}\ln\left(\frac{N_D N_A}{n_i^2}\right). \tag{5}$$

Since at equilibrium $n_{no}p_{no} = n_{po}p_{po} = n_i^2$,

$$\psi_{bi} = \frac{kT}{q}\ln\left(\frac{p_{po}}{p_{no}}\right) = \frac{kT}{q}\ln\left(\frac{n_{no}}{n_{po}}\right). \tag{6}$$

This gives the relationship between carrier densities on either side of the junction.

If one or both sides of the junction are degenerate, care has to be taken in calculating the Fermi-levels and built-in potential. Equation 4 has to be used since Boltzmann statistics cannot be used to simplify the Fermi-Dirac integral. Furthermore, incomplete ionization has to be considered, i.e., $n_{no} \neq N_D$ and/or $p_{po} \neq N_A$ (Eqs. 34 and 35 of Chapter 1).

Next, we proceed to calculate the field and potential distribution inside the depletion region. To simplify the analysis, the depletion approximation is used which assumes that the depleted charge has a box profile. Since in thermal equilibrium the electric field in the neutral regions (far from the junction at either side) of the semiconductor must be zero, the total negative charge per unit area in the p-side must be precisely equal to the total positive charge per unit area in the n-side:

$$N_A W_{Dp} = N_D W_{Dn}. \tag{7}$$

From the Poisson equation we obtain

$$-\frac{d^2\psi_i}{dx^2} = \frac{d\mathscr{E}}{dx} = \frac{\rho(x)}{\varepsilon_s} = \frac{q}{\varepsilon_s}[N_D^+(x) - n(x) - N_A^-(x) + p(x)]. \tag{8}$$

Inside the depletion region, $n(x) \approx p(x) \approx 0$, and assuming complete ionization,

$$\frac{d^2\psi_i}{dx^2} \approx \frac{qN_A}{\varepsilon_s} \qquad \text{for } -W_{Dp} \leq x \leq 0, \tag{9a}$$

$$-\frac{d^2\psi_i}{dx^2} \approx \frac{qN_D}{\varepsilon_s} \qquad \text{for } 0 \leq x \leq W_{Dn}. \tag{9b}$$

The electric field is then obtained by integrating the above equations, as shown in Fig. 1b:

$$\mathscr{E}(x) = -\frac{qN_A(x + W_{Dp})}{\varepsilon_s} \qquad \text{for } -W_{Dp} \leq x \leq 0, \tag{10}$$

$$\mathscr{E}(x) = -\mathscr{E}_m + \frac{qN_D x}{\varepsilon_s}$$

$$= -\frac{qN_D}{\varepsilon_s}(W_{Dn} - x) \qquad \text{for } 0 \leq x \leq W_{Dn} \tag{11}$$

where $\mathscr{E}_m$ is the maximum field that exists at $x = 0$ and is given by

$$|\mathscr{E}_m| = \frac{qN_D W_{Dn}}{\varepsilon_s} = \frac{qN_A W_{Dp}}{\varepsilon_s}. \tag{12}$$

Integrating Eqs. 10 and 11 once again gives the potential distribution $\psi_i(x)$ (Fig. 1c)

$$\psi_i(x) = \frac{qN_A}{2\varepsilon_s}(x + W_{Dp})^2 \qquad \text{for } -W_{Dp} \leq x \leq 0, \tag{13}$$

$$\psi_i(x) = \psi_i(0) + \frac{qN_D}{\varepsilon_s}\left(W_{Dn} - \frac{x}{2}\right)x \qquad \text{for } 0 \leq x \leq W_{Dn}. \tag{14}$$

With these, the potentials across different regions can be found as:

$$\psi_p = \frac{qN_A W_{Dp}^2}{2\varepsilon_s}, \tag{15a}$$

$$|\psi_n| = \frac{qN_D W_{Dn}^2}{2\varepsilon_s}, \tag{15b}$$

(ψ_n is relative to the n-type bulk and is thus negative. See definition in Appendix A)

$$\psi_{bi} = \psi_p + |\psi_n| = \psi_i(W_{Dn}) = \frac{|\mathscr{E}_m|}{2}(W_{Dp} + W_{Dn}) \tag{16}$$

where $\mathscr{E}_m$ can also be expressed as:

$$|\mathscr{E}_m| = \sqrt{\frac{2qN_A\psi_p}{\varepsilon_s}} = \sqrt{\frac{2qN_D|\psi_n|}{\varepsilon_s}}. \tag{17}$$

From Eqs. 16 and 7, the depletion widths are calculated to be:

$$W_{Dp} = \sqrt{\frac{2\varepsilon_s \psi_{bi}}{q} \frac{N_D}{N_A(N_A + N_D)}} \quad , \tag{18a}$$

$$W_{Dn} = \sqrt{\frac{2\varepsilon_s \psi_{bi}}{q} \frac{N_A}{N_D(N_A + N_D)}} , \tag{18b}$$

$$W_{Dp} + W_{Dn} = \sqrt{\frac{2\varepsilon_s}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\psi_{bi}} . \tag{19}$$

The following relationships can be further deduced:

$$\frac{|\psi_n|}{\psi_{bi}} = \frac{W_{Dn}}{W_{Dp} + W_{Dn}} = \frac{N_A}{N_A + N_D} , \tag{20a}$$

$$\frac{\psi_p}{\psi_{bi}} = \frac{W_{Dp}}{W_{Dp} + W_{Dn}} = \frac{N_D}{N_A + N_D} . \tag{20b}$$

For a one-sided abrupt junction (p^+-n or n^+-p), Eq. 4 is used to calculate the built-in potential. In this case, the majority of the potential variation and depletion region will be inside the lightly doped side. Equation 19 reduces to

$$W_D = \sqrt{\frac{2\varepsilon_s \psi_{bi}}{qN}} \tag{21}$$

where N is N_D or N_A depending on whether $N_A \gg N_D$ or vice versa, and

$$\psi_i(x) = |\mathscr{E}_m|\left(x - \frac{x^2}{2W_D}\right). \tag{22}$$

This discussion uses box profiles for the depletion charges, i.e., depletion approximation. A more accurate result for the depletion-layer properties can be obtained by considering the majority-carrier contribution in addition to the impurity concentration in the Poisson equation, that is, $\rho \approx -q[N_A - p(x)]$ on the p-side and $\rho \approx q[N_D - n(x)]$ on the n-side. The depletion width is essentially the same as given by Eq. 19, except that ψ_{bi} is replaced by ($\psi_{bi} - 2kT/q$).* The correction factor $2kT/q$ comes about because of the two majority-carrier distribution tails[5,6] (electrons in n-side and holes in p-side, as shown by the dashed lines in Fig. 1a) near the edges of the depletion region. Each contributes a correction factor kT/q. The depletion-layer width at thermal equilibrium for a one-sided abrupt junction becomes

$$W_D = \sqrt{\frac{2\varepsilon_s}{qN}\left(\psi_{bi} - \frac{2kT}{q}\right)}. \tag{23}$$

Furthermore, when a voltage V is applied to the junction, the total electrostatic potential variation across the junction is given by ($\psi_{bi} - V$) where V is positive for forward bias (positive voltage on p-region with respect to n-region) and negative for reverse bias. Substituting ($\psi_{bi} - V$) for ψ_{bi} in Eq. 23 yields the depletion-layer width as a function of the applied voltage. The results for one-sided abrupt junctions in
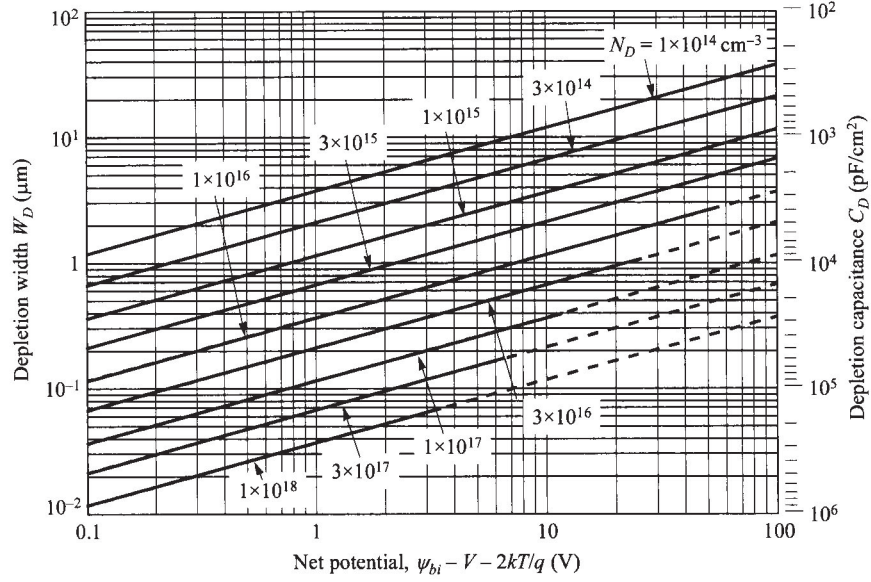
Fig. 2 Depletion-layer width and depletion-layer capacitance per unit area as a function of net potential ($\psi_{bi} - V - 2kT/q$) for one-sided abrupt junctions in Si. Doping N is from the lightly doped side. Dashed lines represent breakdown conditions.

silicon are shown in Fig. 2. The net potential at zero bias is near 0.8 V for Si and 1.3 V for GaAs. This net potential will be decreased under forward bias and increased under reverse bias. These results can also be used for GaAs since both Si and GaAs have approximately the same static dielectric constants. To obtain the depletion-layer width for other semiconductors such as Ge, one must multiply the results of Si by the factor $\sqrt{\varepsilon_s(\text{Ge})/\varepsilon_s(\text{Si})}$ (= 1.16). The simple model above can give adequate predictions for most abrupt *p-n* junctions.

---

* In the *p*-type region, the Poisson equation including the hole concentration is

$$\frac{d^2\psi_i}{dx^2} = \frac{q}{\varepsilon_s}[N_A - p(x)] = \frac{qN_A}{\varepsilon_s}[1 - \exp(-\beta_{th}\psi_i)].$$

Integrating both sides by $d\psi_i$, and using $d\psi_i/dx = -\mathscr{E}$,

$$\int_0^{\psi_p} -\frac{d\mathscr{E}}{dx} \, d\psi_i = \frac{qN_A}{\varepsilon_s}\int_0^{\psi_p}[1 - \exp(-\beta_{th}\psi_i)]d\psi_i,$$

$$\frac{\mathscr{E}_m^2}{2} = \frac{qN_A}{\beta_{th}\varepsilon_s}[\beta_{th}\psi_p + \exp(-\beta_{th}\psi_p) - 1] \approx \frac{qN_A}{\varepsilon_s}\left(\psi_p - \frac{kT}{q}\right).$$

Comparing this to Eq. 17, the potential is decreased by kT/q per side of the junction.

Depletion-Layer Capacitance. The depletion-layer capacitance per unit area is defined as $C_D = dQ_D/dV = \varepsilon_s/W_D$, where dQ_D is the incremental depletion charge on each side of the junction (total charge is zero) upon an incremental change of the applied voltage dV. For one-sided abrupt junctions, the capacitance per unit area is given by

$$C_D = \frac{\varepsilon_s}{W_D} = \sqrt{\frac{q\varepsilon_s N}{2}} \left(\psi_{bi} - V - \frac{2kT}{q} \right)^{-1/2} \tag{24}$$

where V is positive/negative for forward/reverse bias. The results of the depletion-layer capacitance are also shown in Fig. 2. Rearrange the above equation leads to:

$$\frac{1}{C_D^2} = \frac{2}{q\varepsilon_s N} \left(\psi_{bi} - V - \frac{2kT}{q} \right), \tag{25}$$

$$\frac{d(1/C_D^2)}{dV} = -\frac{2}{q\varepsilon_s N}. \tag{26}$$

It is clear from Eqs. 25 and 26 that by plotting $1/C^2$ versus V, a straight line should result from a one-sided abrupt junction (Fig. 3). The slope gives the impurity concentration of the substrate (N), and the extrapolation to $1/C^2 = 0$ gives ($\psi_{bi} - 2kT/q$). Note that, for the forward bias, a diffusion capacitance exists in addition to the depletion capacitance mentioned previously. The diffusion capacitance will be discussed in Section 2.3.4.

Note that the semiconductor potential and the capacitance-voltage data are insensitive to changes in the doping profiles that occur in a distance less than a Debye length.[7] The Debye length L_D is a characteristic length for semiconductors and is defined as

$$L_D \equiv \sqrt{\frac{\varepsilon_s kT}{q^2 N}} = \sqrt{\frac{\varepsilon_s}{qN\beta_{th}}}. \tag{27}$$

This Debye length gives an idea of the limit of the potential change in response to an abrupt change in the doping profile. Consider a case where the doping has a small
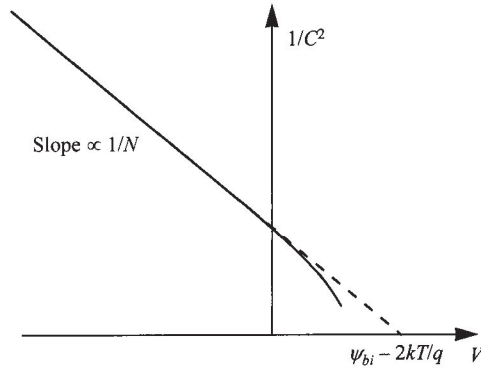

Fig. 3 A $1/C^2$-V plot can yield the built-in potential and doping density N.

increase of ΔN_D in the background of N_D, the change of potential $\Delta \psi_i(x)$ near the step is given by

$$n = N_D \exp\left(\frac{\Delta \psi_i q}{kT}\right), \tag{28}$$

$$\frac{d^2 \Delta \psi_i}{dx^2} = -\frac{q}{\varepsilon_s}(N_D + \Delta N_D - n) = -\frac{q N_D}{\varepsilon_s}\left[1 + \frac{\Delta N_D}{N_D} - \exp\left(\frac{\Delta \psi_i q}{kT}\right)\right]$$

$$\approx -\frac{q N_D}{\varepsilon_s}\left[1 + \frac{\Delta N_D}{N_D} - \left(1 + \frac{\Delta \psi_i q}{kT}\right)\right] \approx \frac{q^2 N_D}{\varepsilon_s kT}\Delta \psi_i \tag{29}$$

whose solution has a decay length given by Eq. 27. This implies that if the doping profile changes abruptly in a scale less than the Debye length, this variation has no effect and cannot be resolved, and that if the depletion width is smaller than the Debye length, the analysis using the Poisson equation is no longer valid. At thermal equilibrium the depletion-layer widths of abrupt junctions are about $8L_D$ for Si, and $10L_D$ for GaAs. The Debye length as a function of doping density is shown in Fig. 4 for silicon at room temperature. For a doping density of 10^{16} cm^{-3}, the Debye length is 40 nm; for other dopings, L_D will vary as $1/\sqrt{N}$, that is, a reduction by a factor of 3.16 per decade.

2.2.2 Linearly Graded Junction

In practical devices, the doping profiles are not abrupt, especially near the metallurgical junction where the two types meet and they compensate each other. When the depletion widths terminate within this transition region, the doping profile can be approximated by a linear function. Consider the thermal-equilibrium case first. The impurity distribution for a linearly graded junction is shown in Fig. 5a. The Poisson equation for this case is
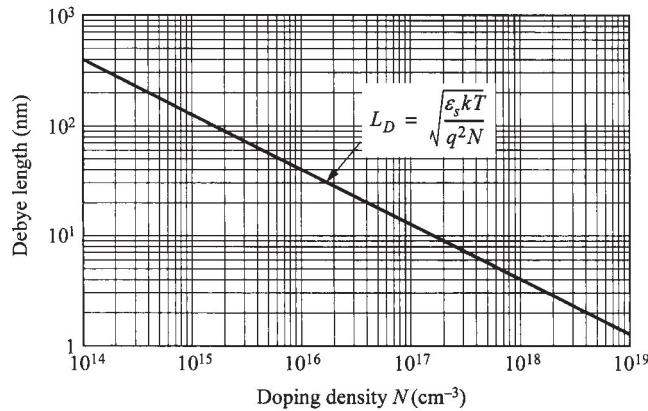

Fig. 4 Debye length in Si at room temperature as a function of doping density N.
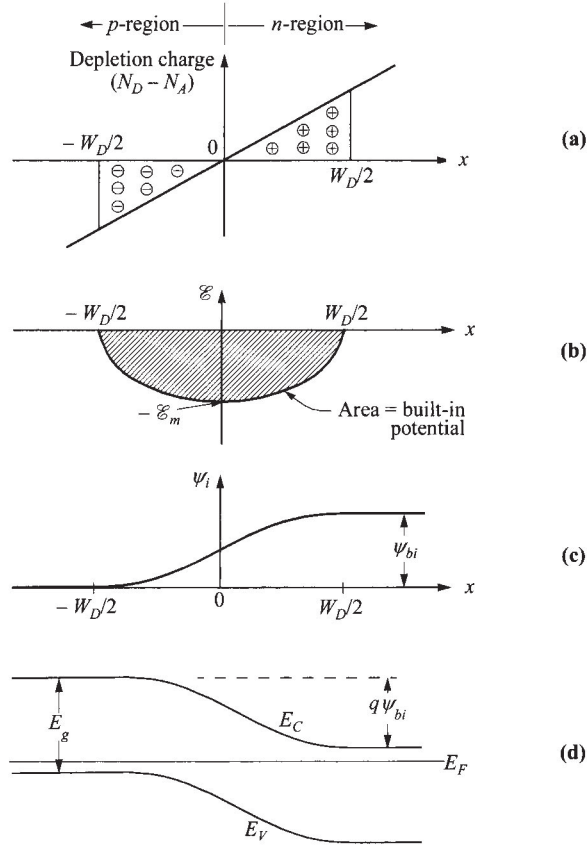
Fig. 5 Linearly graded junction in thermal equilibrium. (a) Space-charge distribution. (b) Electric-field distribution. (c) Potential distribution. (d) Energy-band diagram.

$$-\frac{d^2\psi_i}{dx^2} = \frac{d\mathscr{E}}{dx} = \frac{\rho(x)}{\varepsilon_s} = \frac{q}{\varepsilon_s}(p - n + ax)$$

$$\approx \frac{qax}{\varepsilon_s} \qquad\qquad -\frac{W_D}{2} \le x \le \frac{W_D}{2} \qquad (30)$$

where a is the doping gradient in cm^{-4}. By integrating Eq. 30 once, we obtain the field distribution shown in Fig. 5b:

$$\mathscr{E}(x) = -\frac{qa}{2\varepsilon_s}\left[\left(\frac{W_D}{2}\right)^2 - x^2\right] \qquad -\frac{W_D}{2} \le x \le \frac{W_D}{2} \qquad (31)$$

with the maximum field $\mathscr{E}_m$ at $x = 0$,

$$|\mathscr{E}_m| = \frac{qaW_D^2}{8\varepsilon_s}. \qquad (32)$$

Integrating Eq. 30 once again gives the potential distribution shown in Fig. 5c

$$\psi_i(x) = \frac{qa}{6\varepsilon_s}\Big[2\Big(\frac{W_D}{2}\Big)^3 + 3\Big(\frac{W_D}{2}\Big)^2 x - x^3\Big] \qquad -\frac{W_D}{2} \le x \le \frac{W_D}{2} \qquad (33)$$

from which the built-in potential can be related to the depletion width

$$\psi_{bi} = \frac{qaW_D^3}{12\varepsilon_s} \qquad (34)$$

or

$$W_D = \Big(\frac{12\varepsilon_s\psi_{bi}}{qa}\Big)^{1/3}. \qquad (35)$$

Since the values of the impurity concentrations at the edges of the depletion region ($-W_D/2$ and $W_D/2$) are the same and equal to $aW_D/2$, the built-in potential for a linearly graded junction can be approximated by an expression similar to Eq. 5:

$$\psi_{bi} \approx \frac{kT}{q}\ln\Big[\frac{(aW_D/2)(aW_D/2)}{n_i^2}\Big]$$

$$\approx \frac{2kT}{q}\ln\Big(\frac{aW_D}{2n_i}\Big). \qquad (36)$$

Equations 35 and 36 can thus be used to solve for W_D and ψ_{bi}.

Based on an accurate numerical technique,[8] the built-in potential can be calculated explicitly by an expression as a *gradient voltage* V_g:

$$V_g = \frac{2kT}{3q}\ln\Big(\frac{a^2\varepsilon_s kT}{8n_i^3 q^2}\Big). \qquad (37)$$

The gradient voltages for Si and GaAs as a function of impurity gradient are shown in Fig. 6. These voltages are smaller than the ψ_{bi} calculated from Eq. 36, using the depletion approximation, by more than 100 mV. The depletion-layer width and the corresponding capacitance for silicon using this V_g as the built-in potential are plotted in Fig. 7 as a function of net potential ($V_g - V$).

The depletion-layer capacitance for a linearly graded junction is given by

$$C_D = \frac{\varepsilon_s}{W_D} = \Big[\frac{qa\varepsilon_s^2}{12(\psi_{bi}-V)}\Big]^{1/3} \qquad (38)$$

where V is positive/negative for forward/reverse bias.

2.2.3 Arbitrary Doping Profile

In this section we consider the doping near the junction to be of any arbitrary shape. Limiting the discussion to the *n*-side of a p^+-*n* junction, the net potential change at the junction is given by integrating the total field across the depletion region;
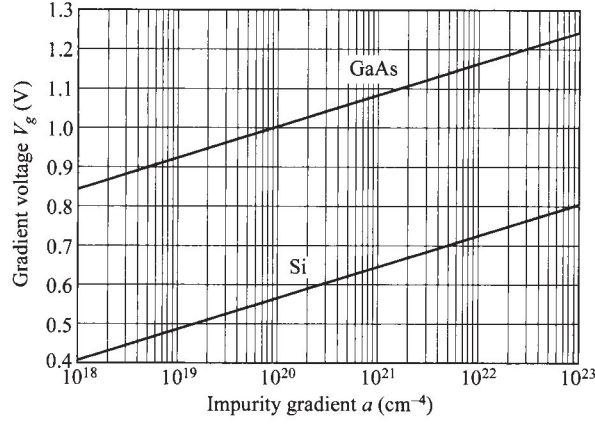
Fig. 6 Gradient voltages for linearly graded junctions in Si and GaAs.

$$\psi_n = \psi_{n0} - V = -\int_0^{W_D} \mathscr{E}(x)dx = -x\mathscr{E}(x)\Big|_0^{W_D} + \int_{\mathscr{E}(0)}^{\mathscr{E}(W_D)} xd\mathscr{E}, \tag{39}$$

where ψ_{n0} is ψ_n at zero bias. The first term becomes zero since the field at the depletion edge $\mathscr{E}(W_D)$ is zero. The interface potential becomes
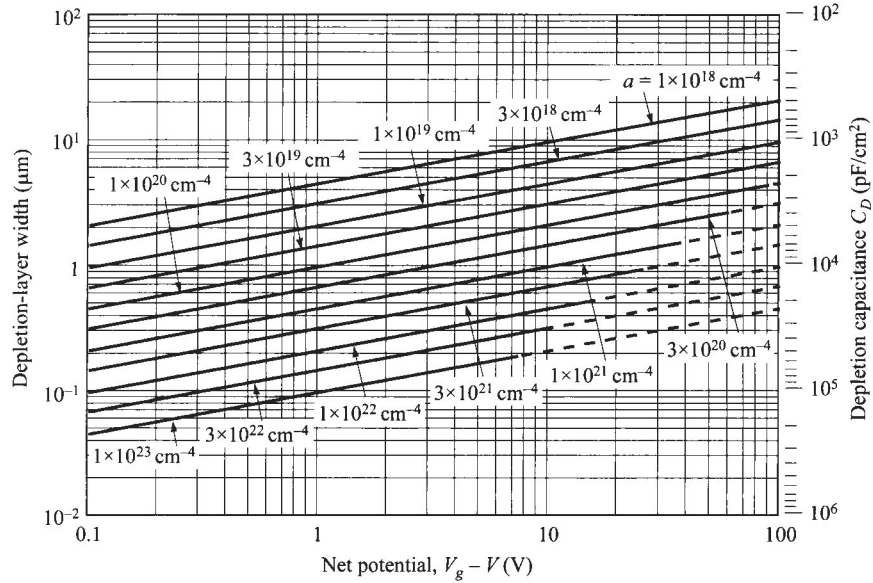

Fig. 7 Depletion-layer width and depletion-layer capacitance per unit area as a function of net potential $(V_g - V)$ for different impurity gradients in linearly graded junctions in Si. Dashed lines represent breakdown conditions.

$$\psi_n = \int_{\mathscr{E}(0)}^{\mathscr{E}(W_D)} x \frac{d\mathscr{E}}{dx} dx = \frac{q}{\varepsilon_s} \int_0^{W_D} x N_D(x) dx. \tag{40}$$

Meanwhile, the total depletion-layer charge is given by

$$Q_D = q \int_0^{W_D} N_D(x) dx. \tag{41}$$

Differentiating the above quantities with respect to the depletion width gives

$$\frac{dV}{dW_D} = -\frac{d\psi_n}{dW_D} = -\frac{q N_D(W_D) W_D}{\varepsilon_s}, \tag{42}$$

$$\frac{dQ_D}{dW_D} = q N_D(W_D). \tag{43}$$

From these we obtain the depletion-layer capacitance,

$$C_D = \left| \frac{dQ_D}{dV} \right| = \left| \frac{dQ_D}{dW_D} \times \frac{dW_D}{dV} \right| = \frac{\varepsilon_s}{W_D}. \tag{44}$$

Again the general expression of ε_s/W_D is obtained and is applicable to any arbitrary doping profile. From this we can derive Eq. 26 for a general nonuniform profile;

$$\frac{d(1/C_D^2)}{dV} = \frac{d(1/C_D^2)}{dW_D} \frac{dW_D}{dV} = \frac{2W_D}{\varepsilon_s^2} \frac{dW_D}{dV}$$

$$= -\frac{2}{q\varepsilon_s N_D(W_D)}. \tag{45}$$

This C-V technique can be used to measure nonuniform doping profile. The $1/C_D^2$-V plot (like that shown in Fig. 3) would deviate from a straight line if the doping is not constant.

2.3   CURRENT-VOLTAGE CHARACTERISTICS

2.3.1   Ideal Case—Shockley Equation[1,2]

The ideal current-voltage characteristics are based on the following four assumptions: (1) the abrupt depletion-layer approximation; that is, the built-in potential and applied voltages are supported by a dipole layer with abrupt boundaries, and outside the boundaries the semiconductor is assumed to be neutral; (2) the Boltzmann approximation, similar to Eqs. 21 and 23 of Chapter 1, is valid; (3) the low-injection assumption; that is, the injected minority carrier densities are small compared with the majority-carrier densities; and (4) no generation-recombination current exists inside the depletion layer, and the electron and hole currents are constant throughout the depletion layer.

We first consider the Boltzmann relation. At thermal equilibrium this relation is given by

$$n = n_i \exp\left(\frac{E_F - E_i}{kT}\right), \tag{46a}$$

$$p = n_i \exp\left(\frac{E_i - E_F}{kT}\right). \tag{46b}$$

Obviously, at thermal equilibrium, the pn product from the above equations is equal to n_i^2. When voltage is applied, the minority-carrier densities on both sides of the junction are changed, and the pn product is no longer equal to n_i^2. We shall now define the quasi-Fermi (imref) levels as follows:

$$n \equiv n_i \exp\left(\frac{E_{Fn} - E_i}{kT}\right), \tag{47a}$$

$$p \equiv n_i \exp\left(\frac{E_i - E_{Fp}}{kT}\right), \tag{47b}$$

where E_{Fn} and E_{Fp} are the quasi-Fermi levels for electrons and holes, respectively. From Eqs. 47a and 47b we obtain

$$E_{Fn} \equiv E_i + kT\ln\left(\frac{n}{n_i}\right), \tag{48a}$$

$$E_{Fp} \equiv E_i - kT\ln\left(\frac{p}{n_i}\right). \tag{48b}$$

The pn product becomes

$$pn = n_i^2 \exp\left(\frac{E_{Fn} - E_{Fp}}{kT}\right). \tag{49}$$

For a forward bias, $(E_{Fn} - E_{Fp}) > 0$ and $pn > n_i^2$; on the other hand, for a reversed bias, $(E_{Fn} - E_{Fp}) < 0$ and $pn < n_i^2$.

From Eq. 156a of Chapter 1, Eq. 47a, and the fact that $\mathscr{E} \equiv \nabla E_i/q$, we obtain

$$J_n = q\mu_n\left(n\mathscr{E} + \frac{kT}{q}\nabla n\right) = \mu_n n \nabla E_i + \mu_n kT\left[\frac{n}{kT}(\nabla E_{Fn} - \nabla E_i)\right]$$

$$= \mu_n n \nabla E_{Fn}. \tag{50}$$

Similarly, we obtain,

$$J_p = \mu_p p \nabla E_{Fp}. \tag{51}$$

Thus, the electron and hole current densities are proportional to the gradients of the electron and hole quasi-Fermi levels, respectively. If $E_{Fn} = E_{Fp}$ = constant (at thermal equilibrium), then $J_n = J_p = 0$.

The idealized potential distributions and the carrier concentrations in a p-n junction under forward-bias and reverse-bias conditions are shown in Fig. 8. The variations of E_{Fn} and E_{Fp} with distance are related to the carrier concentrations as given in
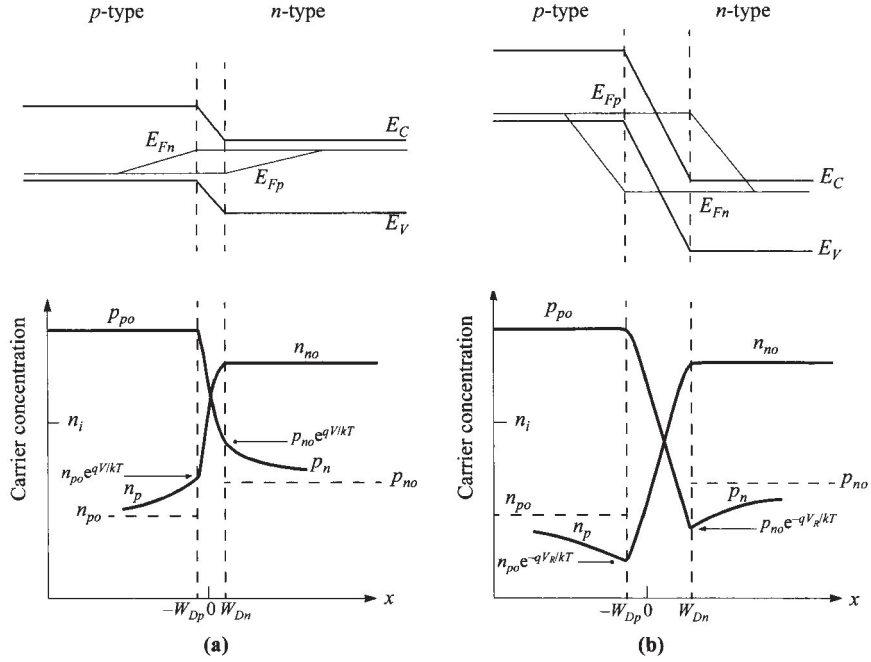
Fig. 8 Energy-band diagram, with quasi-Fermi levels for electrons and holes, and carrier distributions under (a) forward bias and (b) reverse bias.

Eqs. 48a and 48b, and to the current as given by Eqs. 50 and 51. Inside the depletion region, E_{Fn} and E_{Fp} remain relatively constant. This comes about because the carrier concentrations are relatively much higher inside the depletion region, but since the currents remain fairly constant, the gradients of the quasi-Fermi levels have to be small. In addition, the depletion width is typically much shorter than the diffusion length, so the total drop of quasi-Fermi levels inside the depletion width is not significant. With these arguments, it follows that within the depletion region,

$$qV = E_{Fn} - E_{Fp}. \tag{52}$$

Equations 49 and 52 can be combined to give the electron density at the boundary of the depletion-layer region on the *p*-side ($x = -W_{Dp}$):

$$n_p(-W_{Dp}) = \frac{n_i^2}{p_p} \exp\left(\frac{qV}{kT}\right) \approx n_{po} \exp\left(\frac{qV}{kT}\right) \tag{53a}$$

where $p_p \approx p_{po}$ for low-level injection, and n_{po} is the equilibrium electron density on the *p*-side. Similarly,

$$p_n(W_{Dn}) = p_{no} \exp\left(\frac{qV}{kT}\right) \tag{53b}$$

at $x = W_{Dn}$ for the n-type boundary. The preceding equations are the most-important boundary conditions for the ideal current-voltage equation.

From the continuity equations we obtain for the steady-state condition in the n-side of the junction:

$$-U + \mu_n \mathscr{E} \frac{dn_n}{dx} + \mu_n n_n \frac{d\mathscr{E}}{dx} + D_n \frac{d^2 n_n}{dx^2} = 0, \tag{54a}$$

$$-U - \mu_p \mathscr{E} \frac{dp_n}{dx} - \mu_p p_n \frac{d\mathscr{E}}{dx} + D_p \frac{d^2 p_n}{dx^2} = 0. \tag{54b}$$

In these equations, U is the net recombination rate. Note that due to charge neutrality, majority carriers need to adjust their concentrations such that $(n_n - n_{no}) = (p_n - p_{no})$. It also follows that $dn_n/dx = dp_n/dx$. Multiplying Eq. 54a by $\mu_p p_n$ and Eq. 54b by $\mu_n n_n$, and combining with the Einstein relation $D = (kT/q)\mu$, we obtain

$$-\frac{p_n - p_{no}}{\tau_p} - \frac{n_n - p_n}{(n_n/\mu_p) + (p_n/\mu_n)} \frac{\mathscr{E} dp_n}{dx} + D_a \frac{d^2 p_n}{dx^2} = 0 \tag{55}$$

where

$$D_a = \frac{n_n + p_n}{n_n/D_p + p_n/D_n} \tag{56}$$

is the ambipolar diffusion coefficient, and

$$\tau_p \equiv \frac{p_n - p_{no}}{U}. \tag{57}$$

From the low-injection assumption [e.g., $p_n \ll (n_n \approx n_{no})$ in the n-type semiconductor], Eq. 55 reduces to

$$-\frac{p_n - p_{no}}{\tau_p} - \mu_p \mathscr{E} \frac{dp_n}{dx} + D_p \frac{d^2 p_n}{dx^2} = 0 \tag{58}$$

which is Eq. 54b except that the term $\mu_p p_n d\mathscr{E}/dx$ is ignored under the low-injection assumption.

In the neutral region where there is no electric field, Eq. 58 further reduces to

$$\frac{d^2 p_n}{dx^2} - \frac{p_n - p_{no}}{D_p \tau_p} = 0. \tag{59}$$

The solution of Eq. 59, with the boundary conditions of Eq. 53b and $p_n(x = \infty) = p_{no}$, gives

$$p_n(x) - p_{no} = p_{no} \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] \exp\left(-\frac{x - W_{Dn}}{L_p}\right) \tag{60}$$

where

$$L_p \equiv \sqrt{D_p \tau_p}. \tag{61}$$

At $x = W_{Dn}$, the hole diffusion current is

$$J_p = -qD_p\frac{dp_n}{dx}\bigg|_{W_{Dn}} = \frac{qD_p p_{no}}{L_p}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \tag{62a}$$

Similarly, we obtain the electron diffusion current in the *p*-side

$$J_n = qD_n\frac{dn_p}{dx}\bigg|_{-W_{Dp}} = \frac{qD_n n_{po}}{L_n}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \tag{62b}$$

The minority-carrier densities and the current densities for the forward-bias and reverse-bias conditions are shown in Fig. 9. It is interesting to note that the hole current is due to injection of holes from the *p*-side to the *n*-side, but the magnitude is determined by the properties in the *n*-side only (D_p, L_p, p_{no}). The analogy holds for the electron current.

The total current is given by the sum of Eqs. 62a and 62b:

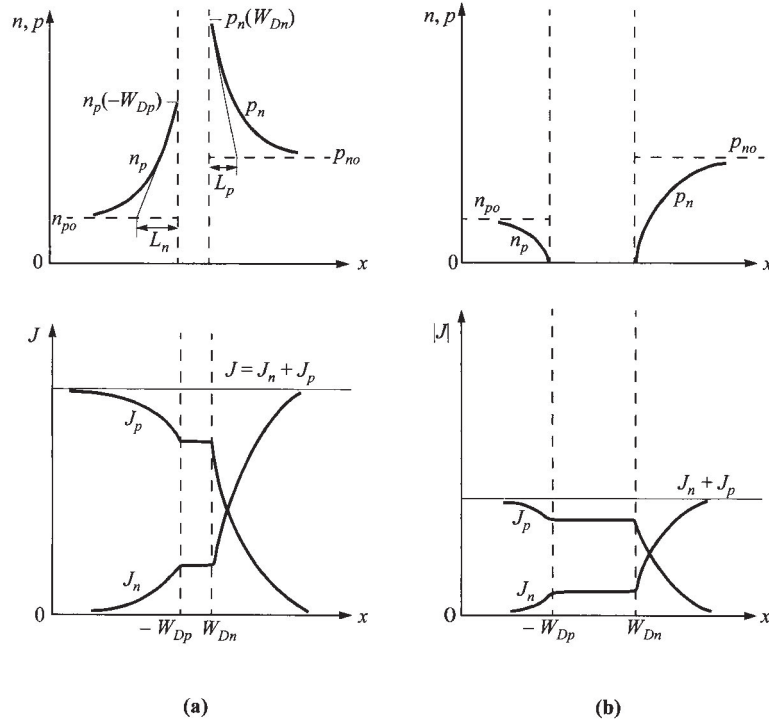$$J = J_p + J_n = J_0\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \tag{63}$$

Fig. 9 Carrier distributions and current densities (both linear plots) for (a) forward-biased conditions and (b) reverse-biased conditions.

$$J_0 \equiv \frac{qD_p p_{no}}{L_p} + \frac{qD_n n_{po}}{L_n} \equiv \frac{qD_p n_i^2}{L_p N_D} + \frac{qD_n n_i^2}{L_n N_A} \; . \tag{64}$$

Equation 63 is the celebrated Shockley equation,[1,2] which is the ideal diode law. The ideal current-voltage relation is shown in Figs. 10a and b in the linear and semilog plots, respectively. In the forward direction (positive bias on the p-side) for $V > 3kT/q$, the rate of current rise is constant (Fig. 10b); at 300 K for every decade change of current, the voltage changes by 59.5 mV ($= 2.3kT/q$). In the reverse direction, the current density saturates at $-J_0$.

We shall now briefly consider the temperature effect on the saturation current density J_0. We shall consider only the first term in Eq. 64, since the second term will behave similarly to the first one. For the one-sided p^+-n abrupt junction (with donor concentration N_D), $p_{no} \gg n_{po}$, the second term can also be neglected. The quantities n_i, D_p, p_{no}, and L_p ($\equiv \sqrt{D_p \tau_p}$) are all temperature-dependent. If D_p/τ_p is proportional to T^γ, where γ is a constant, then

$$J_0 \approx \frac{qD_p p_{no}}{L_p} \approx q \sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} \propto T^{\gamma/2}\left[T^3 \exp\left(-\frac{E_g}{kT}\right)\right]$$

$$\propto T^{(3+\gamma/2)} \exp\left(-\frac{E_g}{kT}\right) \; . \tag{65}$$

The temperature dependence of the term $T^{(3+\gamma/2)}$ is not important compared with the exponential term. The slope of a plot J_0 versus $1/T$ is determined mainly by the energy gap E_g. It is expected that in the reverse direction, where $|J_R| \approx J_0$, the current will increase approximately as $\exp(-E_g/kT)$ with temperature; and in the forward direction, where $J_F \approx J_0\exp(qV/kT)$, the current will increase approximately as $\exp[-(E_g - qV)/kT]$.
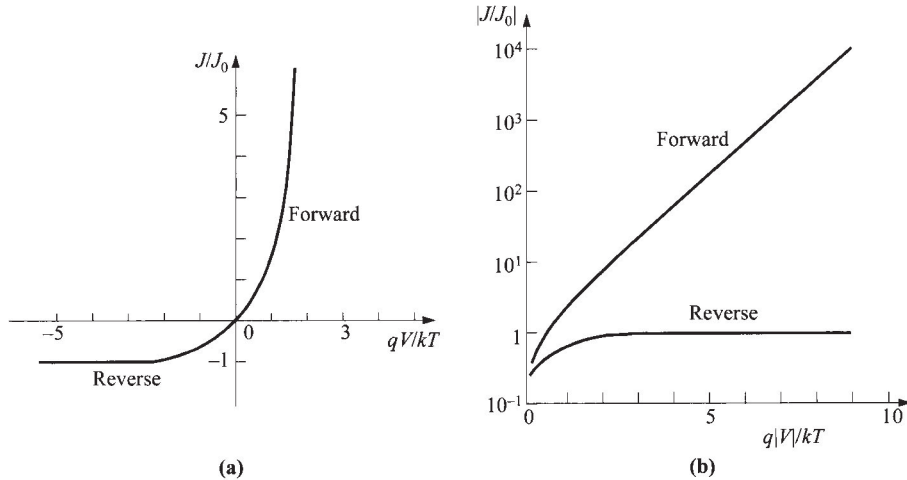

Fig. 10 Ideal current-voltage characteristics. (a) Linear plot. (b) Semilog plot.

The Shockley equation adequately predicts the current-voltage characteristics of germanium *p-n* junctions at low current densities. For Si and GaAs *p-n* junctions, however, the ideal equation can only give qualitative agreement. The departures from the ideal are mainly due to: (1) the generation and recombination of carriers in the depletion layer, (2) the high-injection condition that may occur even at relatively small forward bias, (3) the parasitic *IR* drop due to series resistance, (4) the tunneling of carriers between states in the bandgap, and (5) the surface effects. In addition, under sufficiently larger field in the reverse direction, the junction will breakdown as a result, for example, of avalanche multiplication. The junction breakdown will be discussed in Section 2.4.

The surface effects on *p-n* junctions are primarily due to ionic charges on or outside the semiconductor surface that induce image charges in the semiconductor, and thereby cause the formation of the so-called surface channels or surface depletion-layer regions. Once a channel is formed, it modifies the junction depletion region and gives rise to surface leakage current. For Si planar *p-n* junctions, the surface leakage current is generally much smaller than the generation-recombination current in the depletion region.

2.3.2 Generation-Recombination Process[3]

Consider first the generation current under the reverse-bias condition. Because of the reduction in carrier concentration under reverse bias ($pn \ll n_i^2$), the dominant generation processes, as discussed in Section 1.5.4, are those of emission. The rate of generation of electron-hole pairs can be obtained from Eq. 92 of Chapter 1 with the condition $p \ll n_i$ and $n \ll n_i$:

$$U = -\left\{ \frac{\sigma_p \sigma_n v_{th} N_t}{\sigma_n \exp[(E_t - E_i)/kT] + \sigma_p \exp[(E_i - E_t)/kT]} \right\} n_i \equiv -\frac{n_i}{\tau_g} \tag{66}$$

where τ_g is the generation lifetime and is defined as the reciprocal of the expression in brackets (see Eq. 98 of Chapter 1 and the discussion following). The current due to generation in the depletion region is thus given by

$$J_{ge} = \int_0^{W_D} q|U|dx \approx q|U|W_D \approx \frac{qn_i W_D}{\tau_g} \tag{67}$$

where W_D is the depletion-layer width. If the generation lifetime is a slowly varying function of temperature, the generation current will then have the same temperature dependence as n_i. At a given temperature, J_{ge} is proportional to the depletion-layer width, which in turn is dependent on the applied reverse bias. It is thus expected that

$$J_{ge} \propto (\psi_{bi} + V)^{1/2} \tag{68}$$

for abrupt junctions, and

$$J_{ge} \propto (\psi_{bi} + V)^{1/3} \tag{69}$$

for linearly graded junctions.

The total reverse current (for $p_{no} \gg n_{po}$ and $|V| > 3kT/q$) can be approximated by the sum of the diffusion component in the neutral region and the generation current in the depletion region:

$$J_R = q \sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} + \frac{qn_i W_D}{\tau_g}. \tag{70}$$

For semiconductors with large values of n_i (such as Ge), the diffusion component will dominate at room temperature and the reverse current will follow the Shockley equation; but if n_i is small (such as for Si), the generation current may dominate. A typical result for Si is shown in Fig. 11, curve (e). At sufficiently high temperatures, however, the diffusion current will dominate.

At forward bias, where the major recombination-generation processes in the depletion region are the capture processes, we have a recombination current J_{re} in addition to the diffusion current. Substituting Eq. 49 in Eq. 92 of Chapter 1 yields

$$U = \frac{\sigma_p \sigma_n v_{th} N_t n_i^2 [\exp(qV/kT) - 1]}{\sigma_n \{n + n_i \exp[(E_t - E_i)/kT]\} + \sigma_p \{p + n_i \exp[(E_i - E_t)/kT]\}}. \tag{71}$$
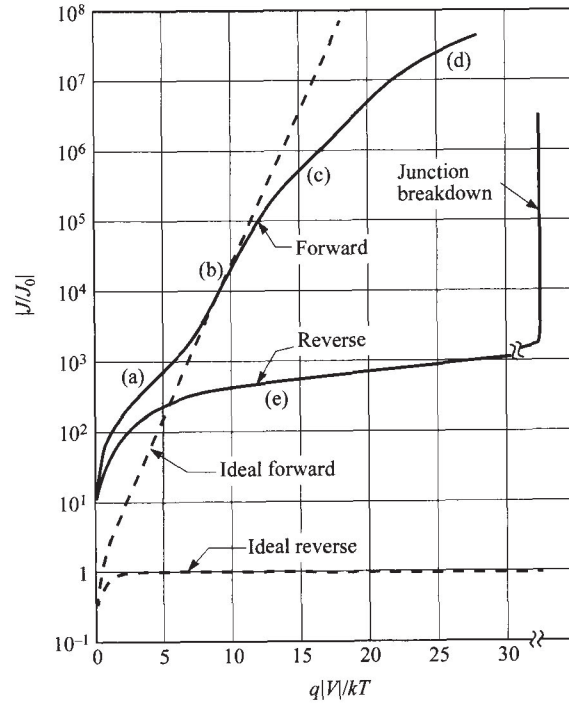

Fig. 11 Current-voltage characteristics of a practical Si diode. (a) Generation-recombination current region. (b) Diffusion-current region. (c) High-injection region. (d) Series-resistance effect. (e) Reverse leakage current due to generation-recombination and surface effects.

Under the assumptions that $E_t = E_i$ and $\sigma_n = \sigma_p = \sigma$, Eq. 71 reduces to

$$U = \frac{\sigma v_{th} N_t n_i^2 [\exp(qV/kT) - 1]}{n + p + 2n_i}$$

$$= \frac{\sigma v_{th} N_t n_i^2 [\exp(qV/kT) - 1]}{n_i \{\exp[(E_{Fn} - E_i)/kT] + \exp[(E_i - E_{Fp})/kT] + 2\}}. \tag{72}$$

The maximum value of U exists in the depletion region where E_i is halfway between E_{Fn} and E_{Fp}, and so the denominator of Eq. 72 becomes $2n_i[\exp(qV/2kT) + 1]$. We obtain for $V > kT/q$,

$$U \approx \frac{1}{2} \sigma v_{th} N_t n_i \exp\left(\frac{qV}{2kT}\right) \tag{73}$$

and

$$J_{re} = \int_0^{W_D} qU dx \approx \frac{qW_D}{2} \sigma v_{th} N_t n_i \exp\left(\frac{qV}{2kT}\right) \approx \frac{qW_D n_i}{2\tau} \exp\left(\frac{qV}{2kT}\right). \tag{74}$$

The above approximation assumes that most part of the depletion layer has this maximum recombination rate, and J_{re} is thus somewhat an overestimate. A more rigorous derivation gives[9]

$$J_{re} = \int_0^{W_D} qU dx = \sqrt{\frac{\pi}{2}} \frac{kTn_i}{\tau \mathscr{E}_0} \exp\left(\frac{qV}{2kT}\right) \tag{75}$$

where $\mathscr{E}_0$ is the electric field at the location of maximum recombination, and it is equal to

$$\mathscr{E}_0 = \sqrt{\frac{qN(2\psi_B - V)}{\varepsilon_s}}. \tag{76}$$

Similar to the generation current in reverse bias, the recombination current in forward bias is also proportional to n_i. The total forward current can be approximated by the sum of Eqs. 63 and 75. For a p^+-n junction ($p_{no} \gg n_{po}$) and $V \gg kT/q$:

$$J_F = q\sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} \exp\left(\frac{qV}{kT}\right) + \sqrt{\frac{\pi}{2}} \frac{kTn_i}{\tau_p \mathscr{E}_0} \exp\left(\frac{qV}{2kT}\right). \tag{77}$$

The experimental results in general can be represented by the empirical form,

$$J_F \propto \exp\left(\frac{qV}{\eta kT}\right) \tag{78}$$

where the ideality factor η equals 2 when the recombination current dominates [Fig. 11, curve (a)] and η equals 1 when the diffusion current dominates [Fig. 11, curve (b)]. When both currents are comparable, η has a value between 1 and 2.

2.3.3 High-Injection Condition

At high current densities (under the forward-bias condition) such that the injected minority-carrier density is comparable to the majority concentration, both drift and diffusion current components must be considered. The individual conduction current densities can always be given by Eqs. 50 and 51. Since J_p, q, μ_p, and p are positive, the quasi-Fermi level for holes E_{Fp} increases monotonically to the right as shown in Fig. 8a. Similarly, the quasi-Fermi level for electrons E_{Fn} decreases monotonically to the left. Thus, everywhere the separation of the two quasi-Fermi levels must be equal to or less than the applied voltage, and therefore[10]

$$pn \leq n_i^2 \exp\left(\frac{qV}{kT}\right) \tag{79}$$

even under the high-injection condition. Note also that the foregoing argument does not depend on recombination in the depletion region.

To illustrate the high-injection case, we present in Fig. 12 plots of numerical simulation results for carrier concentrations and energy-band diagram with quasi-Fermi levels for a silicon p^+-n step junction. The current densities in Figs. 12a, b, and c are 10, 10^3, and 10^4 A/cm^2, respectively. At 10 A/cm^2 the diode is in the low-injection regime. Almost all of the potential drop occurs across the junction. The hole concentration in the n-side is small compared to the electron concentration. At 10^3 A/cm^2 the electron concentration near the junction exceeds the donor concentration appreciably (bear in mind that from charge neutrality, injected carriers $\Delta p = \Delta n$). An ohmic potential drop appears on the n-side. At 10^4 A/cm^2 we have very high injection; the poten-
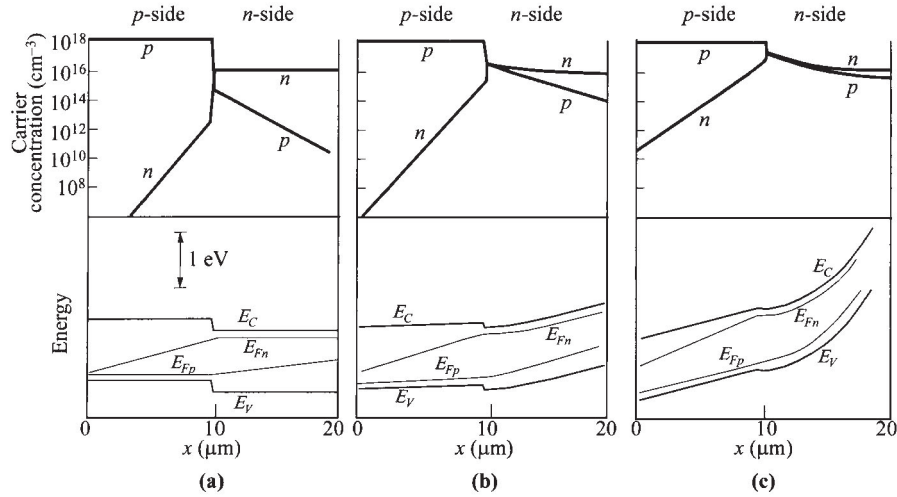

Fig. 12 Carrier concentrations and energy-band diagrams for a Si p^+-n junction operated at different current densities. (a) 10 A/cm^2. (b) 10^3 A/cm^2. (c) 10^4 A/cm^2. Device parameters: $N_A = 10^{18}$ cm^{-3}, $N_D = 10^{16}$ cm^{-3}, $\tau_n = 3\times10^{-10}$ s, and $\tau_p = 8.4\times10^{-10}$ s. (After Ref. 10.)

tial drop across the junction is insignificant compared to ohmic drops on both sides of the neutral regions. Even though only the center region of the diode is shown in Fig. 12, it is apparent that the separation of the quasi-Fermi levels is equal to or less than the applied voltage (qV).

From Fig. 12b and c, the carrier densities at the n-side of the junction are comparable $(n = p)$. Substituting this condition in Eq. 79, we obtain $p_n(x = W_{Dn}) \approx n_i \exp(qV/2kT)$. The current then becomes roughly proportional to $\exp(qV/2kT)$, as shown in Fig. 11, curve (c).

At high-current levels we should consider another effect associated with the finite resistivity in the quasi-neutral regions. This resistance absorbs an appreciable amount of the applied voltage between the diode terminals. This is shown in Fig. 11 as curve-(d). One can estimate the series resistance from comparing the experimental curve to the ideal curve $(\Delta V = IR)$. The series resistance effect can be substantially reduced by the use of epitaxial materials $(p^+\text{-}n\text{-}n^+)$.

2.3.4 Diffusion Capacitance

The depletion-layer capacitance considered previously accounts for most of the junction capacitance when the junction is reverse-biased. When forward-biased, there is, in addition, a significant contribution to junction capacitance from the rearrangement of minority carrier density, the so-called diffusion capacitance. In other words, the latter is due to the injected charge, while the former to the depletion-layer charge.

When a small ac signal is applied to a junction that is forward-biased at a dc voltage V_0 and current density J_0, the total voltage and current are defined by

$$V(t) = V_0 + V_1 \exp(j\omega t), \tag{80}$$

$$J(t) = J_0 + J_1 \exp(j\omega t) \tag{81}$$

where V_1 and J_1 are the small-signal voltage and current density, respectively. The imaginary part of the admittance J_1/V_1 will give the diffusion conductance and diffusion capacitance:

$$Y \equiv \frac{J_1}{V_1} \equiv G_d + j\omega C_d . \tag{82}$$

The electron and hole densities at the depletion region boundaries can be obtained from Eqs. 53a and 53b by using $[V_0 + V_1\exp(j\omega t)]$ instead of V. We obtain for the n-side of the junction and $V_1 \ll V_0$,

$$p_n(W_{Dn}) = p_{no}\exp\left\{\frac{q[V_0 + V_1\exp(j\omega t)]}{kT}\right\}$$

$$\approx p_{no}\exp\left(\frac{qV_0}{kT}\right) + \frac{p_{no}qV_1}{kT}\exp\left(\frac{qV_0}{kT}\right)\exp(j\omega t) \approx p_{no}\exp\left(\frac{qV_0}{kT}\right) + \tilde{p}_n(t) \tag{83}$$

A similar expression can be obtained for the electron density in the p-side. The first term in Eq. 83 is the dc component, and the second term is the small-signal ac com-

ponent. Substituting $\tilde{p}_n$ into the continuity equation (Eq. 158b of Chapter 1 with $G_p = \mathscr{E} = d\mathscr{E}/dx = 0$) yields

$$j\omega\tilde{p}_n = -\frac{\tilde{p}_n}{\tau_p} + D_p\frac{d^2\tilde{p}_n}{dx^2} \tag{84}$$

or

$$\frac{d^2\tilde{p}_n}{dx^2} - \frac{\tilde{p}_n}{D_p\tau_p/(1+j\omega\tau_p)} = 0. \tag{85}$$

Equation 85 is identical to Eq. 59 if the carrier lifetime is expressed as

$$\tau_p^* = \frac{\tau_p}{1+j\omega\tau_p}. \tag{86}$$

We can then obtain the alternating current density from Eq. 63 by making the appropriate substitutions:

$$J = \left(qp_{no}\sqrt{\frac{D_p}{\tau_p^*}} + qn_{po}\sqrt{\frac{D_n}{\tau_n^*}}\right)\exp\left\{\frac{q[V_0 + V_1\exp(j\omega t)]}{kT}\right\}$$

$$\approx \left(qp_{no}\sqrt{\frac{D_p}{\tau_p^*}} + qn_{po}\sqrt{\frac{D_n}{\tau_n^*}}\right)\left[\exp\left(\frac{qV_0}{kT}\right)\right]\left[1 + \frac{qV_1}{kT}\exp(j\omega t)\right], \tag{87}$$

with the ac component being

$$J_1 = \left(\frac{qD_pp_{no}\sqrt{1+j\omega\tau_p}}{L_p} + \frac{qD_nn_{po}\sqrt{1+j\omega\tau_n}}{L_n}\right)\left[\exp\left(\frac{qV_0}{kT}\right)\right]\frac{qV_1}{kT}. \tag{88}$$

From J_1/V_1, both G_d and C_d can be found and they are frequency dependent.

For relatively low frequencies ($\omega\tau_p$, $\omega\tau_n \ll 1$), the diffusion conductance G_{d0} is given by

$$G_{d0} = \frac{q}{kT}\left(\frac{qD_pp_{no}}{L_p} + \frac{qD_nn_{po}}{L_n}\right)\exp\left(\frac{qV_0}{kT}\right) \qquad \text{mho/cm}^2 \tag{89}$$

which has exactly the same value obtained by differentiating Eq. 63. The low-frequency diffusion capacitance C_{d0} can be obtained by using the approximation $\sqrt{1+j\omega\tau} \approx (1 + 0.5j\omega\tau)$

$$C_{d0} = \frac{q^2}{2kT}(L_pp_{no} + L_nn_{po})\exp\left(\frac{qV_0}{kT}\right) \qquad \text{F/cm}^2. \tag{90}$$

This diffusion capacitance is proportional to the forward current. For an n^+-p one-sided junction, it can shown that

$$C_{d0} = \frac{qL_n^2}{2kTD_n}J_F. \tag{91}$$

The frequency dependence of the diffusion conductance and capacitance is shown in Fig. 13 as a function of the normalized frequency $\omega\tau$ where only one term in Eq. 88
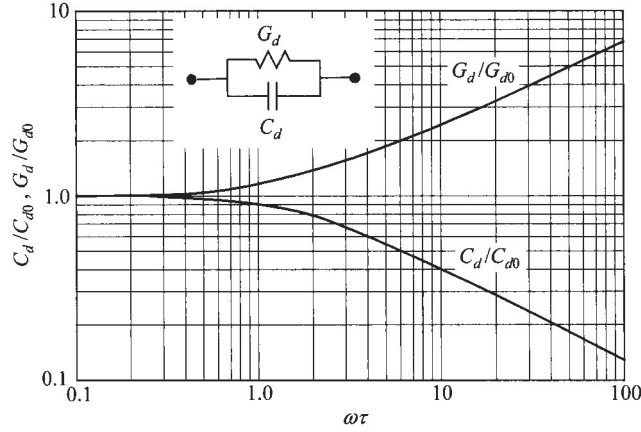
Fig. 13 Normalized diffusion conductance and diffusion capacitance versus $\omega\tau$. Inset shows the equivalent circuit of a *p-n* junction under forward bias.

is considered (e.g., the term contains p_{no} if $p_{no} \gg n_{po}$). The inset shows the equivalent circuit of the ac admittance. It is clear from Fig. 13 that the diffusion capacitance decreases with increasing frequency. For high frequencies, C_d is approximately proportional to $\omega^{-1/2}$. The diffusion capacitance is also proportional to the dc current level [$\propto \exp(qV_0/kT)$]. For this reason, C_d is especially important at low frequencies and under forward-bias conditions.

2.4 JUNCTION BREAKDOWN

When a sufficiently high field is applied to a *p-n* junction, the junction *breaks down* and conducts a very large current.[11] Breakdown occurs only in the reverse-bias regime because high voltage can be applied resulting in high field. There are basically three breakdown mechanisms: (1) thermal instability, (2) tunneling, and (3) avalanche multiplication. We consider the first two mechanisms briefly, and discuss avalanche multiplication in more detail.

2.4.1 Thermal Instability

Breakdown due to thermal instability is responsible for the maximum dielectric strength in most insulators at room temperature, and is also a major effect in semiconductors with relatively small bandgaps (e.g., Ge). Because of the heat dissipation caused by the reverse current at high reverse voltage, the junction temperature increases. This temperature increase, in turn, increases the reverse current in comparison with its value at lower voltages. This positive feedback is responsible for breakdown. The temperature effect on the reverse current-voltage characteristics is explained in Fig. 14. In this figure the reverse currents J_0 are represented by a family
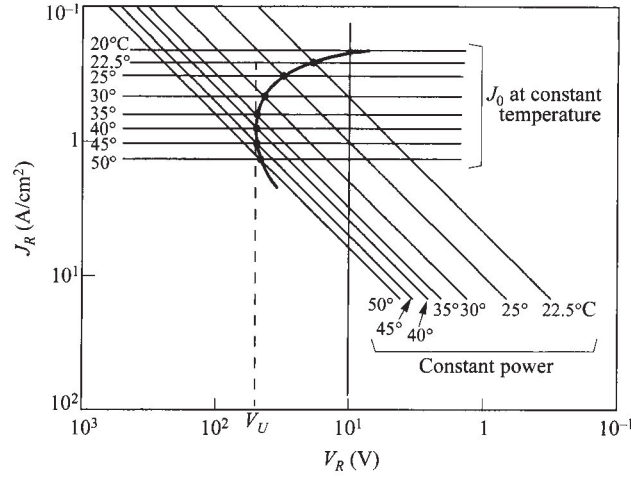
Fig. 14   Reverse current-voltage characteristics of thermal breakdown, where V_U is the turn-over voltage. (Note decreasing values of coordinates.) (After Ref. 12.)

of horizontal lines. Each line represents the current at a constant junction temperature, and the current varies as $T^{3+\gamma/2}\exp(-E_g/kT)$, as discussed previously. The heat dissipation hyperbolas which are proportional to the power, given by the I-V product, are shown as sloped straight lines in the log-log plot. These lines also have to satisfy the curves of constant junction temperature. So the reverse current-voltage characteristics are obtained by the intersection points of these two sets of curves. Because of the heat dissipation at high reverse voltage, the characteristics show a negative differential resistance. In this condition, the diode will be destroyed unless some special measure such as a large series-limiting resistor is used. This effect is called thermal instability or thermal runaway. The voltage V_U is called the turnover voltage. For p-n junctions with relatively large saturation currents (e.g., in Ge), the thermal instability is important at room temperature, but at very low temperatures it becomes less important compared with other mechanisms.

2.4.2 Tunneling

We next consider the tunneling effect (see Section 1.5.7) when the junction is under a large reverse bias. It is well known that carriers can tunnel through a potential barrier if this barrier is sufficiently thin, induced by a large field as shown in Fig. 15a. In this particular case, the barrier has a triangular shape with the maximum height given by the energy gap. The derivation of the tunneling current of a p-n junction (tunnel diode) is considered in details in Chapter 8, and the result is given here as:

$$J_t = \frac{\sqrt{2m^*}q^3 \mathscr{E} V_R}{4\pi^2\hbar^2\sqrt{E_g}}\exp\left(-\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\mathscr{E}\hbar}\right). \tag{92}$$

Since the field is not constant, $\mathscr{E}$ is some average field inside the junction.
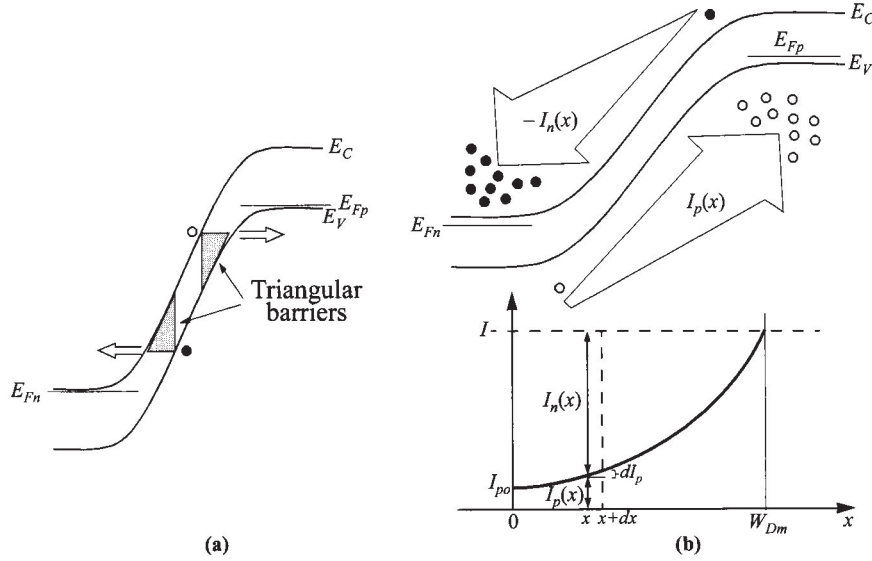
Fig. 15   Energy band diagrams showing breakdown mechanisms of (a) tunneling and (b) avalanche multiplication (example initiated by hole current I_{po}).

When the field approaches 10^6 V/cm in Si, significant current begins to flow by means of this band-to-band tunneling process. To obtain such a high field, the junction must have relatively high impurity concentrations on both the *p*- and *n*-side. The mechanism of breakdown for *p-n* junctions with breakdown voltages less than about $4E_g/q$ is due to the tunneling effect. For junctions with breakdown voltages in excess of $6E_g/q$, the mechanism is caused by avalanche multiplication. At voltages between 4 and 6 E_g/q, the breakdown is due to a mixture of both avalanche and tunneling. Since the energy bandgaps E_g in Si and GaAs decrease with increasing temperature (refer to Chapter 1), the breakdown voltage in these semiconductors due to the tunneling effect has a negative temperature coefficient; that is, the breakdown voltage decreases with increasing temperature. This is because a given breakdown current J_t can be reached at smaller reverse voltages (or fields) at higher temperatures (Eq. 92). This temperature effect is generally used to distinguish the tunneling mechanism from the avalanche mechanism, which has a positive temperature coefficient; that is, the breakdown voltage increases with increasing temperature.

2.4.3 Avalanche Multiplication

Avalanche multiplication, or impact ionization, is the most-important mechanism in junction breakdown. The avalanche breakdown voltage imposes an upper limit on the reverse bias for most diodes, on the collector voltage of bipolar transistors, and on the drain voltages of MESFETs and MOSFETs. In addition, the impact ionization mech-

anism can be used to generate microwave power, as in IMPATT devices, and to amplify optical signals, as in avalanche photodetectors.

We first derive the basic ionization integral which determines the breakdown condition. Assume that a current I_{po} is incident at the left-hand side of the depletion region with width W_{Dm} (Fig. 15b). If the electric field in the depletion region is high enough that electron-hole pairs are generated by the impact ionization process, the hole current I_p will increase with distance through the depletion region and reach a value $M_p I_{po}$ at $x = W_{Dm}$. Similarly, the electron current I_n will increase from $I_n(W_{Dm}) = 0$ to $I_n(0) = I - I_{po}$, where the total current $I (= I_p + I_n)$ is constant at steady state. The incremental hole current is equal to the number of electron-hole pairs generated per second in the distance dx,

$$dI_p = I_p \alpha_p dx + I_n \alpha_n dx \tag{93}$$

or

$$\frac{dI_p}{dx} - (\alpha_p - \alpha_n)I_p = \alpha_n I . \tag{94}$$

The electron and hole ionization rates (α_n and α_p) have been considered in Chapter 1.

The solution of Eq. 94 with the boundary condition of $I = I_p(W_{Dm}) = M_p I_{po}$ is given by*

$$I_p(x) = I\left\{ \int_0^x \alpha_n \exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right]dx + \frac{1}{M_p}\right\} \Big/ \exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right] \tag{95}$$

where M_p is the multiplication factor of holes and is defined as

$$M_p \equiv \frac{I_p(W_{Dm})}{I_p(0)} \equiv \frac{I}{I_{po}} . \tag{96}$$

With a relationship†

$$\int_0^{W_{Dm}} (\alpha_p - \alpha_n) \exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right]dx = -\exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right]\Bigg|_0^{W_{Dm}}$$

$$= -\exp\left(\left[-\int_0^{W_{Dm}} (\alpha_p - \alpha_n)dx'\right] + 1\right), \tag{97}$$

Equation 95 can be evaluated at $x = W_{Dm}$ and be rewritten as

---

* Equation 94 has the form $y' + Py = Q$, where $y = I_p$. The standard solution is

$$y = \left[\int_0^x Q\left(\exp\int_0^x P dx'\right)dx + C\right] \Big/ \exp\int_0^x P dx'$$

where C is the constant of integration.

$$1 - \frac{1}{M_p} = \int_0^{W_{Dm}} \alpha_p \exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right]dx. \tag{98}$$

Note that M_p is a function of α_n in addition to α_p. The avalanche breakdown voltage is defined as the voltage where M_p approaches infinity. Hence the breakdown condition is given by the ionization integral

$$\int_0^{W_{Dm}} \alpha_p \exp\left[-\int_0^x (\alpha_p - \alpha_n)dx'\right]dx = 1. \tag{99a}$$

If the avalanche process is initiated by electrons instead of holes, the ionization integral is given by

$$\int_0^{W_{Dm}} \alpha_n \exp\left[-\int_x^{W_{Dm}} (\alpha_n - \alpha_p)dx'\right]dx = 1. \tag{99b}$$

Equations 99a and 99b are equivalent;[13] that is, the breakdown condition depends only on what is happening within the depletion region and not on the carriers (or primary current) that initiate the avalanche process. The situation does not change when a mixed primary current initiates the breakdown, so either Eq. 99a or Eq. 99b gives the breakdown condition. For semiconductors with equal ionization rates ($\alpha_n = \alpha_p = \alpha$) such as GaP, Eq. 99a or 99b reduces to the simple expression

$$\int_0^{W_{Dm}} \alpha dx = 1. \tag{100}$$

From the breakdown conditions described above and the field dependence of the ionization rates, the breakdown voltage, maximum electric field, and depletion-layer width can be calculated. As discussed previously, the electric field and potential in the depletion layer are determined from the solutions of the Poisson equation. Depletion-layer boundaries that satisfy Eq. 99a or 99b can be obtained numerically using an iteration method. With known boundaries we obtain the breakdown voltage

$$V_{BD} = \frac{\mathscr{E}_m W_{Dm}}{2} = \frac{\varepsilon_s \mathscr{E}_m^2}{2qN} \tag{101}$$

for one-sided abrupt junctions, and

---

† Let

$$U = \int_0^x y dx' \quad , \quad \frac{dU}{dx} = y \quad , \quad \frac{d}{dU}e^U = e^U.$$

The integral can be simplified to

$$\int y\left(\exp \int_0^x y dx'\right)dx = \int y e^U dx = \int e^U dU = e^U = \exp \int_0^x y dx'.$$

$$V_{BD} = \frac{2\mathscr{E}_m W_{Dm}}{3} = \frac{4\mathscr{E}_m^{3/2}}{3}\left(\frac{2\varepsilon_s}{qa}\right)^{1/2} \tag{102}$$

for linearly graded junctions, where N is the ionized background impurity concentration of the lightly doped side, a the impurity gradient, and $\mathscr{E}_m$ the maximum field.

Figure 16a shows the calculated breakdown voltage as a function of N for abrupt junctions in Si, $\langle 100 \rangle$-oriented GaAs, and GaP. The experimental results are generally in good agreement with the calculated values.[15] The dashed line in the figure indicates the upper limit of N for which the avalanche breakdown calculation is valid. This limitation is based on the criterion of $6E_g/q$. Above these corresponding values of N, the tunneling mechanism will contribute to the breakdown process and eventually dominates.

In GaAs, the ionization rates and thus breakdown voltage depend on crystal orientations, besides doping concentration (refer to Chapter 1).[16] At a doping concentration of around 10^{16} cm^{-3}, the breakdown voltages are essentially independent of orientations. At lower dopings, V_{BD} in $\langle 111 \rangle$ becomes the largest whereas at higher dopings, V_{BD} in $\langle 100 \rangle$ is the largest.

Figure 16b shows the calculated breakdown voltage versus the impurity gradient for linearly graded junctions. The dashed line indicates the upper limit of a for which the avalanche breakdown calculation is valid.

The calculated values of the maximum field $\mathscr{E}_m$ and the depletion-layer width at breakdown for the three semiconductors above are shown in Fig. 17a for the abrupt junctions, and in Fig. 17b for the linearly graded junctions. For the Si abrupt junctions, the maximum field at breakdown can be expressed as[17]

$$\mathscr{E}_m = \frac{4\times 10^5}{1-(1/3)\log_{10}(N/10^{16}\text{ cm}^{-3})} \qquad \text{V/cm} \tag{103}$$

where N is in cm^{-3}.

Because of the strong dependence of the ionization rates on the field, the maximum field at breakdown, sometimes called the *critical field*, varies very slowly with either N or a (within a factor of 4 over many orders of magnitude in N and a). Thus, as a first approximation, we can assume that for a given semiconductor, $\mathscr{E}_m$ has a fixed value. Then from Eqs. 101 and 102 we obtain $V_{BD} \propto N^{-1.0}$ for abrupt junctions and $V_{BD} \propto a^{-0.5}$ for linearly graded junctions. Figure 16 shows that the foregoing patterns are generally followed (within a factor of 3). Also as expected, for a given N or a, the breakdown voltage increases with the energy bandgap of the material, since the avalanche process requires band-to-band excitations. It should be cautioned that the critical field is only a rough guide line but not a fundamental material property. It assumes a uniform field over a large distance. For example, if there is a high field but only occurring over a small distance, breakdown would not happen since Eq. 100 cannot be satisfied. Also, the total voltage (field times distance) needs to be larger than the bandgap for band-to-band carrier multiplication. An example is the high field but small voltage drop in an accumulation layer.
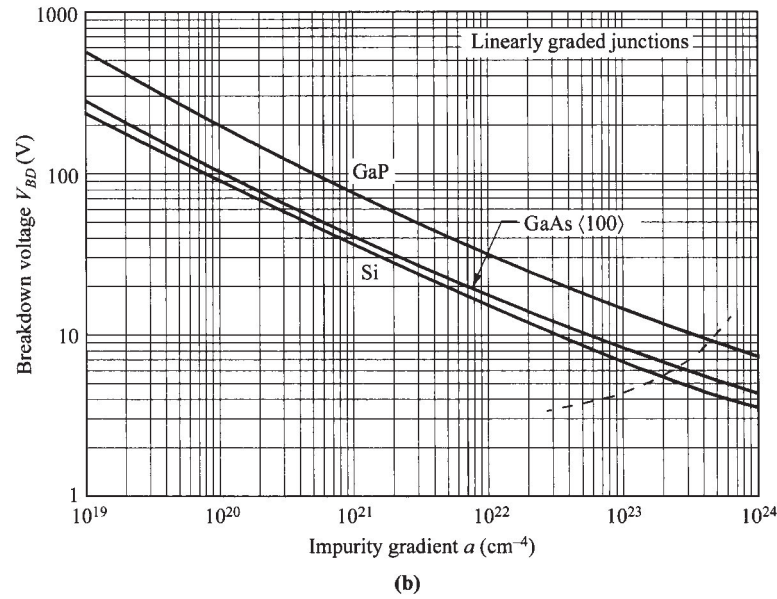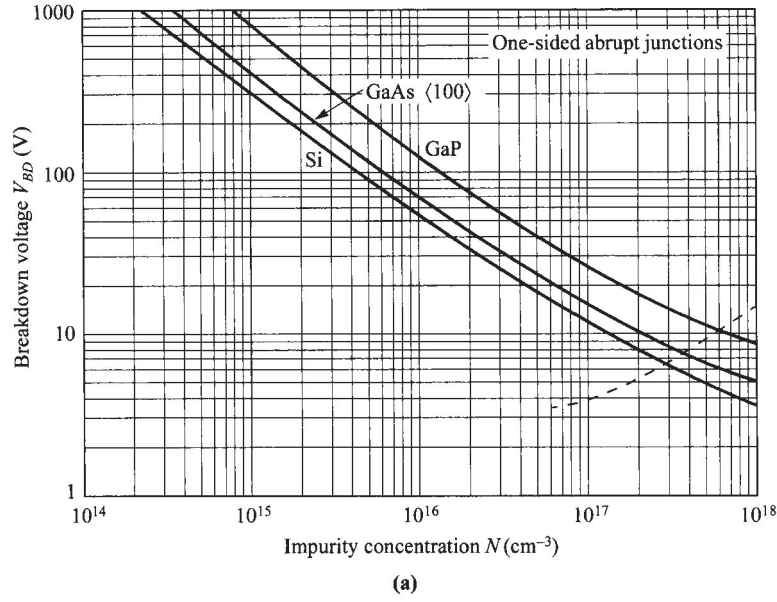
Fig. 16 Avalanche breakdown voltage in Si, ⟨100⟩-oriented GaAs, and GaP, for (a) one-sided abrupt junctions (vs. impurity concentration) and (b) linearly graded junctions (vs. impurity gradient). The dashed lines indicate the maximum doping or doping gradient beyond which tunneling will dominate the breakdown characteristics. (After Ref. 14.)
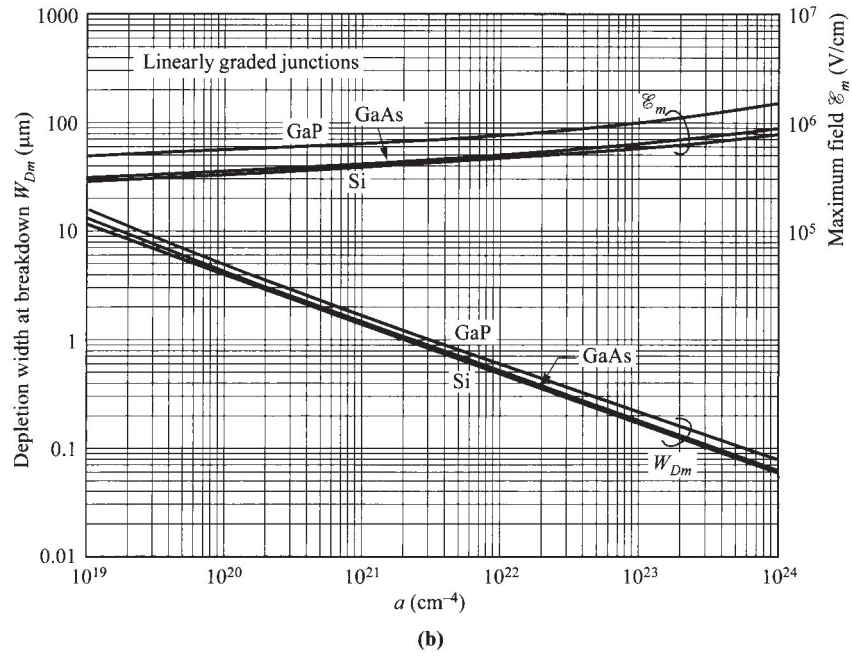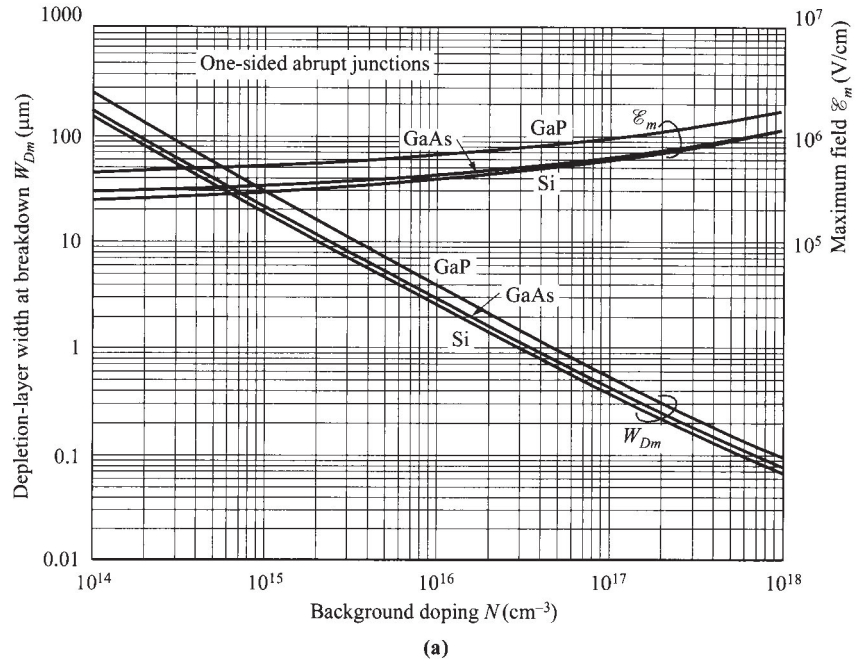
Fig. 17 Depletion-layer width and maximum field at breakdown in Si, $\langle 100 \rangle$-oriented GaAs, and GaP for (a) one-sided abrupt junctions and (b) linearly graded junctions. (After Ref. 14.)

An approximate universal expression can be given as follows for the results above comprising all semiconductors studied:

$$V_{BD} \approx 60\left(\frac{E_g}{1.1 \text{ eV}}\right)^{3/2}\left(\frac{N}{10^{16} \text{ cm}^{-3}}\right)^{-3/4} \quad \text{V} \tag{104}$$

for abrupt junctions where E_g is the room-temperature bandgap in eV, and N is the background doping in cm^{-3}; and

$$V_{BD} \approx 60\left(\frac{E_g}{1.1 \text{ eV}}\right)^{6/5}\left(\frac{a}{3\times10^{20} \text{ cm}^{-4}}\right)^{-2/5} \quad \text{V} \tag{105}$$

for linearly graded junctions where a is the impurity gradient in cm^{-4}.

For diffused junctions with a linear gradient near the junction and a constant doping on one side (Fig. 18 inset), the breakdown voltage lies between the two limiting cases considered previously[18] (Fig. 16). As shown in Fig. 18, for large a, the breakdown voltage of these junctions is given by the abrupt junction results (bottom line); on the other hand, for small a, V_{BD} will be given by the linearly graded junction results (parallel lines) and is independent of N_B.

In Figs. 16 and 17, it is assumed that the semiconductor layer is thick enough to support the maximum depletion-layer width W_{Dm} at breakdown. If, however, the semiconductor layer W is smaller than W_{Dm} (shown in Fig. 19, inset), the device will be punched through (i.e., the depletion layer reaches the n^+ substrate) prior to breakdown. As the reverse bias increases further, the depletion width cannot continue to expand and the device will break down prematurely. The maximum electric field $\mathscr{E}_m$
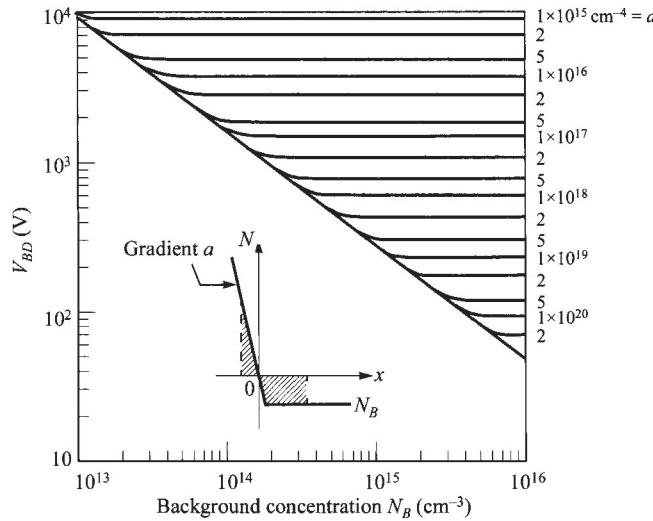

Fig. 18 Breakdown voltage for Si diffused junctions at 300 K. The inset shows the space-charge distribution. (After Ref. 18.)
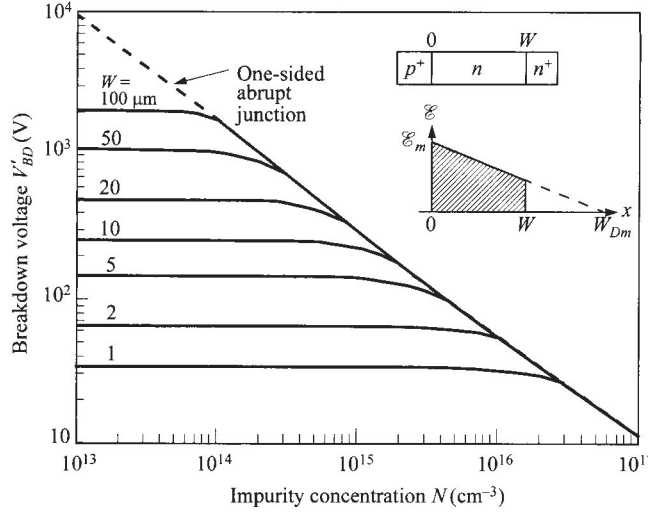
Fig. 19 Breakdown voltage for Si p^+-π-n^+ and p^+-v-n^+ junctions, where π stands for lightly doped p-type and v for lightly doped n-type. W is the thickness of the π- or v-region.

is essentially the same as for the nonpunched-through diode. Therefore, the reduced breakdown voltage V'_{BD} for the punched-through diode, compared to a regular device with V_{BD} for the same doping, can be given by

$$\frac{V'_{BD}}{V_{BD}} = \frac{\text{Shaded area in figure insert}}{(\mathscr{E}_m W_{Dm})/2}$$
$$= \left(\frac{W}{W_{Dm}}\right)\left(2 - \frac{W}{W_{Dm}}\right) . \tag{106}$$

Punch-through usually occurs when the doping concentration N becomes sufficiently low as in a p^+-π-n^+ or p^+-v-n^+ diode, where π stands for a lightly doped p-type and v for a lightly doped n-type semiconductor. The breakdown voltages for such diodes as calculated from Eq. 106 are shown in Fig. 19 as a function of the background doping for Si one-sided abrupt junction formed on epitaxial substrates (e.g., v on n^+ with the epitaxial-layer thickness W as a parameter). For a given thickness, the breakdown voltage approaches a constant value as the doping decreases, corresponding to the punch-through of the epitaxial layer.

The results shown so far are for avalanche breakdowns at room temperature. At higher temperatures the breakdown voltage increases. A qualitative explanation of this increase is that hot carriers passing through the depletion layer under a high field lose part of their energy to optical phonons via scattering, resulting in a smaller ionization rate (see Fig. 24 of Chapter 1). Therefore, the carriers lose more energy to the crystal lattice along a given distance at a constant field. Hence, the carriers must pass through a greater potential difference (or higher voltage) before they can acquire sufficient energy to generate an electron-hole pair. The predicted values of V_{BD} normal-

ized to the room-temperature value are shown in Fig. 20 for silicon. Note that there are substantial increases of the breakdown voltage, especially for lower dopings (or small gradient) at higher temperatures.[20]

Edge Effects. For junctions formed by a planar process, a very important junction curvature effect at the perimeter should be considered. A schematic diagram of a planar junction is shown in Fig. 21a. Note that at the perimeter, the depletion region is narrower and the field is higher. Since the cylindrical and/or spherical regions of the junction have a higher field intensity, the avalanche breakdown voltage is determined by these regions. The potential $\psi(r)$ and the electric field $\mathscr{E}(r)$ in a cylindrical or spherical *p-n* junction can be calculated from Poisson equation:

$$\frac{1}{r^n}\frac{d}{dr}[r^n\mathscr{E}(r)] = \frac{\rho(r)}{\varepsilon_s} \tag{107}$$

where n equals 1 for the cylindrical junction, and 2 for the spherical junction. The solution for $\mathscr{E}(r)$ can be obtained from this equation and is given by

$$\mathscr{E}(r) = \frac{1}{\varepsilon_s r^n}\int_{r_j}^{r} r^n\rho(r)dr + \frac{C_1}{r^n} \tag{108}$$

where r_j is the radius of curvature of the metallurgical junction, and the constant C_1 must be adjusted so that the integration of the field is equal to the built-in potential.

The calculated results for Si one-sided abrupt junctions at 300 K can be expressed by a simple equation:[18]

$$\frac{V_{CY}}{V_{BD}} = \left[\frac{1}{2}(\eta^2 + 2\eta^{6/7})\ln(1 + 2\eta^{-8/7}) - \eta^{6/7}\right] \tag{109}$$

for cylindrical junctions, and

$$\frac{V_{SP}}{V_{BD}} = [\eta^2 + 2.14\eta^{6/7} - (\eta^3 + 3\eta^{13/7})^{2/3}] \tag{110}$$

for spherical junctions, where V_{CY} and V_{SP} are the breakdown voltages of cylindrical and spherical junctions, respectively, V_{BD} and W_{Dm} are the breakdown voltage and maximum depletion width of a plane junction having the same background doping, and $\eta \equiv r_j/W_{Dm}$. Figure 21b illustrates the numerical results as a function of η. Clearly, as the radius of curvature becomes smaller, so does the breakdown voltage. However, for linearly graded cylindrical or spherical junctions, the calculated results show that the breakdown voltage is relatively independent of its radius of curvature.[21]

Another edge effect that causes premature breakdown is due to an MOS (metal-oxide semiconductor) structure over the junction at the surface. Such a configuration is often called a gated diode. At certain gate biases, the field near the gate edge is higher than in the planar portion of the junction and breakdown changes location from the surface area of the metallurgical junction to the edge of the gate. This gate-voltage dependence of breakdown is shown in Fig. 22. At high positive gate bias on a p^+-n junction, the p^+-surface is depleted while the n-surface is accumulated. Breakdown occurs near the metallurgical junction at the surface. As the gate bias is swept
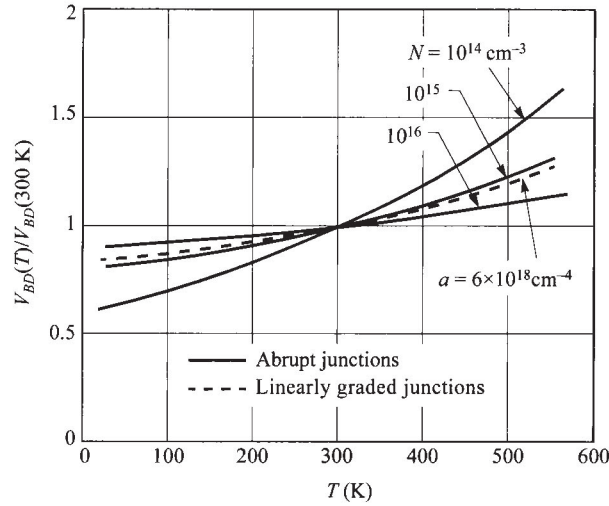
Fig. 20 Normalized avalanche breakdown voltage versus lattice temperature, in silicon. The breakdown voltage generally increases with temperature. (After Ref. 19.}
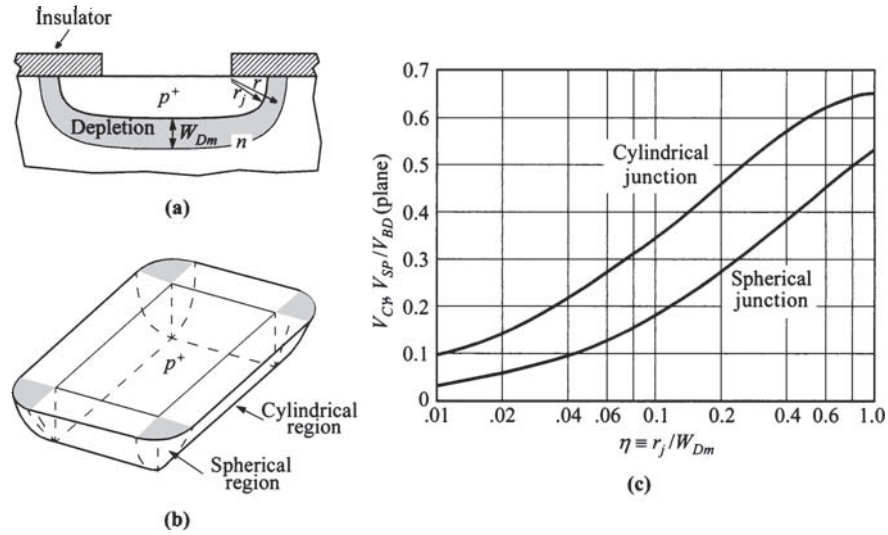

Fig. 21 (a) A planar diffusion or implantation process forms a junction curvature near the edges of the mask with r_j the radius of curvature. (b) Three-dimensional view of the junction curvature showing the spherical region at the corners. (c) Normalized breakdown voltage of cylindrical and spherical junctions as a function of the normalized radius of curvature. (After Ref. 18.)
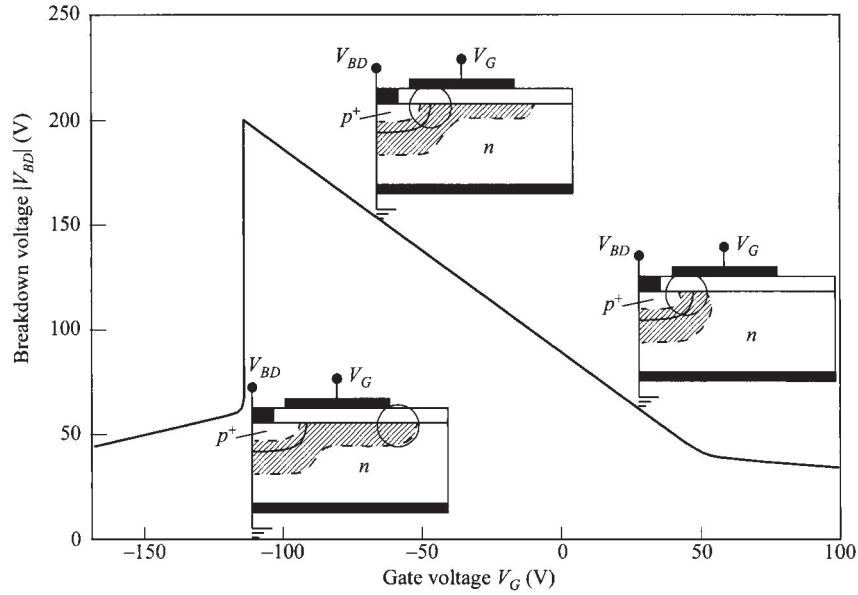
113

Fig. 22 Gate-voltage dependence of breakdown in a gated diode. The location of high-field breakdown shifts with gate bias. (After Ref. 22.)

more negatively, the location of breakdown moves toward the *n*-side (to the right). In the middle gate-bias range, the breakdown voltage has a linear dependence on the gate bias, with[23]

$$V_{BD} = mV_G + \text{constant} \tag{111}$$

and $m \leq 1$. At some high negative gate bias, the field directly under the gate edge is high enough to cause breakdown, and the breakdown voltage collapses. This gated-diode breakdown phenomenon is reversible and the measurement can be repeated. To minimize this edge effect, the oxide thickness should be above a critical value.[22] This mechanism is also responsible for the gate-induced drain leakage (GIDL) of the MOSFET (see Section 6.4.5).

2.5 TRANSIENT BEHAVIOR AND NOISE

2.5.1 Transient Behavior

For switching applications the transitions from forward bias to reverse bias and vice versa must be nearly abrupt and the transient time short. For a *p-n* junction, while the latter is reasonably fast, the response from forward to reverse is limited by minority-carrier charge storage. Figure 23a shows a simple circuit in which a forward current I_F flows in the *p-n* junction; at time $t = 0$, the switch S is suddenly thrown to the right,
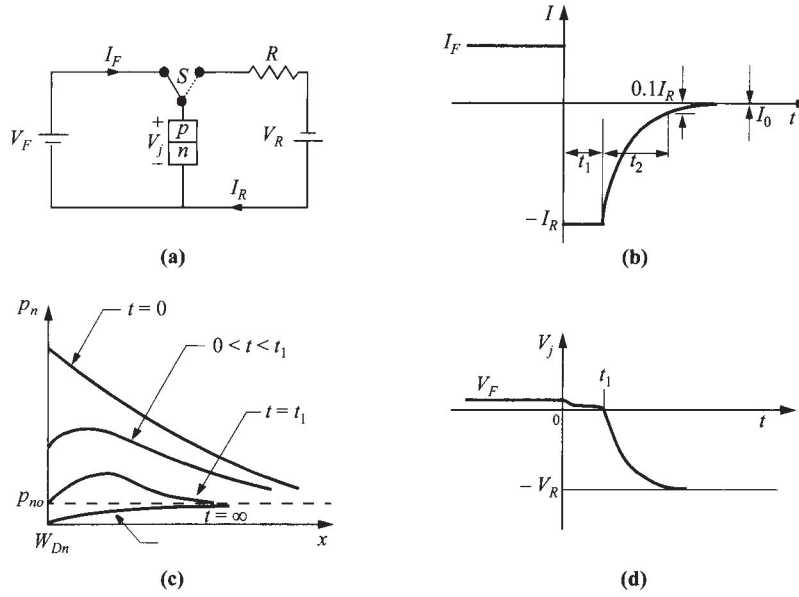
Fig. 23 Transient behavior of a *p-n* junction. (a) Basic switching circuit. (b) Transient current response. (c) Minority-carrier distribution outside depletion edge for various time intervals. (d) Transient junction-voltage response. (After Ref. 24.)

and an initial reverse current $I_R = (V_R - V_F)/R$ flows. The transient time is defined as the time in which the current drops to 10% of the initial reverse current I_R, and is equal to the sum of t_1 and t_2 as shown in Fig. 23b, where t_1 and t_2 are the time intervals for the constant-current phase and the decay phase, respectively.

Consider the constant-current phase (also called storage phase) first. The continuity equation as given in Chapter 1 can be written for the *n*-type side of a p^+-*n* junction $(p_{po} \gg n_{no})$ as

$$\frac{\partial p_n(x, t)}{\partial t} = D_p \frac{\partial^2 p_n(x, t)}{\partial x^2} - \frac{p_n(x, t) - p_{no}}{\tau_p}. \tag{112}$$

The boundary conditions are that at $t = 0$ the initial distribution of holes is a steady-state solution to the diffusion equation, and that under forward bias the voltage across the junction is given from Eq. 53b as

$$V_j(t) = \frac{kT}{q} \ln\left[\frac{p_n(0, t)}{p_{no}}\right]. \tag{113}$$

The distribution of the minority-carrier density p_n with time is shown in Fig. 23c. From Eq. 113 it can be calculated that, as long as $p_n(0,t)$ is greater than p_{no} (in the interval $0 < t < t_1$), the junction voltage V_j remains of the order of kT/q, as shown in Fig. 23d. In this time interval the reverse current is approximately constant and we

have the constant-current phase. The solution of the time-dependent continuity equation gives t_1 by the transcendental equations[24]

$$\text{erf}\sqrt{\frac{t_1}{\tau_p}} = \frac{1}{1 + (I_R/I_F)}.$$ (114)

However, an explicit expression for t_1 can be obtained, using a charge-control model which can also provide some insight into the problem. The stored minority-carrier charge in the lightly doped side is given by the integral

$$Q_s = qA\int \Delta p_n dx.$$ (115)

Integration of the continuity equation, after the current is switched to the reversed mode, becomes

$$-I_R = \frac{dQ_s}{dt} + \frac{Q_s}{\tau_p}.$$ (116)

With the initial condition given by the forward current $Q_s(0) = I_F\tau_p$, the solution is given by

$$Q_s(t) = \tau_p\left[-I_R + (I_F + I_R)\exp\left(\frac{-t}{\tau_p}\right)\right].$$ (117)

By setting $Q_s = 0$, t_1 can be obtained as

$$t_1 = \tau_p\ln\left(1 + \frac{I_F}{I_R}\right).$$ (118)

A comparison of Eq. 118 to the exact solution of Eq. 114 shows that this estimate gives higher values by a factor of ≈ 2 for $I_F/I_R = 0.1$ and ≈ 20 for $I_F/I_R = 10$.

After t_1, the hole density starts to decrease below its equilibrium value p_{no}. The junction voltage tends to reach $-V_R$ and a new boundary condition now holds. This phase is the decay phase with the initial boundary condition $p_n(0,t_1) = p_{no}$. The solution for t_2 is given by another transcendental equation

$$\text{erf}\sqrt{\frac{t_2}{\tau_p}} + \frac{\exp(-t_2/\tau_p)}{\sqrt{\pi t_2/\tau_p}} = 1 + 0.1\left(\frac{I_R}{I_F}\right).$$ (119)

The total results for t_1 and t_2 are shown in Fig. 24 where the solid lines are for the plane junction with the length of the n-type material W much greater than the diffusion length ($W \gg L_p$), and the dashed lines are for the narrow-base junction with $W \ll L_p$. For a large I_R/I_F ratio, the transient time can be approximated by

$$t_1 + t_2 \approx \frac{\tau_p}{2}\left(\frac{I_R}{I_F}\right)^{-2}$$ (120)
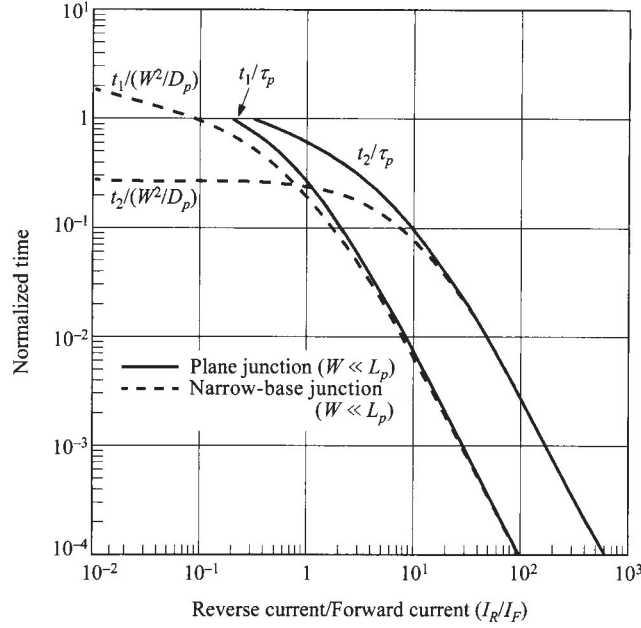
for $W \gg L_p$, or

Fig. 24 Normalized time versus the ratio of reverse current to forward current. W is width of the n-region in a p^+-n junction. (After Ref. 24.)

$$t_1 + t_2 \approx \frac{W^2}{2D_p}\left(\frac{I_R}{I_F}\right)^{-2} \tag{121}$$

for $W \ll L_p$. For example, if one switches a junction (of $W \gg L_p$) from forward 10 mA to reverse 10 mA ($I_R/I_F = 1$), the time for the constant-current phase is $0.3\,\tau_p$, and that for the decay phase is about $0.6\,\tau_p$. Total transient time is then $0.9\,\tau_p$. A fast switch requires that τ_p be small for all cases. The lifetime τ_p can be substantially reduced by introducing impurities with deep levels in the forbidden gap, such as gold in silicon.

2.5.2 Noise

The term "noise" refers to spontaneous fluctuations in the current passing through, or the voltage developed across, semiconductor bulk materials or devices. Since semiconductor devices are mainly used to amplify small signals or to measure small physical quantities, spontaneous fluctuations in current or voltage set a lower limit to these signals. It is important to know the factors contributing to these limits, to use this knowledge to optimize operating conditions, and to find new methods and new technologies to reduce noise.

Observed noise is generally classified into (1) thermal noise or Johnson noise, (2) flicker noise, and (3) shot noise. Thermal noise occurs in any conductor or semiconductor device and is caused by the random thermal motion of the current carriers. It

is also called white noise because its level is the same at all frequencies. The open-circuit mean-square voltage of thermal noise is given by[25,26]

$$\langle V_n^2 \rangle = 4kTBR \tag{122}$$

where B is the bandwidth in Hz, and R the real part of the dynamic impedance (dV/dI) between terminals. At room temperature, for a semiconductor device with 1 kΩ resistance, the root-mean-square voltage $\sqrt{\langle V_n^2 \rangle}$ measured with a 1-Hz bandwidth is only about 4 nV.

Flicker noise is distinguished by its peculiar spectral distribution which is proportional to $1/f^\alpha$ with α generally close to unity (the so-called $1/f$ noise). Flicker noise is thus important at lower frequencies. For most semiconductor devices, the origin of flicker noise is the surface effect. The $1/f$ noise-power spectrum has been correlated both qualitatively and quantitatively with the lossy part of the metal-insulator-semiconductor (MIS) gate impedance due to carrier recombination at the interface traps.

Shot noise is due to the discreteness of charge carriers that contribute to current flow, and it constitutes the major noise in most semiconductor devices. It is independent of frequency (white spectrum) at low and intermediate frequencies. At higher frequencies the shot-noise spectrum also becomes frequency-dependent. The mean-square noise current of shot noise for a *p-n* junction is given by

$$\langle i_n^2 \rangle = 2qB|I| \tag{123}$$

where I can be forward or reverse current. For low injection the total mean-square noise current (neglecting $1/f$ noise) is the sum

$$\langle i_n^2 \rangle = \frac{4kTB}{R} + 2qB|I| . \tag{124}$$

From the Shockley equation we obtain

$$\frac{1}{R} = \frac{dI}{dV} = \frac{d}{dV}\left\{ I_0 \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] \right\} = \frac{qI_0}{kT}\exp\left(\frac{qV}{kT}\right) . \tag{125}$$

Substituting Eq. 125 into Eq. 124 yields for the forward-bias condition,

$$\langle i_n^2 \rangle = 4qI_0 B \exp\left(\frac{qV_F}{kT}\right) + 2qI_0 B\left[\exp\left(\frac{qV_F}{kT}\right) - 1 \right]$$

$$\approx 6qI_0 B \exp\left(\frac{qV_F}{kT}\right) . \tag{126}$$

Experimental measurements indeed confirm that the mean-square noise current is proportional to the saturation current I_0, which can be increased by irradiation.

2.6 TERMINAL FUNCTIONS

A *p-n* junction is a two-terminal device that can perform various terminal functions, depending upon its biasing condition as well as its doping profile and device geometry. In this section we discuss briefly some interesting device performances based on

its current-voltage, capacitance-voltage, and breakdown characteristics discussed in previous sections. Many other related two-terminal devices will be considered in subsequent chapters (e.g., tunnel diode in Chapter 8 and IMPATT diode in Chapter 9).

2.6.1 Rectifier

A rectifier is a two-terminal device that gives a very low resistance to current flow in one direction and a very high resistance in the other direction, i.e., it allows current in only one direction. The forward and reverse resistances of a rectifier can be derived from the current-voltage relationship of a practical diode,

$$I = I_0 \left[\exp\left(\frac{qV}{\eta kT}\right) - 1 \right] \tag{127}$$

where I_0 is the saturation current and the ideality factor η generally has a value between 1 (for diffusion current) and 2 (for recombination current). The forward dc (or static) resistance R_F and small-signal (or dynamic) resistance r_F are obtainable from Eq. 127:

$$R_F \equiv \frac{V_F}{I_F} \approx \frac{V_F}{I_0} \exp\left(\frac{-qV_F}{\eta kT}\right), \tag{128}$$

$$r_F \equiv \frac{dV_F}{dI_F} \approx \frac{\eta kT}{qI_F}. \tag{129}$$

The reverse dc resistance R_R and small-signal resistance r_R are given by

$$R_R \equiv \frac{V_R}{I_R} \approx \frac{V_R}{I_0}, \tag{130}$$

$$r_R \equiv \frac{dV_R}{dI_R} = \frac{\eta kT}{qI_0} \exp\left(\frac{q|V_R|}{\eta kT}\right). \tag{131}$$

Comparing Eqs. 128–131 shows that the dc rectification ratio R_R/R_F varies with the factor $(V_R/V_F)\exp(qV_F/\eta kT)$, while the ac rectification ratio r_R/r_F varies with $(I_F/I_0)\exp(q|V_R|/\eta kT)$.

p-n junction rectifiers generally have slow switching speeds; that is, a significant time delay is necessary to obtain high impedance after switching from the forward-conduction state to the reverse-blocking state. This time delay (proportional to the minority-carrier lifetime as shown in Fig. 24) is of little consequence in rectifying 60-Hz currents. For high-frequency applications, the lifetime should be sufficiently reduced to maintain rectification efficiency. The majority of rectifiers have power-dissipation capabilities from 0.1 to 10 W, reverse breakdown voltages from 50 to 2500 V (for a high-voltage rectifier two or more *p-n* junctions are connected in series), and switching times from 50 ns for low-power diodes to about 500 ns for high-power diodes.

A rectifier has many circuit applications.[27] It is used to transform ac signals into different specials waveforms. Examples are half-wave and full-wave rectifiers,

clipper and clamper circuits, peak detector (demodulator), etc. It can also be used as a ESD (electrostatic discharge) protection device.

2.6.2 Zener Diode

A Zener diode (also called voltage regulator) has a well-controlled breakdown voltage, called the Zener voltage, with sharp breakdown characteristics in the reverse-bias region. Prior to breakdown, the diode has a very high resistance; after breakdown the diode has a very small dynamic resistance. The terminal voltage is thus limited (or regulated) by the breakdown voltage, and this is used to establish a fixed reference voltage.

Most Zener diodes are made of Si, because of the low saturation current in Si diodes and the advanced Si technology. They are special *p-n* junctions with higher doping concentrations on both sides. As discussed in Section 2.4, for breakdown voltage V_{BD} larger than $6E_g/q$ ($\approx$ 7 V for Si), the breakdown mechanism is mainly avalanche multiplication, and the temperature coefficient of V_{BD} is positive. For $V_{BD} < 4E_g/q$ ($\approx$ 5 V for Si), the breakdown mechanism is band-to-band tunneling, and the temperature coefficient of V_{BD} is negative. For $4E_g/q < V_{BD} < 6E_g/q$, the break-down is due to a combination of these two mechanisms. One can connect, for example, a negative-temperature-coefficient diode in series with a positive-tempera-ture-coefficient diode to produce a temperature-independent regulator (with a tem-perature coefficient of the order of 0.002% per °C), which is suitable as a voltage reference.

2.6.3 Varistor

A varistor (variable resistor) is a two-terminal device that shows nonohmic behavior, i.e., voltage-dependent resistance.[28] Equations 128 and 129 show the nonohmic char-acteristics of a *p-n* junction diode in the forward-bias region. Similar nonohmic char-acteristics are obtainable from metal-semiconductor contacts considered in Chapter 3. An interesting application of varistors is their use as a symmetrical frac-tional-voltage ($\approx$ 0.5 V) limiter by connecting two diodes in parallel, oppositely poled. The two-diode unit will exhibit the forward *I-V* characteristics in either direc-tion. A varistor, being a nonlinear device, is also useful in microwave modulation, mixing, and detection (demodulation). Varistors based on metal-semiconductor con-tacts are more common due to their higher speed from the absence of minority-charge storage.

2.6.4 Varactor

The term *varactor* comes from *variable reactor* and means a device whose reactance (or capacitance) can be varied in a controlled manner with a dc bias voltage. Varactor diodes are widely used in parametric amplification, harmonic generation, mixing, detection, and voltage-variable tuning.

For this application, the forward bias is to be avoided because of excessive current which is undesirable for any capacitor. The basic capacitance-voltage relationships in

reverse bias have already been derived in Section 2.2. We shall now extend the previous derivations of abrupt and linearly graded doping distributions to a more general case. The one-dimensional Poisson equation is given as

$$\frac{d^2 \psi_i}{dx^2} = - \frac{qN}{\varepsilon_s} \tag{132}$$

where N is the generalized doping distribution (negative sign for donors) as shown in Fig. 25a (assuming one side is heavily doped):

$$N = Bx^m \quad \text{for} \quad x \geq 0. \tag{133}$$

For $m = 0$ we have $N = B$, corresponding to the uniformly doped (or one-sided abrupt junction) case. For $m = 1$, the doping profile corresponds to a one-sided linearly graded case. For $m < 0$, the device is called a "hyper-abrupt" junction. The hyper-abrupt doping profile can be achieved by an epitaxial process or by ion implantation. The boundary conditions are $\psi(x = 0) = 0$ and $\psi(x = W_D) = V_R + \psi_{bi}$, where V_R is the applied reverse voltage and ψ_{bi} is the built-in potential. Integrating the Poisson equation with the boundary conditions, we obtain for the depletion-layer width and the differential capacitance per unit area[29]

$$W_D = \left[\frac{\varepsilon_s(m + 2)(V_R + \psi_{bi})}{qB} \right]^{1/(m+2)}, \tag{134}$$

$$C_D \equiv \frac{\varepsilon_s}{W_D} = \left[\frac{qB\varepsilon_s^{m+1}}{(m+2)(V_R + \psi_{bi})} \right]^{1/(m+2)} \propto (V_R + \psi_{bi})^{-s}, \tag{135}$$

$$s \equiv \frac{1}{m+2}. \tag{136}$$

One important parameter in characterizing the varactor is the sensitivity defined by[30]

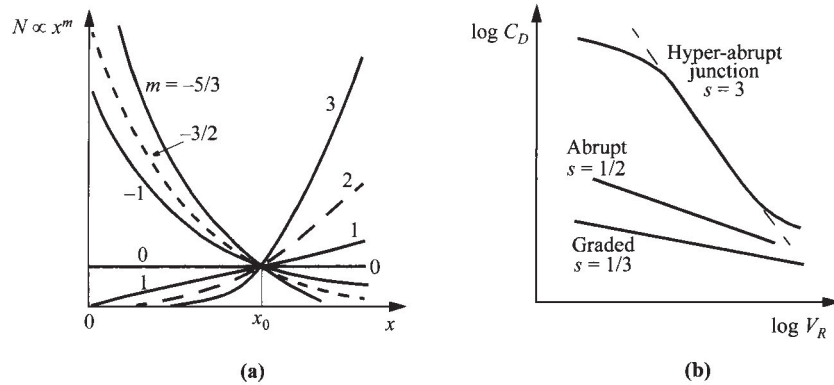(a)   (b)

Fig. 25 (a) Various impurity distributions (normalized at x_0) for varactors. (b) Log-log plot of depletion-layer capacitance versus reverse bias. (After Refs. 29 and 30.)

$$-\frac{dC_D}{C_D}\frac{V_R}{dV_R} = -\frac{d(\log C_D)}{d(\log V_R)} = \frac{1}{m+2} = s.$$ (137)

The larger the s, the larger will be the capacitance variation with biasing voltage. For linearly graded junctions, $m = 1$ and $s = 1/3$; for abrupt junctions, $m = 0$ and $s = 1/2$; for hyper-abrupt junction with $m = -1, -3/2,$ or $-5/3$, the value of s is 1,2, or 3, respectively. The capacitance-voltage relationships for these junction diodes are shown in Fig. 25b. The hyper-abrupt junction, as expected, has the highest sensitivity and gives rise to the largest capacitance variation.

2.6.5 Fast-Recovery Diode

Fast-recovery diodes are designed to give ultrahigh switching speed. The devices can be classified into two types: *p-n* junction diodes and metal-semiconductor diodes. The general switching behavior of both types can be described by Fig. 23b. The total recovery time $(t_1 + t_2)$ for a *p-n* junction diode can be substantially reduced by introducing recombination centers, such as Au in Si, to reduce the carrier lifetime. Although the recovery time is directly proportional to the lifetime τ, as shown in Fig. 24, it is not possible, unfortunately, to reduce the recovery time indefinitely by introducing an extremely large number of recombination centers N_t, because the reverse generation current of a *p-n* junction is proportional to N_t (Eqs. 66 and 67). For direct bandgap semiconductors, such as GaAs, the minority-carrier lifetimes are generally much smaller than that of Si. This results in ultra-high-speed GaAs *p-n* junction diodes with recovery times of the order of 0.1 ns or less. For Si the practical recovery time is in the range of 1 to 5 ns.

The metal-semiconductor diodes (Schottky diodes) fundamentally exhibit ultrahigh-speed characteristics, because they are majority-carrier devices and the minority-carrier storage effect is negligible. We discuss metal-semiconductor contacts in detail in Chapter 3.

2.6.6 Charge-Storage Diode

In contrast to fast-recovery diodes, a charge-storage diode is designed to store a charge while conducting in the forward direction and, upon switching to the reverse direction, to conduct a reverse current for a short period. A particularly interesting charge-storage diode is the step-recovery diode (also called the snapback diode) that conducts in the reverse direction for a short period and then abruptly cuts off the current as the stored charge has been dissipated. In other words, it is desirable here to reduce the decay phase or t_2 without shortening the storage phase or t_1. Most charge-storage diodes are made from Si with relatively long minority-carrier lifetimes ranging from 0.5 to 5 μs. Note that the lifetimes are about 1000 times longer than for fast-recovery diodes. The mechanism to reduce the decay phase is by a special doping profile such that the injected charge is confined closer to the junction. This cutoff occurs in the range of picoseconds and results in a fast-rising wavefront which is rich in harmonics. Because of these characteristics, step-recovery diodes are used in harmonic generation and pulse shaping.

2.6.7 *p-i-n* Diode

A *p-i-n* diode is a *p-n* junction with an intrinsic layer (*i*-region) sandwiched between the *p*-layer and the *n*-layer. In practice, however, the idealized *i*-region is approximated by either a high-resistivity *p*-layer (referred to as π-layer) or a high-resistivity *n*-layer (*ν*-layer). The *p-i-n* diode has found wide applications in microwave circuits. Its special feature is a wide intrinsic layer that provides unique properties such as low and constant capacitance, high breakdown voltage in reverse bias, and most interestingly, as a variolosser (variable attenuator) by controlling the device resistance which varies approximately linearly with the forward bias current. The switching time is approximately given by $W/2v_s$, where W is the width of the *i*-region.[31] It can modulate signals up to the GHz range. Furthermore, the forward characteristics of a thyristor (refer to Chapter 11) in its on-state closely resemble those of a *p-i-n* diode.

At near zero or low reverse bias, the lightly doped intrinsic layer starts to be fully depleted, and the capacitance is given by

$$C = \frac{\varepsilon_s}{W}. \tag{138}$$

Once fully depleted, its capacitance is independent of reverse bias. Figure 19 gives the breakdown voltage of a *p-i-n* diode under reverse bias. Since there is little net charge within the intrinsic layer, the electric field is constant and the breakdown voltage can be estimated by

$$V_{BD} \approx \mathscr{E}_m W \tag{139}$$

where the maximum breakdown field $\mathscr{E}_m$ for Si at lower dopings is about 2.5×10^5 V/cm. These two equations show that the width of the *i*-region W controls the trade-off between frequency response and power (from maximum voltage).

Under forward conditions, holes are injected from the *p*-region and electrons from the *n*-region. As the injected carrier densities are nearly equal (and uniform) due to charge neutrality, they are much higher than the *i*-region doping concentration, so the *p-i-n* diode is generally operated in the high-injection condition, $\Delta p = \Delta n \gg n_i$. The current conduction is via recombination within the *i*-region and is given by (see Eq. 74)

$$J_{re} = \int_0^W qU dx = \frac{qWn_i}{2\tau} \exp\!\left(\frac{qV_F}{2kT}\right). \tag{140}$$

For a detailed discussion of the dc *I-V* characteristics, the readers are referred to Section 11.2.4.

The most interesting phenomenon for a *p-i-n* diode, however, is for small signals at high frequencies ($> 1/2\pi\tau$) at which the stored carriers within the intrinsic layer are not completely swept away by the RF signal or by recombination. At these frequencies there is no rectification and the *p-i-n* diode behaves like a pure resistor whose value is determined solely by the injected charge, proportional to the dc bias current. This dynamic RF resistance is simply given by

$$R_{RF} = \rho\frac{W}{A} = \frac{W}{q\Delta n(\mu_n + \mu_p)A}$$
$$= \frac{W^2}{J_F\tau(\mu_n + \mu_p)A} \;.$$

(141)

Here the relationship $J_F = qW\Delta n/\tau$ has been assumed. The RF resistance is controlled by the dc bias current, and typical characteristics are shown in Fig. 26.

2.7 HETEROJUNCTIONS

Some properties of heterojunctions have been discussed in Section 1.7. When the two semiconductors have the same type of conductivity, the junction is called an *isotype* heterojunction. When the conductivity types differ, the junction is called an *anisotype* heterojunction which is a much more useful and common structure than its counterpart. In 1951, Shockley proposed the abrupt heterojunction to be used as an efficient emitter-base injector in a bipolar transistor.[33] In the same year, Gubanov published a theoretical paper on heterojunctions.[34] Kroemer later analyzed a similar, although graded, heterojunction as a wide-bandgap emitter.[35] Since then, heterojunctions have been extensively studied, and many important applications have been made, among them the room-temperature injection laser, light-emitting diode (LED), photodetector, and solar cell, to name a few. In many of these applications, by forming periodic heterojunctions with layer thickness of the order of 10 nm, we utilize the interesting properties of quantum wells and superlattices. Additional information on heterojunctions can be found in Refs. 36–39.

2.7.1 Anisotype Heterojunction

The energy-band model of an idealized anisotype abrupt heterojunction without interface traps was proposed by Anderson[40] based on the previous work of Shockley. We consider this model next, since it can adequately explain most transport processes,
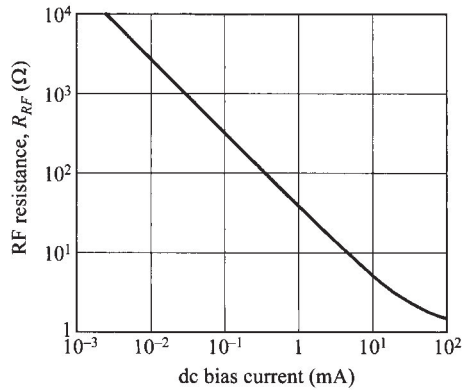

Fig. 26 Typical RF resistance as a function of dc forward current. (After Ref. 32.)

and only slight modification of the model is needed to account for nonideal cases such as interface traps. Figures 27a and c show the energy-band diagrams of two isolated semiconductors of opposite types. The two semiconductors are assumed to have different bandgaps E_g, different permittivities ε_s, different work functions ϕ_m, and different electron affinities χ. Work function and electron affinity are defined as the energy required to remove an electron from the Fermi level E_F and from the bottom of the conduction band E_C, respectively, to a position just outside the material (vacuum level). The difference in energy of the conduction-band edges in the two semiconductors is represented by ΔE_C and that in the valence-band edges by ΔE_V. The
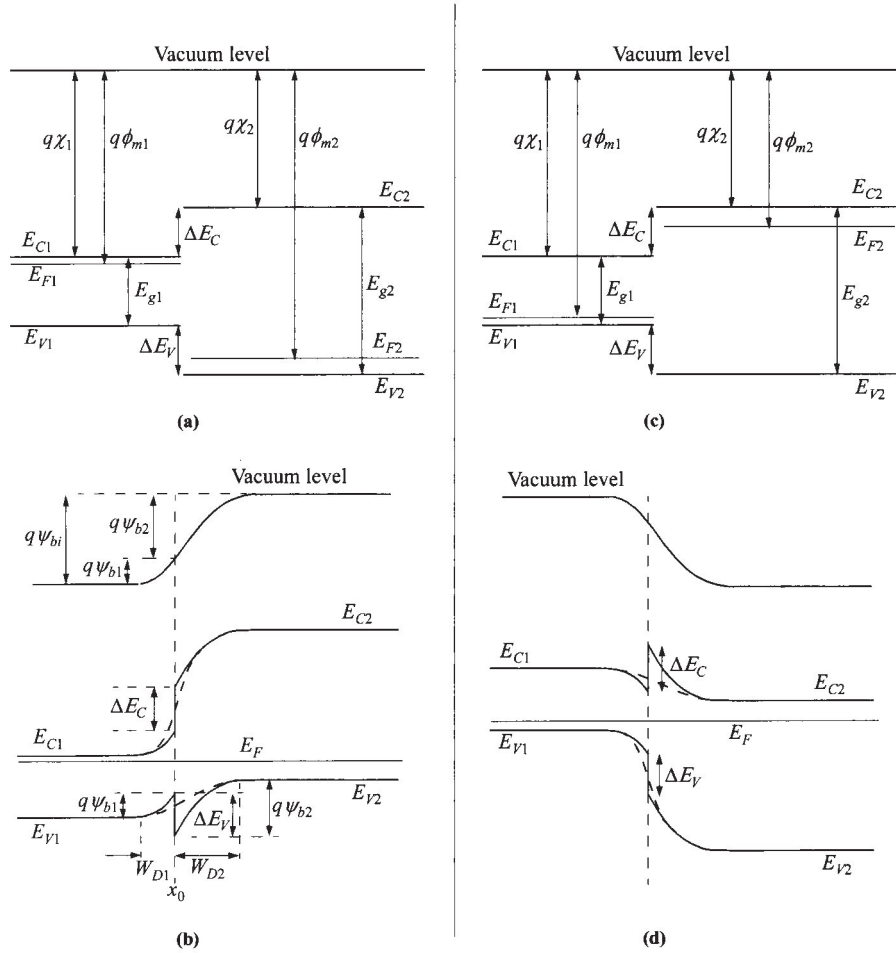

Fig. 27 Energy-band diagrams for (a) two isolated semiconductors of opposite types and different E_g (of which the smaller bandgap is n-type) and (b) their idealized anisotype heterojunction at thermal equilibrium. In (c) and (d), the smaller bandgap is p-type. In (b) and (d), the dashed lines across the junctions represent graded composition. (After Ref. 40.)

electron affinity rule ($\Delta E_C = q\Delta\chi$) shown in Figure 27 may not be a valid assumption in all cases. However, by choosing ΔE_C as an empirical quantity, the Anderson model remains satisfactory and unaltered.[41]

When a junction is formed between these semiconductors, the energy-band profile at equilibrium is as shown in Fig. 27b for an *n-p* anisotype heterojunction where, in this example, the narrow-bandgap material is *n*-type. Since the Fermi level must coincide on both sides in equilibrium and the vacuum level is everywhere parallel to the band edges and is continuous, the discontinuity in the conduction-band edges (ΔE_C) and valence-band edges (ΔE_V) is invariant with doping in those cases where E_g and χ are not functions of doping (i.e., nondegenerate semiconductors). The total built-in potential ψ_{bi} is equal to the sum of the partial built-in voltages ($\psi_{b1} + \psi_{b2}$), where ψ_{b1} and ψ_{b2} are the electrostatic potentials supported at equilibrium by semiconductors 1 and 2, respectively.* From Fig. 27, it is apparent that since at equilibrium, $E_{F1} = E_{F2}$, the total built-in potential is given by

$$\psi_{bi} = |\phi_{m1} - \phi_{m2}|. \tag{142}$$

The depletion widths and capacitance can be obtained by solving the Poisson equation for the step junction on either side of the interface. One boundary condition is the continuity of electric displacement, that is, $\mathscr{D}_1 = \mathscr{D}_2 = \varepsilon_{s1}\mathscr{E}_1 = \varepsilon_{s2}\mathscr{E}_2$ at the interface. We obtain

$$W_{D1} = \left[\frac{2N_{A2}\varepsilon_{s1}\varepsilon_{s2}(\psi_{bi} - V)}{qN_{D1}(\varepsilon_{s1}N_{D1} + \varepsilon_{s2}N_{A2})}\right]^{1/2}, \tag{143a}$$

$$W_{D2} = \left[\frac{2N_{D1}\varepsilon_{s1}\varepsilon_{s2}(\psi_{bi} - V)}{qN_{A2}(\varepsilon_{s1}N_{D1} + \varepsilon_{s2}N_{A2})}\right]^{1/2}, \tag{143b}$$

and

$$C_D = \left[\frac{qN_{D1}N_{A2}\varepsilon_{s1}\varepsilon_{s2}}{2(\varepsilon_{s1}N_{D1} + \varepsilon_{s2}N_{A2})(\psi_{bi} - V)}\right]^{1/2}. \tag{144}$$

The relative voltage supported in each semiconductor is

$$\frac{\psi_{b1} - V_1}{\psi_{b2} - V_2} = \frac{N_{A2}\varepsilon_{s2}}{N_{D1}\varepsilon_{s1}} \tag{145}$$

where the applied voltage is divided into the two regions $V = V_1 + V_2$. It is apparent that the foregoing expressions will reduce to the expression for the *p-n* junction (homojunction) discussed in Section 2.3, when both sides of the heterojunction become the same materials.

In considering the current flow, the example in Fig. 27b shows that the conduction-band edge E_C increases monotonically while the valence-band edge E_V goes through some peak near the junction. The hole current could become complicated

---

* The convention is to list the material with the smaller bandgap as the first symbol.

because of the added barrier which might present a bottle-neck in thermionic emission, in series with diffusion. The analysis can be greatly simplified by assuming a graded junction where ΔE_C and ΔE_V become smooth transitions inside the depletion region. With this assumption, the diffusion currents are similar to a regular p-n junction but with the appropriate parameters in place. The electron and hole diffusion currents are:

$$J_n = \frac{qD_{n2}n_{i2}^2}{L_{n2}N_{A2}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right],$$ (146a)

$$J_p = \frac{qD_{p1}n_{i1}^2}{L_{p1}N_{D1}}\left[\exp\left(\frac{qV}{kT}\right) - 1\right].$$ (146b)

Note that the band offsets ΔE_C and ΔE_V are not in these equations, and also that each diffusion current component depends on the properties of the receiving side only, as in the case of a homojunction. The total current becomes

$$J = J_n + J_p = \left(\frac{qD_{n2}n_{i2}^2}{L_{n2}N_{A2}} + \frac{qD_{p1}n_{i1}^2}{L_{p1}N_{D1}}\right)\left[\exp\left(\frac{qV}{kT}\right) - 1\right].$$ (147)

Of particular interest is the ratio of the two diffusion currents.

$$\frac{J_n}{J_p} = \frac{L_{p1}D_{n2}N_{D1}n_{i2}^2}{L_{n2}D_{p1}N_{A2}n_{i1}^2} = \frac{L_{p1}D_{n2}N_{D1}N_{C2}N_{V2}\exp(-E_{g2}/kT)}{L_{n2}D_{p1}N_{A2}N_{C1}N_{V1}\exp(-E_{g1}/kT)}$$

$$\approx \frac{N_{D1}}{N_{A2}}\exp\left(\frac{-\Delta E_g}{kT}\right) .$$ (148)

Therefore the injection ratio depends exponentially on the bandgap difference, in addition to their doping ratio. This is critical in designing a bipolar transistor where the injection ratio is directly related to the current gain. The heterojunction bipolar transistor (HBT) uses a wide-bandgap emitter to suppress the base current and will be discussed in more details in Chapter 5.

2.7.2 Isotype Heterojunction

The case of an isotype heterojunction is somewhat different. In an n-n heterojunction, since the work function of the wide-bandgap semiconductor is smaller, the energy bands will be bent oppositely to those for the n-p case (Fig. 28a).[42] The relation between ($\psi_{b1} - V_1$) and ($\psi_{b2} - V_2$) can be found from the boundary condition of continuity of electric displacement ($\mathscr{D} = \varepsilon_s\mathscr{E}$) at the interface. For an accumulation (increase of carriers at the interface) in Region-1 governed by Boltzmann statistics, the electric field at x_0 is given by (for detailed derivation see footnote on p. 84)

$$\mathscr{E}_1(x_0) = \sqrt{\frac{2qN_{D1}}{\varepsilon_{s1}}\left\{\frac{kT}{q}\left[\exp\frac{q(\psi_{b1} - V_1)}{kT} - 1\right] - (\psi_{b1} - V_1)\right\}}.$$ (149)

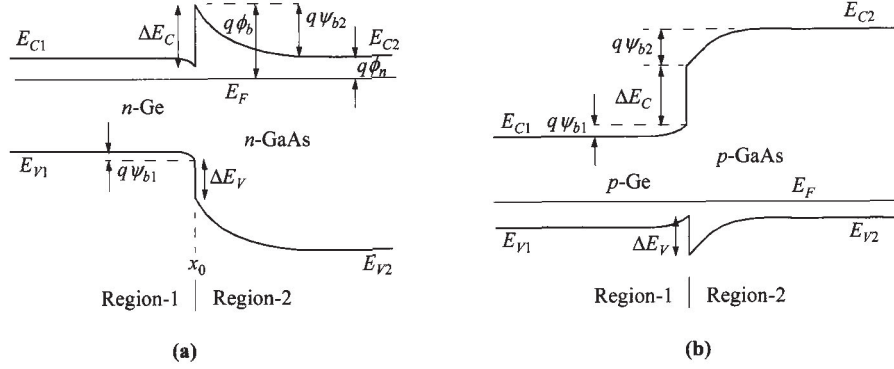The electric field at the interface for a depletion in Region-2 is given by

Fig. 28   Energy-band diagrams for ideal (a) *n-n* and (b) *p-p* isotype heterojunctions. (After Refs. 40 and 42.)

$$\mathscr{E}_2(x_0) = \sqrt{\frac{2qN_{D2}(\psi_{b2} - V_2)}{\varepsilon_{s2}}}. \tag{150}$$

Equating the electric displacement $\mathscr{D} = \mathscr{E}\varepsilon_s$ of Eqs. 149 and 150 gives a relation between $(\psi_{b1} - V_1)$ and $(\psi_{b2} - V_2)$ that is quite complicated. However, if the ratio $\varepsilon_{s1}N_{D1}/\varepsilon_{s2}N_{D2}$ is of the order of unity and $\psi_{bi}(\equiv \psi_{b1} + \psi_{b2}) \gg kT/q$, we obtain[42]

$$\exp\left[\frac{q(\psi_{b1} - V_1)}{kT}\right] \approx \frac{q}{kT}(\psi_{bi} - V) \tag{151}$$

where V is the total applied voltage and is equal to $(V_1 + V_2)$. Also shown in Fig. 28b is the idealized equilibrium energy-band diagram for *p-p* heterojunctions.

For the carrier transport, because of the potential barrier as shown in Fig. 28a, the conduction mechanism is governed by thermionic emission of majority carriers, electrons in this case (refer to Chapter 3 for details). The current density is given by[42]

$$J = qN_{D2}\sqrt{\frac{kT}{2\pi m_2^*}}\exp\left(\frac{-q\psi_{b2}}{kT}\right)\left[\exp\left(\frac{qV_2}{kT}\right) - \exp\left(\frac{-qV_1}{kT}\right)\right]. \tag{152}$$

Substituting Eq. 151 into Eq. 152 yields the current-voltage relationship:

$$J = \frac{q^2 N_{D2}\psi_{bi}}{\sqrt{2\pi m_2^* kT}}\exp\left(\frac{-q\psi_{bi}}{kT}\right)\left(1 - \frac{V}{\psi_{bi}}\right)\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \tag{153}$$

Since the current is thermionic emission as in a metal-semiconductor contact, the pre-exponential factor is often expressed in terms of the effective Richardson constant A^* and the barrier height ϕ_b. With substitution for A^* and the appropriate expression for N_{D2}, the current equation above becomes

$$J = \frac{q\,\psi_{bi}A^*T}{k}\left(1 - \frac{V}{\psi_{bi}}\right)\exp\left(\frac{-q\,\psi_{b1}}{kT}\right)\exp\left(\frac{-q\,\phi_b}{kT}\right)\left[\exp\left(\frac{qV}{kT}\right) - 1\right]$$

$$= J_0\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \tag{154}$$

This expression is quite different from that for metal-semiconductor contact. The value of J_0 is different [from $A^*T^2\exp(-q\phi_B/kT)$] and so is its temperature dependence. The reverse current never saturates but increases linearly with voltage at large $-V$. In the forward direction, the dependence of J on V can be approximated by an exponential function $J \propto \exp(qV/\eta kT)$.
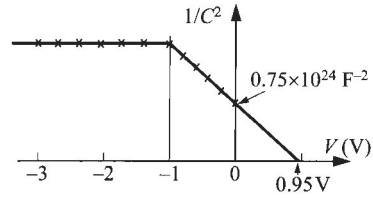
REFERENCES

1. W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," *Bell Syst. Tech. J.*, **28**, 435 (1949);

2. W. Shockley, *Electrons and Holes in Semiconductors*, D. Van Nostrand, Princeton, New Jersey, 1950.

3. 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 (1957).

4. J. L. Moll, "The Evolution of the Theory of the Current-Voltage Characteristics of p-n Junctions," *Proc. IRE*, **46**, 1076 (1958).

5. C. G. B. Garrett and W. H. Brattain, "Physical Theory of Semiconductor Surfaces," *Phys. Rev.*, **99**, 376 (1955).

6. C. Kittel and H. Kroemer, *Thermal Physics*, 2nd Ed., W. H. Freeman and Co., San Francisco, 1980.

7. W. C. Johnson and P. T. Panousis, "The Influence of Debye Length on the C-V Measurement of Doping Profiles," *IEEE Trans. Electron Devices*, **ED-18**, 965 (1971).

8. B. R. Chawla and H. K. Gummel, "Transition Region Capacitance of Diffused p-n Junctions," *IEEE Trans. Electron Devices*, **ED-18**, 178 (1971).

9. M. Shur, *Physics of Semiconductor Devices*, Prentice-Hall, Englewood Cliffs, New Jersey, 1990.

10. H. K. Gummel, "Hole-Electron Product of p-n Junctions," *Solid-State Electron.*, **10**, 209 (1967).

11. J. L. Moll, *Physics of Semiconductors*, McGraw-Hill, New York, 1964.

12. M. J. O. Strutt, *Semiconductor Devices*, Vol. 1, *Semiconductor and Semiconductor Diodes*, Academic, New York, 1966, Chapter 2.

13. P. J. Lundberg, private communication.

14. S. M. Sze and G. Gibbons, "Avalanche Breakdown Voltages of Abrupt and Linearly Graded p-n Junctions in Ge, Si, GaAs, and GaP," *Appl. Phys. Lett.*, **8**, 111 (1966).

15. R. M. Warner, Jr., "Avalanche Breakdown in Silicon Diffused Junctions," *Solid-State Electron.*, **15**, 1303 (1972).

16. M. H. Lee and S. M. Sze, "Orientation Dependence of Breakdown Voltage in GaAs," *Solid-State Electron.*, **23**, 1007 (1980).

17. F. Waldhauser, private communication.

18. S. K. Ghandhi, *Semiconductor Power Devices*, Wiley, New York, 1977.

19. C. R. Crowell and S. M. Sze, "Temperature Dependence of Avalanche Multiplication in Semiconductors," *Appl. Phys. Lett.*, **9**, 242 (1966).

20. C. Y. Chang, S. S. Chiu, and L. P. Hsu, "Temperature Dependence of Breakdown Voltage in Silicon Abrupt *p-n* Junctions," *IEEE Trans. Electron Devices*, **ED-18**, 391 (1971).

21. S. M. Sze and G. Gibbons, "Effect of Junction Curvature on Breakdown Voltages in Semiconductors," *Solid-State Electron.*, **9**, 831 (1966).

22. A. Rusu, O. Pietrareanu, and C. Bulucea, "Reversible Breakdown Voltage Collapse in Silicon Gate-Controlled Diodes," *Solid-State Electron.*, **23**, 473 (1980).

23. A. S. Grove, O. Leistiko, Jr., and W. W. Hooper, "Effect of Surface Fields on the Breakdown Voltage of Planar Silicon *p-n* Junctions," *IEEE Trans. Electron Devices*, **ED-14**, 157 (1967).

24. R. H. Kingston, "Switching Time in Junction Diodes and Junction Transistors," *Proc. IRE*, **42**, 829 (1954).

25. A. Van der Ziel, *Noise in Measurements*, Wiley, New York, 1976.

26. A. Van der Ziel and C. H. Chenette, "Noise in Solid State Devices," in *Advances in Electronics and Electron Physics*, Vol. 46, Academic, New York, 1978.

27. K. K. Ng, *Complete Guide to Semiconductor Devices*, 2nd Ed., Wiley, New York, 2002.

28. J. P. Levin, "Theory of Varistor Electronic Properties", *Crit. Rev. Solid State Sci.*, **5**, 597 (1975).

29. M. H. Norwood and E. Shatz, "Voltage Variable Capacitor Tuning—A Review," *Proc. IEEE*, **56**, 788 (1968).

30. R. A. Moline and G. F. Foxhall, "Ion-Implanted Hyperabrupt Junction Voltage Variable Capacitors," *IEEE Trans. Electron Devices*, **ED-19**, 267 (1972).

31. G. Lucovsky, R. F. Schwarz, and R. B. Emmons, "Transit-Time Considerations in *p-i-n* Diodes," *J. Appl. Phys.*, **35**, 622 (1964).

32. A. G. Milnes, *Semiconductor Devices and Integrated Electronics*, Van Nostrand, New York, 1980

33. W. Shockley, U.S. Patent 2,569,347 (1951).

34. A. I. Gubanov, *Zh. Tekh. Fiz.*, **21**, 304 (1951); *Zh. Eksp. Teor. Fiz.*, **21**, 721 (1951).

35. H. Kroemer, "Theory of a Wide-Gap Emitter for Transistors," *Proc. IRE*, **45**, 1535 (1957).

36. H. C. Casey, Jr., and M. B. Panish, *Heterostructure Lasers*, Academic, New York, 1978.

37. A. G. Milnes and D. L. Feucht, *Heterojunctions and Metal-Semiconductor Junctions*, Academic, New York, 1972.

38. B. L. Sharma and R. K. Purohit, *Semiconductor Heterojunctions*, Pergamon, London, 1974.

39. P. Bhattacharya, Ed., *III-V Quantum Wells and Superlattices*, INSPEC, London, 1996.

40. R. L. Anderson, "Experiments on Ge-GaAs Heterojunctions," *Solid-State Electron.*, **5**, 341 (1962).

41. W. R. Frensley and H. Kroemer, "Theory of the Energy-Band Lineup at an Abrupt Semiconductor Heterojunction," *Phys. Rev. B*, **16**, 2642 (1977).

42. L. L. Chang, "The Conduction Properties of Ge-GaAs$_{1-x}$P$_x$ *n-n* Heterojunctions," *Solid-State Electron.*, **8**, 721 (1965).
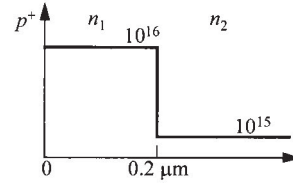
PROBLEMS

1. A silicon p-n junction of 1 cm^2 area consists of a two-sided step junction with an n-region of 10^{17} donors/cm^3 and a p-region of 2×10^{17} acceptors/cm^3. All donors and acceptors are ionized. Find the built-in potential.

2. The measured depletion capacitance of a p^+-n silicon junction (formed in an n-type epitaxial layer) is shown. The device area is 10^{-5} cm^2 and the p^+-layer thickness is 0.07 μm. Find the thickness of the epitaxial layer.
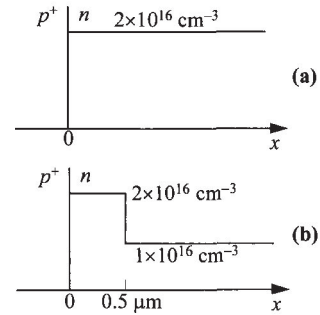
3. A silicon p-n junction has a linearly graded junction on the p-side with an impurity gradient of 10^{19} cm^{-4}, and a uniform doping of 3×10^{14} cm^{-3} on the n-side. (a) If the depletion width of the p-side is 0.8 μm at zero bias, find the total depletion width, the built-in potential and the maximum field at thermal equilibrium. (b) Plot the impurity and field distribution of this junction.

4. Find the depletion-layer width and the maximum field at thermal equilibration for the p^+-n_1-n_2 structure.

5. (a) A silicon p^+-n junction has the following parameters at 300 K: $\tau_p = \tau_g = 10^{-6}$ s, $N_D = 10^{15}$ cm^{-3}. Find the generation current density in the depletion region and the total reverse current density at a bias of 5 V.
   (b) Will there be any significant change of the total reverse current density if τ_p is reduced by a factor of 100 while τ_g remains the same?

6. A p^+-n junction is formed in an n-type substrate with $N_D = 10^{15}$ cm^{-3}. If the junction contains 10^{15} cm^{-3} generation-recombination centers located 0.02 eV above the intrinsic Fermi level of silicon with $\sigma_n = \sigma_p = 10^{-15}$ cm^2 ($v_{th} = 10^7$ cm/s), calculate the generation and recombination current at –0.5 V.

7. For a p-n junction with the p-side doped to 1×10^{17} cm^{-3}, the n-side doped to 1×10^{19} cm^{-3}, and a reverse bias of –2 V, calculate the generation-recombination current density, assuming that the effective lifetime is 1×10^{-5} s.

8. Design an abrupt Si p^+-n junction diode that has a reverse breakdown voltage of 130 V and has a forward-bias current of 2.2 mA at $V = 0.7$ volt. Assume $\tau_p = 10^{-7}$ s.

9. (a) Assume $\alpha = \alpha_0(\mathscr{E}/\mathscr{E}_0)^m$ where α_0, $\mathscr{E}_0$, and m are constants. Also assume $\alpha_n = \alpha_p = \alpha$. Derive an expression for the avalanche breakdown voltage of an n^+-p junction with a uniform acceptor concentration N_A and a dielectric permittivity ε_s.
   (b) If $\alpha_0 = 10^4$ cm^{-1}, $\mathscr{E}_0 = 4\times10^5$ V/cm, $m = 6$, $N_A = 2\times10^{16}$ cm^{-3} and $\varepsilon_s = 10^{-12}$ F/cm, what is the breakdown voltage?

10. When a silicon p^+-n junction is reverse-biased to 30 V, the depletion-layer capacitance is 1.75 nF/cm². If the maximum electric field at avalanche breakdown is 3.1×10^5 V/cm, find the breakdown voltage.

11. A silicon junction diode has a doping profile of p^+-i-n^+-i-n^+ which contains a very narrow n^+-region sandwiched between two i-regions. This narrow region has a doping of 10^{18} cm⁻³ and a width of 10 nm. The first i-region has a thickness of 0.2 μm, and the second i-region is 0.8 μm in thickness. Find the electric field in the second i-region (i.e., in the n^+-i-n^+) when a reverse bias of 20 V is applied to the junction diode.

12. For a silicon one-sided p^+-n-n^+ abrupt junction with a donor concentration of 5×10^{14} cm⁻³, the maximum field at breakdown is 3×10^5 V/cm. If the thickness of the n-type epitaxial layer is reduced to 5 μm, find the breakdown voltage.

13. For a Si p^+-n one-sided abrupt junction with $N_D = 2 \times 10^{16}$ cm⁻³, the breakdown voltage is 32 V (Fig. a). If the doping distribution is modified to Fig. b, find the breakdown voltage.

(a)

(b)

14. Find the value of the electron multiplication factor M_n for a silicon p^+-i-n^+ diode having a reverse bias of 200 V. The corresponding capacitance of the diode is 1.05 nF/cm².

15. In an "ideal" silicon n^+-p junction with $N_A = 10^{16}$ cm⁻³, a minority carrier lifetime of 10^{-8} s, and a mobility of 966 cm²/V-s, find the stored minority carriers in the neutral p-region of 1 μm, under a forward bias of 1 V.

16. For an ideal abrupt silicon p^+-n junction with $N_D = 10^{15}$ cm⁻³, find the stored minority carriers (in C/cm²) in the neutral region when a forward bias of 1 V is applied. Assume the length of the neutral region is 1 μm and the diffusion length of holes is 5 μm. The hole distribution is given by

$$p_n - p_{no} = p_{no}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]\exp\left[\frac{-(x - x_n)}{L_p}\right].$$

17. For a hyperabrupt p^+-n junction varactor, the n-side doping profile is given by $n(x) = Bx^m$ where B is a constant and $m = -3/2$. Derive the express for the differential capacitance.

18. Consider an ideal abrupt heterojunction with a built-in potential of 1.6 V. The impurity concentrations in semiconductor 1 and 2 are 1×10^{16} donors/cm³ and 3×10^{19} acceptors/cm³, and dielectric constants are 12 and 13, respectively. Find the electrostatic potential and depletion width in each material for applied voltages of 0.5 V and –5 V.

19. For an n-GaAs/p-Al$_{0.3}$Ga$_{0.7}$As heterojunctIon at room-temperature, $\Delta E_C = 0.21$ eV. (1) What type of heterojunction is this? (2) Based on the Anderson Model, find the total depletion width at thermal equilibrium when both sides have impurity concentration of 5×10^{15} cm⁻³. (3) Draw the band diagram. [*Hint*: For the bandgap of AlGaAs, refer to Fig. 32 of Chapter 1. The dielectric constant is $(12.9 - 3.12x)$ for Al$_x$Ga$_{1-x}$As. Assume N_C and N_V are the same for Al$_x$Ga$_{1-x}$As with $0 < x < 0.4$.]

20. The alignment of heterojunction between GaAs and $Al_{0.4}Ga_{0.6}As$ is Type-I. The doping concentration is 10^{20} cm^{-3} in $Al_{0.4}Ga_{0.6}As$ and 10^{16} cm^{-3} in GaAs, both doped with carbon. (a) Find the total depletion width under thermal equilibrium condition, assuming the dielectric constant is the same for both semiconductors. (b) Draw the band diagram for $V = 0$.