CHAPTER

1

p-n JUNCTION DIODE

1.1 HISTORY

The *p-n junction* was discovered by Ohl in 1940 when he observed the photovoltaic effect when light was flashed onto a silicon rod.[1,2] Since crystals were not as pure at the time, different parts of the same crystal had different impurities and a natural *p-n* junction was formed unintentionally. Ohl also notice that when a metal whisker was pressed against different parts of the crystal, opposite behaviors were observed. He coined the material *p*-type when "positive" bias was put on the crystal relative to the whisker to produce a large current, and conversely, *n*-type when "negative" bias was needed to conduct similar current. This research group at Bell Laboratories later made the connection between *p*-type to acceptor impurities and *n*-type to donor impurities. The theory for the *p-n* junction diode was developed by Shockley in 1949,[3] and it was instrumental for the invention of the bipolar junction transistor. The theory was subsequently refined by Sah et al.[4] and Moll.[5] More recent review articles on the device may be found in Refs. 6–9. The *p-n* junction has been the most common rectifier used in the electronics industry. It also serves as a very important fundamental building block for many other devices.

1.2 STRUCTURE

The early version of the structure was made by pressing a metal wire onto the surface of a semiconductor. A junction was then formed by passing a pulse of current through the wire and semiconductor. It is believed that doping is diffused from the metal wire as shown in Fig. 1.1(a). Such a structure is referred to as a *point contact* and the metal wire as a *cat's whisker*. (A point contact has the characteristics of either a *p-n* junction or a Schottky barrier, depending on the forming process; see Section 3.2.) Another old process is the alloy method, in which a metal containing the appropriate impurity is placed onto the semiconductor surface. Heating above the eutectic temperature would form an alloy with a thin heavily doped region at the interface. This technique, along with the point contact, is no longer used. A planar structure is shown in Fig. 1.1(b). The surface doping is usually introduced by ion implantation. Diffusion at high temperature can also be used, and the impurity source can be in a carrying gas or deposited material. Another common technique is to incorporate doping during epitaxial growth. The area of the diode is usually defined by an opening in an insulator layer during implantation or diffusion.

1.3 CHARACTERISTICS

A *p-n* junction can be viewed as isolated *p-* and *n*-type materials brought into intimate contact (Fig. 1.2). Being abundant in *n*-type material, electrons diffuse to the *p*-type material. The same process happens for holes from the *p*-type material. This flow of charges sets up an electric field that starts to hinder further diffusion until an equilibrium is struck. The energy-band diagram under equilibrium is shown in Fig. 1.2(b). (Notice that when $N_A \neq N_D$, where E_i crosses E_F does not coincide with the metallurgical junction.) Since the overall charge has to be conserved, it follows that for an abrupt (step) junction,

$$W_{dp}N_A = W_{dn}N_D \tag{1.1}$$

as shown in Fig. 1.2(c). An important parameter is the built-in potential ψ_{bi}. According to Fig. 1.2(b), it is the sum of ψ_{Bn} and ψ_{Bp}, given by

$$\psi_{bi} = \psi_{Bn} + \psi_{Bp} = \frac{kT}{q}\ln\left(\frac{N_D N_A}{n_i^2}\right) \tag{1.2}$$

which is the total band bending at equilibrium by definition.

Under bias, the following can be obtained using the Poisson equation with appropriate boundary conditions,

$$W_{dn} = \sqrt{\frac{2\varepsilon_s\psi_n}{qN_D}} \qquad W_{dp} = \sqrt{\frac{2\varepsilon_s\psi_p}{qN_A}} \tag{1.3}$$

$$|\mathscr{E}_m| = \frac{qN_A W_{dp}}{\varepsilon_s} = \frac{qN_D W_{dn}}{\varepsilon_s} \tag{1.4}$$

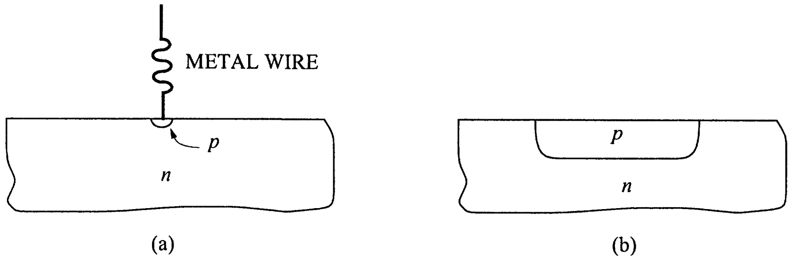

FIGURE 1.1
Cross-section structure of a p-n junction as in (a) point contact and (b) planar technology.

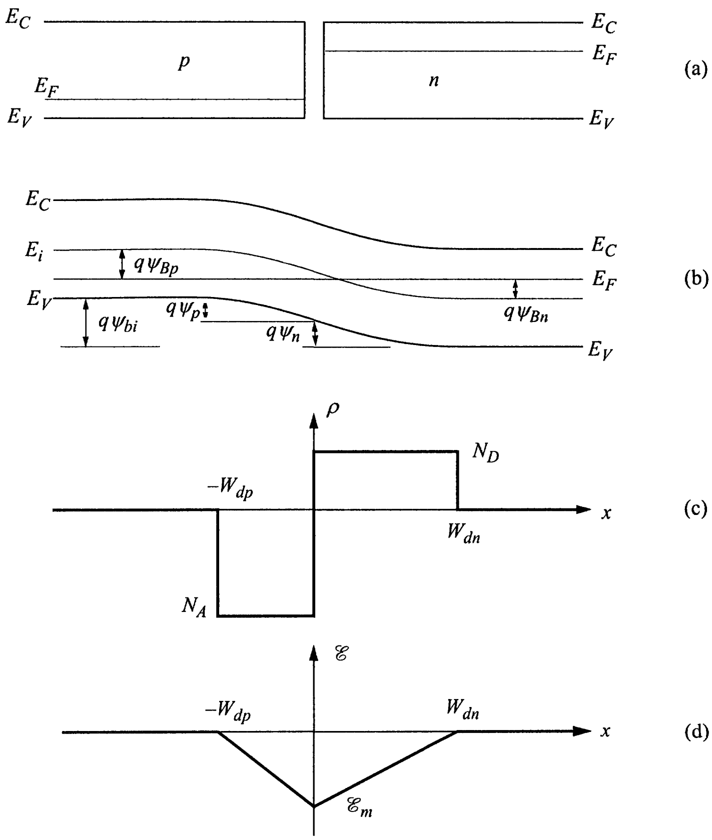

FIGURE 1.2
Formation of a p-n junction by bringing (a) isolated materials into (b) intimate contact. The potential variation is a result of (c) charge distribution or (d) field distribution in the depletion layer.

$$\psi_T = \psi_p + \psi_n = (\psi_{bi} - V_f) \text{ or } (\psi_{bi} + V_r) = \frac{1}{2}\mathscr{E}_m(W_{dn} + W_{dp}) \quad . \tag{1.5}$$

Equation (1.5) can be interpreted as the area under the field-distance curve in Fig. 1.2(d). The partition of band bending and depletion width between the n- and p-regions can be related by

$$\begin{aligned} \frac{W_{dn}}{W_{dn} + W_{dp}} &= \frac{\psi_n}{\psi_T} = \frac{N_A}{N_A + N_D} \\ \frac{W_{dp}}{W_{dn} + W_{dp}} &= \frac{\psi_p}{\psi_T} = \frac{N_D}{N_A + N_D} \quad . \end{aligned} \tag{1.6}$$

It can further be shown that

$$W_{dp} + W_{dn} = \sqrt{\frac{2\varepsilon_s}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)\psi_T} \quad . \tag{1.7}$$

In practical devices, one side usually has a doping concentration much higher than the other, and the junction can be treated as a one-sided junction. The depletion width and potential variation in the heavily doped side can then be neglected.

Figure 1.3, which shows the energy-band diagram and the carrier concentrations under bias, is used to derive the I-V characteristics. The forward current of a p-n junction under bias is determined by diffusion of injected minority carriers. The carrier concentration at the edge of the depletion region is given by

$$\begin{aligned} n_p(W_{dp}) &= n_{po} \exp\left(\frac{qV_f}{kT}\right) \\ p_n(W_{dn}) &= p_{no} \exp\left(\frac{qV_f}{kT}\right) \quad . \end{aligned} \tag{1.8}$$

Combining the continuity equation with the current equation, assuming steady state, zero generation rate, and zero drift current, one gets

$$\begin{aligned} D_n \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{po}}{\tau_n} &= 0 \\ D_p \frac{d^2 p_n}{dx^2} - \frac{p_n - p_{no}}{\tau_p} &= 0 \end{aligned} \tag{1.9}$$

where $x = 0$ now corresponds to the edge of the depletion region. (Notice the x-coordinate in Fig. 1.3(c).) Solving these differential equations gives the minority-carrier profiles

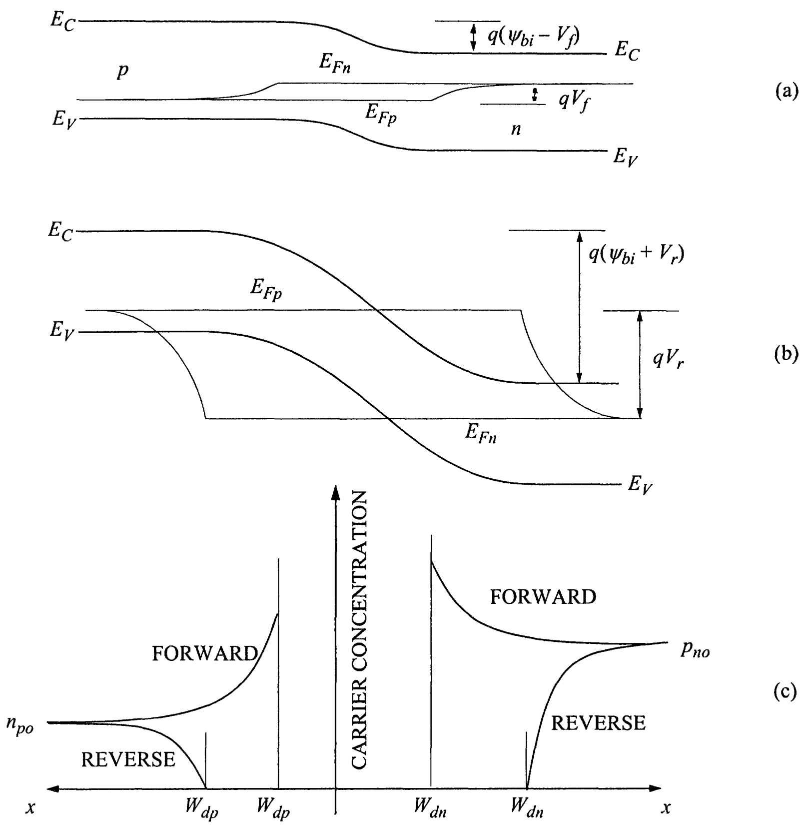

FIGURE 1.3
Energy-band diagram showing a *p-n* junction (a) under forward bias (positive voltage applied to *p*-type material and (b) under reverse bias. (c) Minority-carrier concentration profiles under forward and reverse bias.

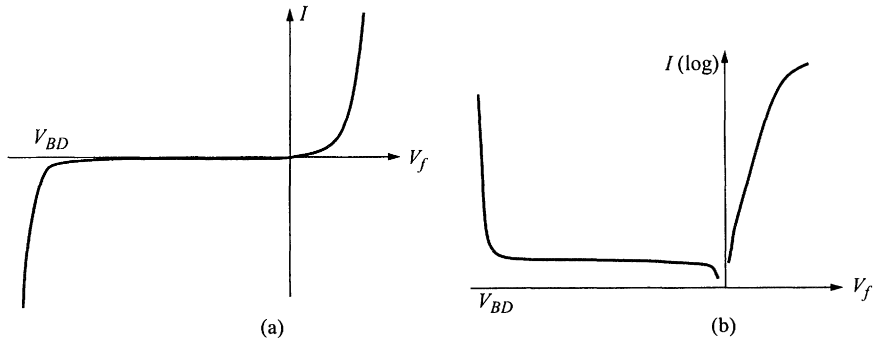

FIGURE 1.4
I-V characteristics of a *p-n* junction in (a) linear current scale and (b) logarithmic current scale.

$$n_p(x) = n_{po} + n_{po}\left[\exp\left(\frac{qV_f}{kT}\right) - 1\right]\exp\left(\frac{-x}{L_n}\right) \quad , \tag{1.10}$$

$$p_n(x) = p_{no} + p_{no}\left[\exp\left(\frac{qV_f}{kT}\right) - 1\right]\exp\left(\frac{-x}{L_p}\right) \quad . \tag{1.11}$$

The two diffusion currents give a total of

$$\begin{aligned} J &= q\left(\frac{D_p p_{no}}{L_p} + \frac{D_n n_{po}}{L_n}\right)\left[\exp\left(\frac{qV_f}{kT}\right) - 1\right] \\ &= qn_i^2\left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A}\right)\left[\exp\left(\frac{qV_f}{kT}\right) - 1\right] \quad . \end{aligned} \tag{1.12}$$

It is interesting to note that each diffusion component is controlled by the doping level in the opposite side of the junction. For example, the electron diffusion from the n-side is determined by the acceptor concentration (N_A) of the p-side, and is independent of its own doping level (N_D). At each side of the junction the diffusion current is a function of distance. It maximizes at $x = 0$, where Eq. (1.12) is obtained. Since the current has to be continuous, the diffusion current is supplemented by the majority-carrier drift current. This equation is also valid for reverse bias when V_f is negative. In cases where the thickness of the p- or n-type material is less than the diffusion length L_p or L_n, the latter parameter should be replaced by the corresponding thickness in Eq. (1.12), thereby increasing the current.

The I-V characteristics described by Eq. (1.12) is shown in Fig. 1.4. In both the linear current scale and the logarithmic current scale, additional features at high forward bias and reversed bias are to be noticed. In the forward direction, current rises exponentially with V_f until the slope becomes more gradual. This can be due to high-level injection of carriers such that the applied voltage is no longer totally developed across the depletion region. Series resistance, R_s, can also cause the same effect. At high reverse bias, breakdown can occur due to impact ionization (see Appendix B6) or Zener tunneling. These mechanisms can be separated by temperature dependence. At higher temperature, the ionization rate decreases and the breakdown voltage due to avalanche multiplication increases. The opposite dependence holds for Zener breakdown. Normally, avalanche multiplication occurs first with a breakdown voltage, to be shown later.

An additional current component besides Eq. (1.12) is due to recombination/generation through midgap states within the depletion region (see Appendix B5). This mechanism gives rise to a current described by

$$J = \frac{qn_i W_d}{2\tau}\left[\exp\left(\frac{qV_f}{2kT}\right) - 1\right] \quad . \tag{1.13}$$

If the term $qn_iW_d/2\tau$ is comparable to or larger than the preexponential factor in Eq. (1.12), the current for small V_f as well as the reverse current will be increased. Notice that the forward recombination current has a slope (in the $\ln(I)$–V curve) half that of diffusion current (see Fig. B5.2(a)).

A common use of the p-n junction requires it to switch between the on-state and the off-state. Because of minority-carrier storage under forward bias, the immediate response to reverse bias is shown in Fig. 1.5, with $I_r = V_r/R_s$,

$$t_d \approx \tau \ln\left(1 + \frac{I_f}{I_r}\right) , \tag{1.14}$$

$$\operatorname{erf}\sqrt{\frac{t_{tr}}{\tau}} + \frac{\exp(-t_{tr}/\tau)}{\sqrt{\pi t_{tr}/\tau}} = 1 + 0.1\left(\frac{I_r}{I_f}\right) . \tag{1.15}$$

This reverse recovery limits a p-n junction to about 1-GHz operation. In order to increase the frequency response, the carrier lifetime τ can be intentionally shortened by introducing impurities for recombination. The penalty for this is increased leakage current. An alternative approach is to use a step-recovery diode (Section 1.5.2).

The equivalent circuit for a p-n junction is shown in Fig. 1.6. Since capacitance is defined by dQ/dV, the depletion-layer capacitance C_d is associated with the depletion-layer charge, while the diffusion capacitance C_D is related to injected carriers. The C_D is significant only under forward-bias conditions and is proportional to the forward current, given by

$$C_D = \frac{q^2}{2kT}(L_p p_{no} + L_n n_{po})\exp\left(\frac{qV_f}{kT}\right) . \tag{1.16}$$

The C_d is determined by the depletion width, and for a one-sided step junction,

$$C_d = \frac{\varepsilon_s}{W_d} = \sqrt{\frac{q\varepsilon_s N}{2(\psi_{bi} + V_r)}} \tag{1.17}$$

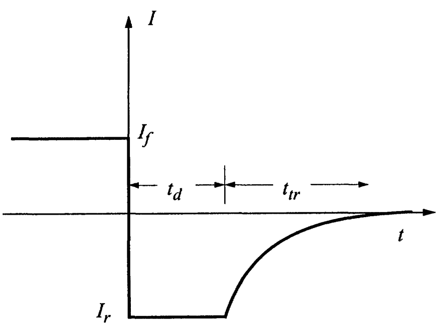

FIGURE 1.5
Transient current characteristics of a p-n junction when switched from forward to reverse direction. t_d and t_{tr} are called delay time and transition time, respectively.

where N is from the lightly doped side. A measurement of $1/C^2$ vs. V_r, as shown in Fig. 1.7, can extrapolate ψ_{bi}, and its slope can determine the doping concentration (or area). This technique can be extended to obtain a nonuniform doping profile,

$$\frac{dC_d^{-2}}{dV_r} = \frac{2}{q\varepsilon_s N(x)} \ . \tag{1.18}$$

Finally, the breakdown voltage of an one-sided step p-n junction can be evaluated based on Eq. (B6.15). However, since the ionization coefficient is a function of the electric field which varies inside the depletion region, the breakdown voltage is not straightforward to calculate and the values are shown in Fig. 1.8 for different materials. They can be shown to fit an empirical formula[10]

$$V_{BD} \approx 60\left[\frac{E_g(\text{in eV})}{1.1}\right]^{3/2}\left[\frac{N(\text{in cm}^{-3})}{10^{16}}\right]^{-3/4} \tag{1.19}$$

where N is the concentration in the lightly doped side.

1.4 APPLICATIONS

1. Because it is the most common rectifier, a p-n junction has many circuit applications (see Appendix C1).
2. Many devices are special forms of p-n junction. Examples are LED, laser, solar cell, and photodiode. A p-n junction also serves as a building block for many other devices, such as the bipolar transistor, MOSFET, junction FET, etc.

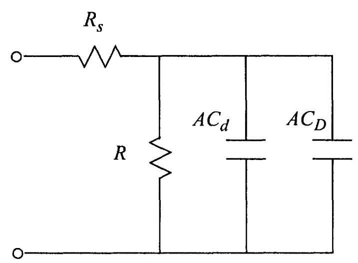

FIGURE 1.6
Equivalent circuit of a p-n junction. A is the area of the diode.

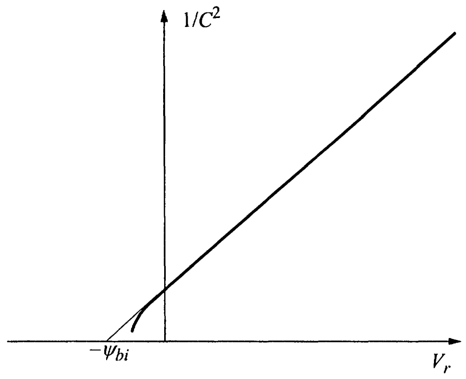

FIGURE 1.7
A plot of capacitance ($1/C^2$) under reverse bias yields ψ_{bi} and doping concentration (or area).

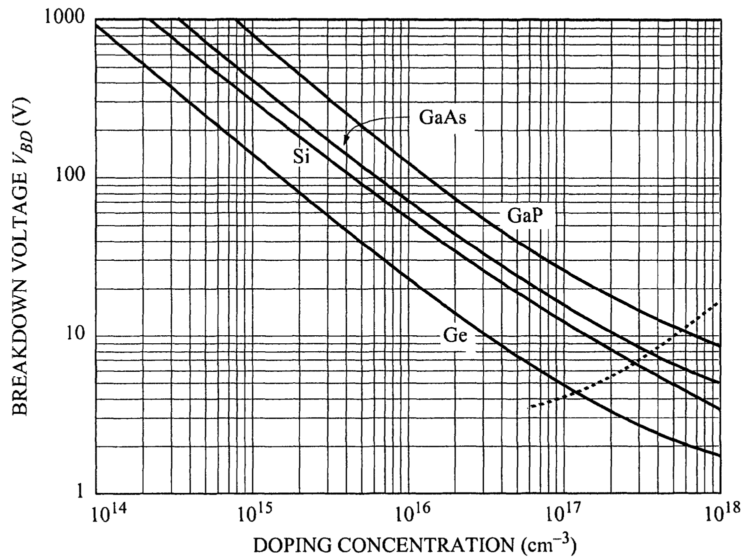

FIGURE 1.8
Breakdown voltage of abrupt one-sided *p-n* junctions for various materials. The dotted line indicates the onset of tunneling due to high doping. (After Ref. 10)

3. Due to the nonlinear, exponential nature of the current, the *p-n* junction can be used as a varistor.
4. The variable depletion capacitance at reverse bias can be utilized as a varactor.
5. A *p-n* junction is a very common protection device for electrostatic discharge (ESD). It discharges a voltage surge when it exceeds a certain value comparable to the built-in potential.
6. A *p-n* junction is a robust device and is a good choice for a diode required in power electronics.
7. The *p-n* junction can be used to isolate devices or regions of semiconductors. An example may be found in the tub isolation for CMOS circuits.
8. The well-behaved forward characteristics of a *p-n* diode enable it to be used as a temperature sensor. In operation, a constant current is applied and the voltage is monitored. This forward voltage drop is a fairly linear function of temperature. GaAs diodes can be good sensors in a wide temperature range, from a few Kelvin to ≈ 400 K, and Si diodes from ≈ 20 K.

1.5 RELATED DEVICES

1.5.1 Zener Diode

A *Zener diode* has a well-controlled breakdown voltage, called *Zener voltage*, and sharp breakdown characteristics in the reverse-bias region. In spite of the name, the breakdown can be due to either impact ionization or Zener tunneling. Zener breakdown is caused by quantum-mechanical tunneling of carriers between the conduction band and the valence band (see Appendix B8). It occurs in junctions with higher doping concentrations, and the critical field required is approximately 1 MV/cm. A Zener diode is generally used to establish a fixed reference voltage.

1.5.2 Step-Recovery Diode

The *step-recovery diode* is sometimes called a *fast-recovery diode*, *snap-off diode*, or *snap-back diode*. The response of a standard *p-n* junction is limited by the minority-carrier storage, with the reverse recovery represented by Fig. 1.5. A step-recovery diode has a special doping profile such that the field confines the injected carriers much closer to the vicinity of the junction. This results in a much shorter transition time t_{tr} (but with the same delay time t_d). The sharp turnoff of current approaches a square waveform which contains rich harmonics and is often used in applications of harmonic generation and pulse shaping.

1.5.3 Anisotype Heterojunction

An *anisotype heterojunction* is a junction not only of opposite types, but also of different semiconductor materials. The structure requires good lattice match between the two materials, and Ge/GaAs can be used as an example.[11] The distinct features are the discontinuity in the conduction band ΔE_C and valence band ΔE_V, as shown in Fig. 1.9. These values can be determined graphically to be

$$\Delta E_C = q(\chi_1 - \chi_2) \tag{1.20}$$

$$\Delta E_V = (E_{g2} - E_{g1}) - \Delta E_C \ . \tag{1.21}$$

The static characteristics described by Eqs. (1.1)–(1.7) have to be modified by the two dielectric constants K_1 and K_2 in the two materials. Specifically, $K_1 \mathscr{E}_1 = K_2 \mathscr{E}_2$ has to be satisfied at the interface. The potential variation across the *n*- and *p*-type materials are given by

$$\frac{\psi_n}{\psi_T} = \frac{K_2 N_A}{K_1 N_D + K_2 N_A} \qquad \frac{\psi_p}{\psi_T} = \frac{K_1 N_D}{K_1 N_D + K_2 N_A} \ . \tag{1.22}$$

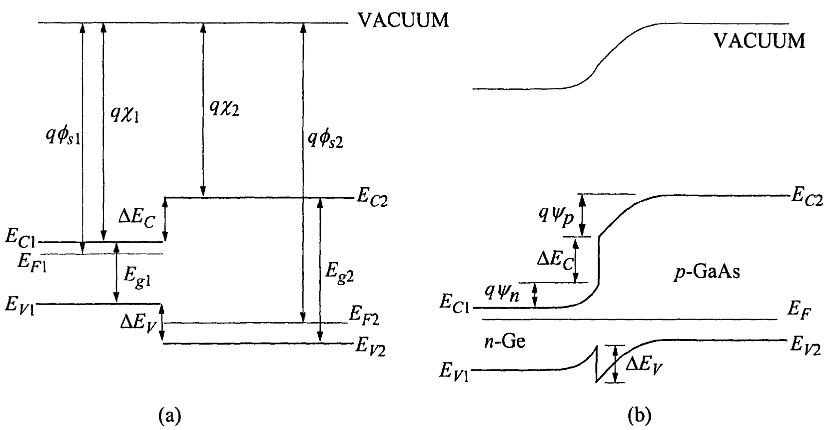

FIGURE 1.9
(a) Energy-band diagram of two isolated different semiconductor materials with opposite types. (b) Example of an anisotype heterojunction between *n*-type Ge and *p*-type GaAs. (After Ref. 11)

According to Fig. 1.9(b), ψ_T at equilibrium is the difference between the two work functions, $q\phi_{s2} - q\phi_{s1}$.

The current conduction, however, can be either diffusion limited or thermionic emission limited. In the example shown in Fig. 1.9, the barrier for holes is similar to a standard *p-n* junction, and hole transport from GaAs to Ge is diffusion limited. Under forward bias, this component is similar to a homojunction:

$$J_p = \frac{q n_i^2 D_p}{L_p N_D}\left[\exp\left(\frac{qV_f}{kT}\right) - 1\right] . \tag{1.23}$$

The barrier for electrons is increased by ΔE_C and the current is greatly reduced. Unlike a homostructure *p-n* junction in which current is dominated by the diffusion component in the lightly doped side, an anisotype heterojunction usually favors the injection of carriers from the material of larger energy gap. Other current components are due to tunneling and recombination arising from a nonideal interface. The suppression of one type of carriers improves the injection efficiency, which makes it beneficial for the emitter-base junction of a bipolar transistor.[12] Other applications include photodetectors in which a local absorption coefficient can be optimized.

1.5.4 Varactor

The word *varactor* comes from *variable reactor*. A varactor, also called a *varactor diode* or *varicap* (variable capacitance) *diode*, is in principle any two-terminal device whose capacitance varies with the dc bias. A *p-n* junction is the simplest structure. A Schottky-barrier diode can perform the same function and is used especially in ultrahigh-speed operations.

When a *p-n* junction is under reverse bias, the depletion layer widens and its capacitance changes according to Eq. (1.17). Forward bias is to be avoided from excessive current which is undesirable for any capacitor. The dependence of capacitance on the dc reverse bias is determined by the doping profile near the junction. It can be described by the form

$$C = C_1(\psi_{bi} + V_r)^{-s} \quad . \tag{1.24}$$

For a one-sided junction, if the profile of the lighter doping is approximated by

$$N(x) = C_2 x^m \quad , \tag{1.25}$$

it can be shown that [8]

$$s = \frac{1}{m+2} \quad . \tag{1.26}$$

For a one-sided step profile, $m = 0$ and $s = 1/2$. For a linearly graded junction, $m = 1$ and $s = 1/3$. If $m < 0$, the junction is said to be hyperabrupt. Specific cases of interest are $m = -1, -3/2, -5/3$ and $s = 1, 2, 3$, respectively.

The applications of a varactor are in filters, oscillators, tuning circuits of radio and TV receivers, parametric amplifiers, and automatic frequency control circuits.

REFERENCES

1. R. S. Ohl, "Light-sensitive electric device," U.S. Patent 2,402,662. Filed May 27, 1941. Granted June 25, 1946.
2. M. Riordan and L. Hoddeson, "The origins of the pn junction," *IEEE Spectrum*, **34**, 46 (1997).
3. W. Shockley, "The theory of *p-n* junctions in semiconductors and *p-n* junction transistors," *Bell Syst. Tech. J.*, **28**, 435 (1949).
4. C. T. Sah, R. N. Noyce and W. Shockley, "Carrier generation and recombination in *p-n* junctions and *p-n* junction characteristics," *Proc. IRE*, **45**, 1228 (1957).
5. J. L. Moll, "The evolution of the theory for the voltage-current characteristic of *p-n* junctions," *Proc. IRE*, **46**, 1076 (1958).
6. A. Nussbaum, "The theory of semiconducting junctions," in R. K. Willardson and A. C. Beer, Eds., *Semiconductors and semimetals*, Vol. 15, p. 39, Academic Press, New York, 1981.
7. M. P. Shaw, "Properties of junctions and barriers," in C. Hilsum, Vol. Ed., T. S. Moss, Ser. Ed., *Handbook on semiconductors*, Vol. 4, p. 1, North-Holland, Amsterdam, 1981.
8. S. M. Sze, *Physics of semiconductor devices*, 2nd Ed., Wiley, New York, 1981.
9. E. S. Yang, *Microelectronic devices*, McGraw-Hill, New York, 1988.

10. 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).
11. R. L. Anderson, "Experiments on Ge–GaAs heterojunction," *Solid-State Electron.*, **5**, 341 (1962).
12. H. Kroemer, "Theory of a wide-gap emitter for transistors," *Proc. IRE*, **45**, 1535 (1957).