II. Metal–Semiconductor Junctions

G. E. Ponchak

The earliest solid-state device was reported in 1874. It consisted of a wire tip pressed into a lead-sulfide crystal. This simple metal–semiconductor junction was the first solid-state device and became known as a whisker contact rectifier. Although whisker contact rectifiers are rarely used anymore, the metal–semiconductor junction is the most important solid-state component in microwave integrated circuits. A few examples of circuit elements that include metal–semiconductor junctions are Schottky diodes, varactor diodes, metal–semiconductor field-effect transistors (MESFETs), high-electron-mobility transistors (HEMTs), and heterojunction bipolar transistors (HBTs).

Using modern semiconductor fabrication processes, the metal–semiconductor junction is very easy to create. Metal is selectively deposited onto an n-GaAs region and an alloying bake is performed if it is required. In other words, fabrication of this junction requires only one mask level and possibly a bake. Besides its ease of fabrication, the junction is very versatile. By varying the type of metal or the semiconductor doping level, the junction can be made into a rectifying or a nonrectifying junction. Rectifying junctions preferentially permit current to flow in one direction versus the other. For example, electrons may flow easier from the metal into the semiconductor than the opposite. Therefore, a rectifying junction acts as a gate keeper to stop current from flowing in the reverse direction. The rectifying junction is commonly called a Schottky contact or a Schottky barrier junction. The nonrectifying junction or ohmic contact permits current to flow across the junction in both directions with very low resistance.

Metal–semiconductor junctions represent the essential and basic building blocks of GaAs-based devices. Therefore, it is essential to get an understanding of the metal–semiconductor junction structure and operation, and the reliability issues related to them. It will become clear throughout this text that a large volume of data has been collected on the reliability issues and failure mechanisms related to metal–semiconductor junctions. This section will introduce the reader to the metal–semiconductor junction and its characteristics, and it will present an introduction to the related failure mechanisms and reliability concerns. Chapter 4 will provide a more detailed discussion of metal–semiconductor-related failure mechanisms.

A. Junction Physics

Figure 3-6 shows a schematic of a metal–semiconductor junction formed on an n-type GaAs substrate with an external bias supply connected to the metal. Although the schematic is simple, it is also an accurate representation of the junction. To understand the junction dynamics, it is necessary to examine the energy-band diagram of the junction. It helps to first study the energy band diagram for a metal and an n-type semiconductor separated from each other such that neither material is influenced by the other. Figure 3-7(a) shows such a case. As discussed in Section 3-I, a finite number of electrons exist in the conduction band of the semiconductor, and the number of these free electrons is dependent on the temperature and doping concentration or purity of the material. Likewise, there are a number of free electrons in the metal, and the number of free electrons is dependent on the metal and the temperature. The only new parameter introduced in Figure 3-7(a) is the metal work function, ϕ_m. The work function is the energy required to remove an electron from the Fermi level of the metal to a vacuum potential. Most of the metals commonly used in GaAs circuits and devices have work functions between 4 and 5.5 eV.

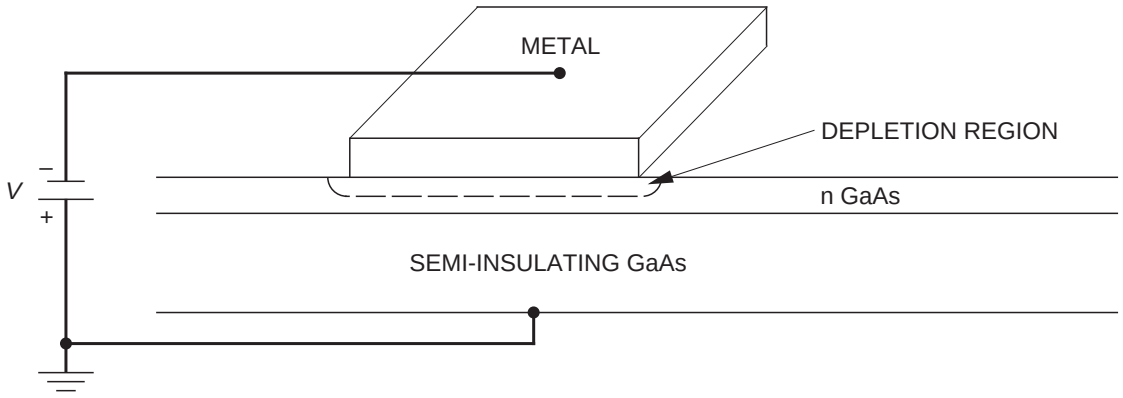

Figure 3-6. Schematic and cross section of metal–GaAs junction.

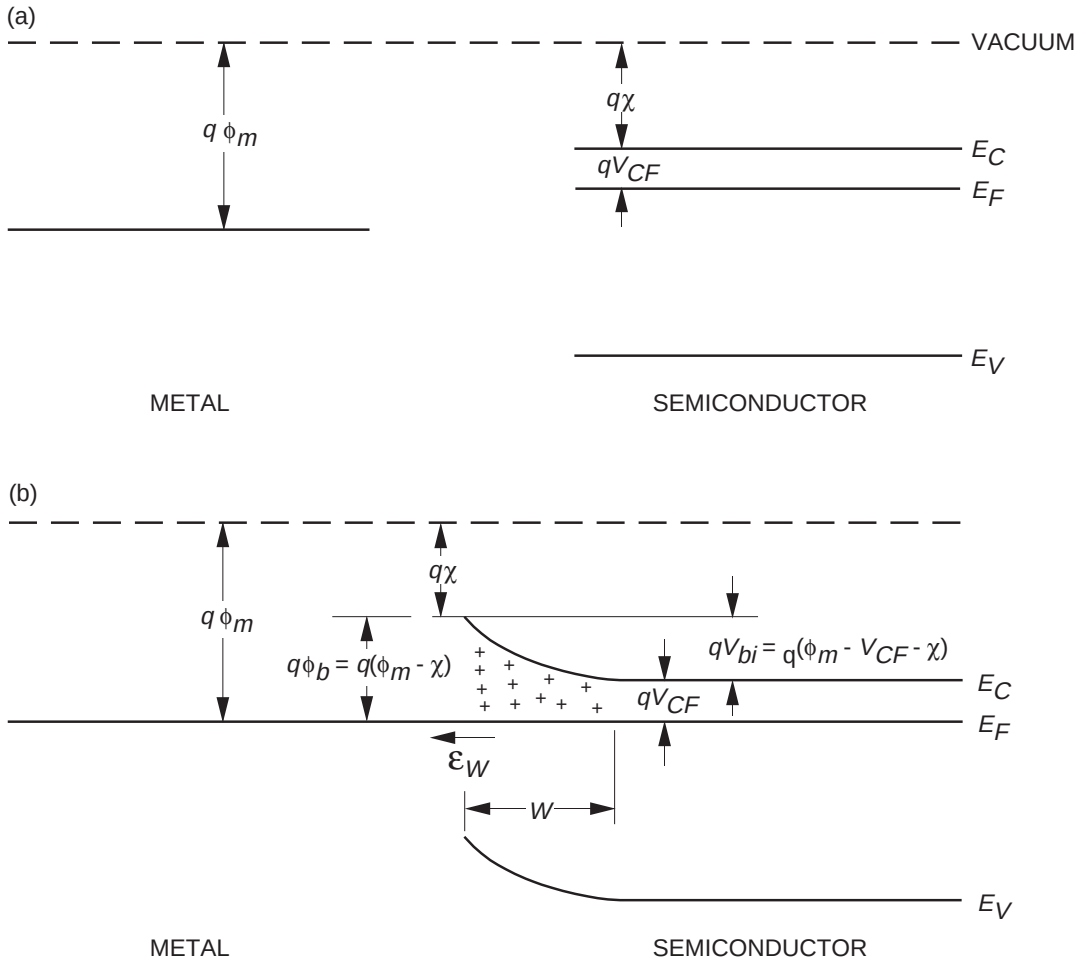

Figure 3-7. Energy band diagram of metal and semiconductor (a) separate from each other and (b) in intimate contact.

If the semiconductor Fermi level is greater than the metal Fermi level, $\chi + V_{CF} < \phi_m$, as is shown in Figure 3-7(a), then when the metal and semiconductor are put in intimate contact, electrons will diffuse from the semiconductor to the metal. As electrons

are depleted from the semiconductor, a net positive charge is created in the semiconductor at the junction. This positive charge will exert a force on the electrons that opposes the diffusion current. Equilibrium is established when these two forces are equal. Figure 3-7(b) shows the contact in equilibrium. Notice that the semiconductor energy bands bend in response to the forces just described. It is within this region, called the depletion region, that all of the junction's electrical properties are established. The amount of band bending is called the built-in potential, V_{bi} . For an electron to cross from the semiconductor to the metal, it must overcome V_{bi}, whereas an electron moving from the metal to the semiconductor must overcome the barrier potential, ϕ_b. To a first approximation, the barrier height is independent of the semiconductor properties, whereas V_{bi} is dependent on the doping level.

If an external potential is applied across the junction, the added electric field will disturb the equilibrium conditions. Consider first a positive external potential (see Figure 3-6). This will create an electric field across the junction that is opposite to the electric field caused by the depleted GaAs atoms. The result is that the diffusion current will not be sufficiently opposed, and current will flow across the junction. This is shown schematically in Figure 3-8(a). Note the reduction in the barrier for electrons flowing

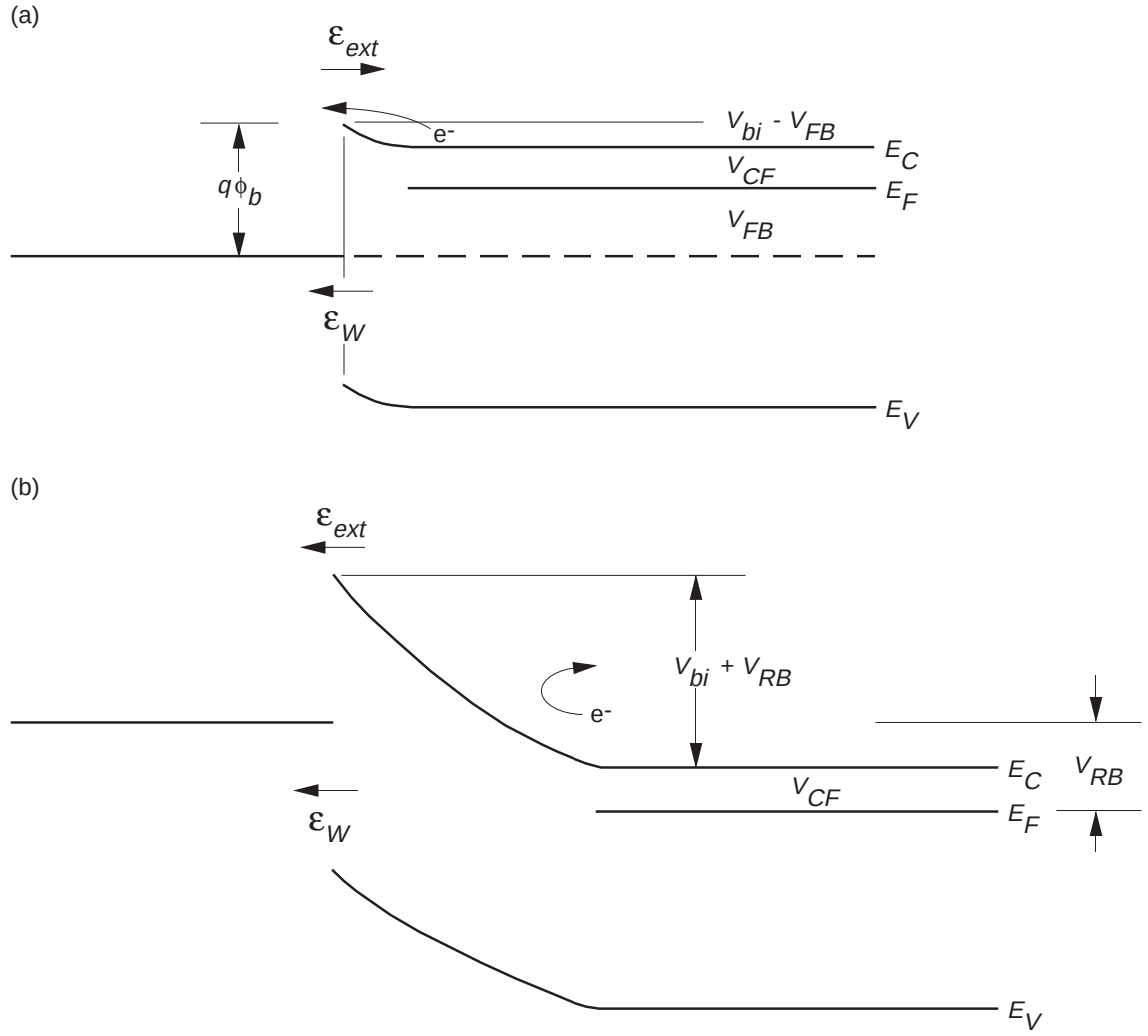

Figure 3-8. Energy band diagram of metal–semiconductor junction under (a) forward bias and (b) reverse bias.

from the semiconductor to the metal, but not for electrons flowing from the metal to the semiconductor. If a negative voltage is applied to the metal, the external field will reinforce the electric field caused by the depleted carriers, increase the band bending at the junction, and prevent the diffusion current from flowing. (See Figure 3-8(b).)

The preceding description assumed ideal material conditions. Specifically, it was assumed that the semiconductor lattice structure was uniform and perfect, even at the surface of the material. In practical cases, this is not possible. The atoms on the exposed surface do not have the required neighboring atom to complete all of the covalent bonds. Therefore, these surface atoms may either give up an electron and become a positively charged donor ion, or accept an electron and become a negatively charged acceptor ion. Surface states and their associated charge cause the energy bands of the semiconductor to bend even before the metal is introduced, as shown in Figure 3-9. Furthermore, when the metal is brought into contact with the semiconductor, the surface states may be able to accommodate all of the charge movement required to equalize the free electrons between the two materials. When this occurs, the barrier potential is no longer dependent on the metal work function. Also, no additional band bending of the semiconductor occurs because of the metal–semiconductor contact. In other words, the junction characteristics are not dependent on the metal interface. Surface states can create severe reliability problems for GaAs devices since they are generally planar devices that use only the upper few thousand angstroms of the substrate. Therefore, besides altering the built-in voltage of the contact, surface states may also provide leakage paths for current [1].

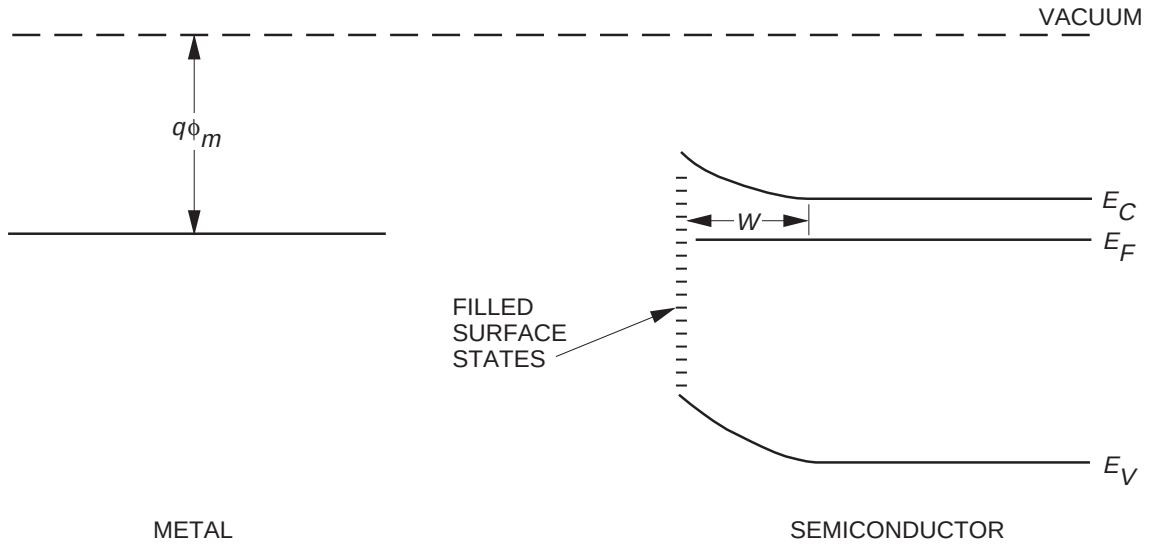

Figure 3-9. Energy band diagram of metal and semiconductor separate from each other when semiconductor surface states exist.

B. Junction Characteristics

Now that the critical parameters have been introduced, their dependence on the semiconductor and metal properties can be examined. First, consider the depletion width. Under abrupt barrier approximations, which are valid for junctions between metals and semiconductors, the width is given by

$$W = \sqrt{\frac{2\varepsilon_r \varepsilon_0}{qN_d}\left(V_{bi} - V - \frac{kT}{q}\right)} \tag{3-2}$$

where N_d is the donor doping concentration, k is Boltzmann's constant, and q is the charge of an electron. The term kT/q, often referred to as V_T, is approximately 0.026 V at room temperature whereas V_{bi} is approximately 1 V. From Equation (3-2), it is seen that the depletion width is smaller for highly doped semiconductors, and that the depletion width varies inversely with the applied bias. Based on the preceding discussion relating to Figure 3-8, it is noted that a positive bias increases current flow and decreases the depletion width. The opposite occurs for a negative bias.

Depletion widths can be quite large. As an example, consider two GaAs substrates at room temperature with an aluminum contact. Let the first have a typical MESFET channel doping of $N_d = 10^{17}/\text{cm}^2$ and the second have a typical ohmic contact doping of $N_d = 10^{19}/\text{cm}^2$. With no external bias supplied, the depletion widths for these two samples are approximately 0.048 μm and 0.006 μm, respectively. Although these appear to be very small quantities, it will become apparent throughout the rest of this chapter that these depletion widths are in fact large compared to the device dimensions required for microwave circuits.

Another critical parameter is the electric field across the depletion region. The concern is that the maximum electric field that occurs at the metal–semiconductor interface must be kept smaller than the breakdown field of GaAs, approximately 4×10^5 V/cm. If $E_m > 4 \times 10^5$ V/cm, electrons have enough kinetic energy to create electron/hole pairs during electron/atom collisions at a faster rate than the free charges can recombine. These new electrons also are accelerated by the electric field and create more electron/hole pairs. This runaway process is called “avalanche breakdown.” The result of avalanche breakdown is often a catastrophic junction failure. The maximum electric field is given by

$$E_m = \frac{qN_d}{\varepsilon_r \varepsilon_0} W = \sqrt{\frac{2qN_d}{\varepsilon_r \varepsilon_0}\left(V_{bi} - V - kT/q\right)} \tag{3-3}$$

The field is stronger when large doping concentrations are used or if a large reverse bias is applied across the junction.

The charge storage in the depletion region also creates a capacitance across the junction, which is given by

$$C = \frac{\varepsilon_r \varepsilon_0}{W} = \sqrt{\frac{q\varepsilon_r \varepsilon_0 N_d}{2\left(V_{bi} - V - \frac{kT}{q}\right)}} \tag{3-4}$$

Note two things about Equation (3-4). First, the capacitance is a function of the applied voltage. Therefore, the junction behaves as a voltage-controlled capacitance. It is this feature of the junction that is exploited in varactor diodes, which are commonly used in phase shifters and voltage controlled oscillators (VCOs). The second thing to note is that the capacitance is dependent on the doping concentration. Therefore, by varying the doping profile across the junction, the capacitance-voltage curve can be varied. Alternatively, if the doping concentration is altered during the life of the diode,

the capacitance will change, and a frequency shift in the VCO or a phase change from the phase shifter will occur.

Although an understanding of the depletion width and its associated capacitance are critical for the gate design of a field effect transistor, it is the current flow through the junction that the circuit designer is ultimately concerned with. In general, current flow through the junction is due to several mechanisms. It is necessary to examine only two of these for the purposes of this text. The first is the transport of electrons over the potential barrier, usually called thermionic emission. Thermionic emission current assumes that only electrons with energies greater than the energy of the potential barrier add to the current flow (see Figure 3-8). Several methods of analysis have been proposed to determine the current density, and although each uses different assumptions and boundary conditions, they all result in an equation of the form:

$$J = J_0 * e^{-q\phi_b/kT} * \left[e^{qV/kT} - 1\right] \tag{3-5}$$

J_0 increases with the doping concentration, N_d, and temperature. Note that J is exponentially dependent on the barrier potential, temperature, and the applied voltage. It is this strong dependence on the applied voltage that makes the junction a good rectifier. Furthermore, the dependence on temperature makes this current mechanism dominant at higher temperatures. When Schottky diodes are characterized, the measured current does not fit Equation (3-5) exactly but rather

$$J = J_0 * e^{-q\phi_b/kT} * \left[e^{qV/nkT} - 1\right] \tag{3-6}$$

where the parameter n is called the "ideality factor." An ideal diode would have $n = 1$, but for actual diodes, $n > 1$. A change in the ideality factor over the life of the diode is an indication that the metal–semiconductor interface is changing.

The second current mechanism that needs to be described is due to quantum mechanical tunneling through the potential barrier. Recall from quantum mechanics that the position of a particle is not absolute, but described by a distribution function. Therefore, although the majority of electrons will be confined by a potential barrier, there is a probability that some of the electrons will exist in the region of the potential barrier. Furthermore, if the potential barrier is thin enough, there is a probability that some electrons will travel through the barrier. This current component is referred to as tunneling current. Figure 3-10 shows the band diagrams for two cases where tunneling current is dominant. The first is a contact between a metal and a highly doped semiconductor. In this case, the depletion width, or the barrier width, is small; recall the example given earlier for the depletion width as a function of doping concentration. The second example shown in Figure 3-10(b) is an extremely reversed biased junction. The tunneling current may be given by

$$J \propto e^{\frac{-4k}{h}\sqrt{\varepsilon_r \varepsilon_0 m^*}\frac{\phi_b}{\sqrt{N_d}}} \tag{3-7}$$

which shows that the tunneling current will increase exponentially with the ratio

$$\sqrt{N_d} \,/\, \phi_b$$

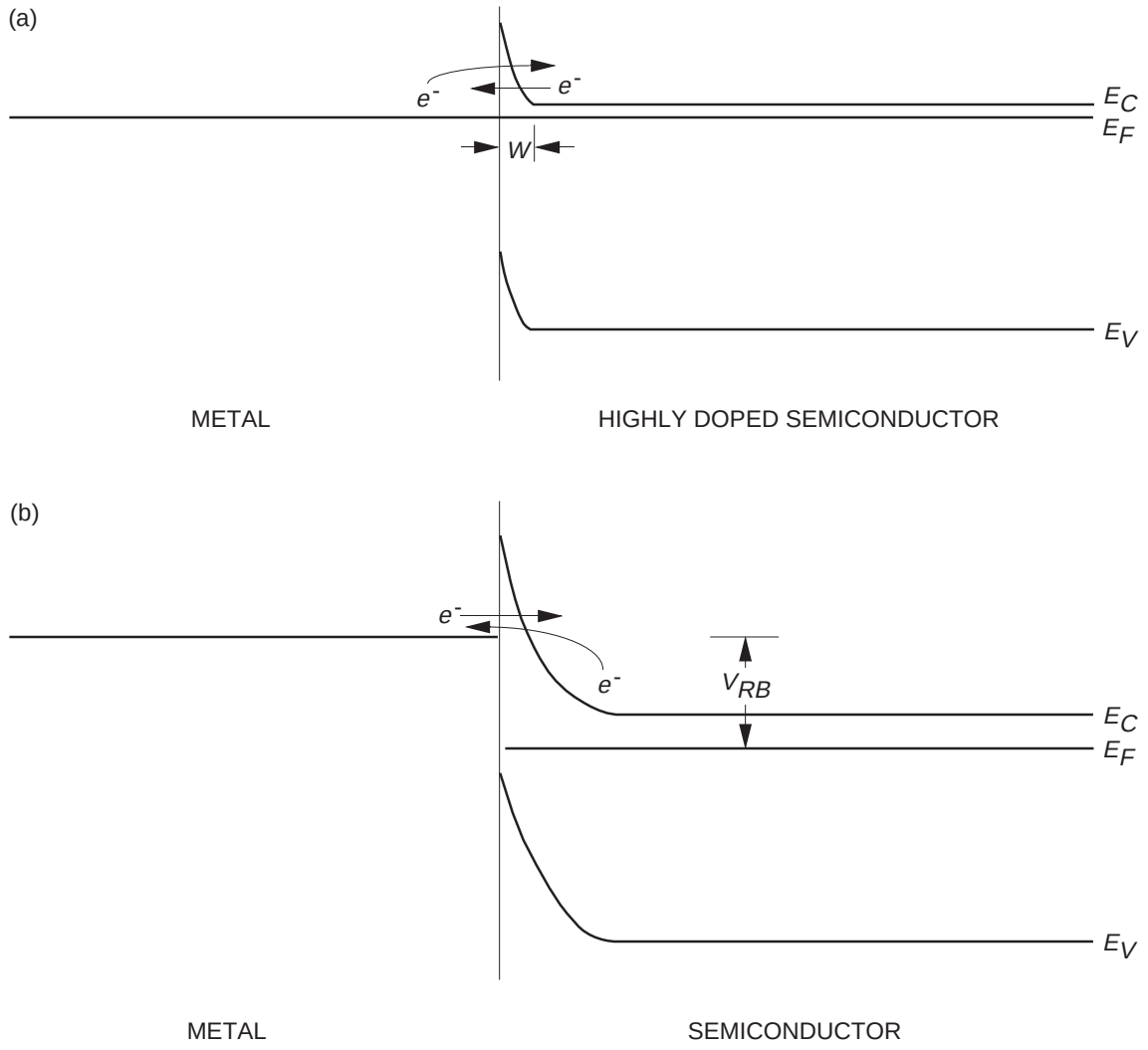

Figure 3-10. Energy band diagram of (a) metal-n^+–semiconductor junction and (b) metal–semiconductor junction under reverse bias.

For doping concentrations greater than 10^{17} cm^{-3} and for low temperatures, the tunneling current can be dominant. Since J is independent of V, this junction makes good ohmic contacts.

C. Device Structures

The practical implementation of a planar diode is shown in Figure 3-11(a). The diode is fabricated either on a molecular-beam-epitaxy- (MBE-) grown n layer or by ion implantation of an n region in the semi-insulating GaAs substrate. This is followed by the deposition of an ohmic contact metal, normally AuGe, and an ohmic contact alloying bake. Lastly, the Schottky contact metal is deposited. A simple equivalent circuit for the diode is shown in Figure 3-11(b). R_{ohm} refers to the ohmic contact junction resistance and R_{chan} refers to the resistance between the two metal contacts. Although both resistances are parasitic and ideally would be eliminated, practical limitations do not permit this.

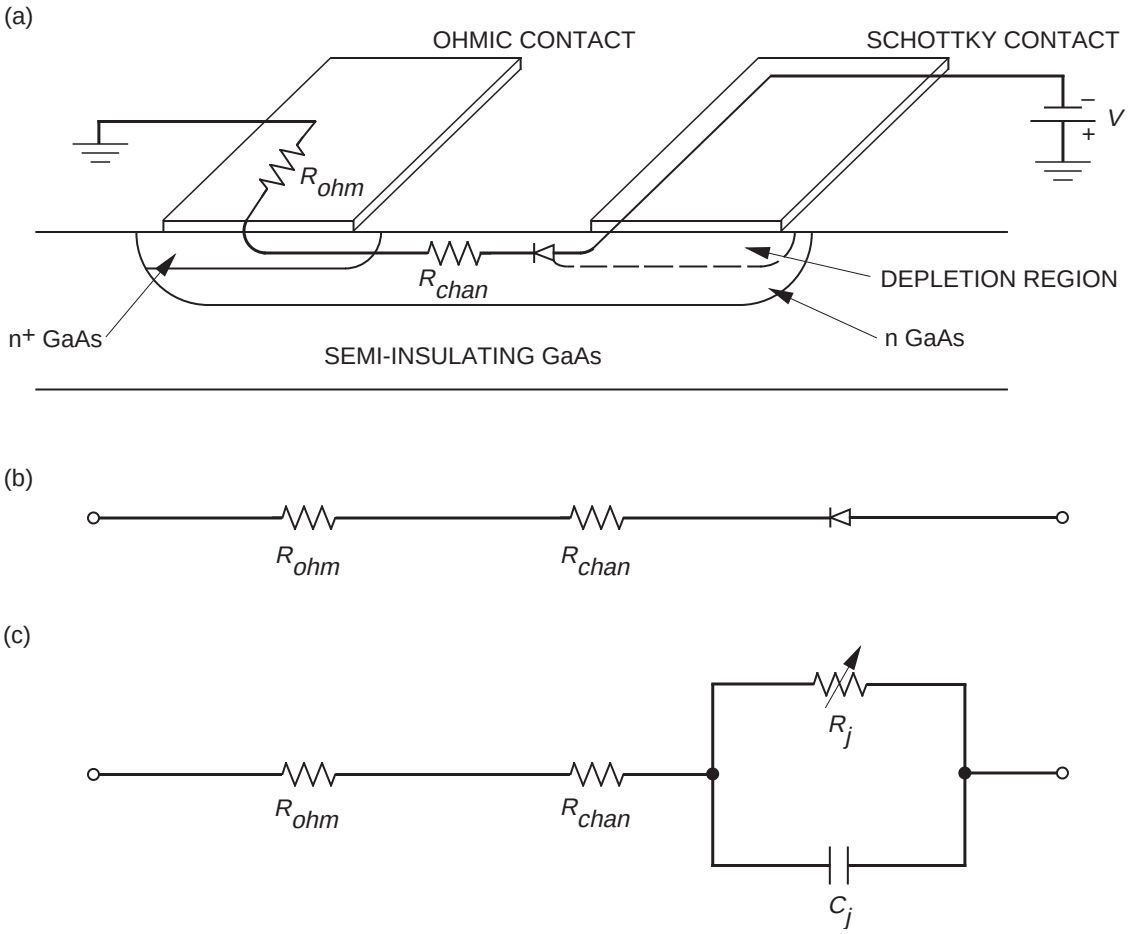

Figure 3-11. GaAs planar diode: (a) schematic, (b) simple equivalent circuit, and (c) equivalent circuit for a planar Schottky diode.

The diode electrical specifications will normally determine the doping concentration of the n region. Therefore, the ohmic contact resistance cannot be altered unless an n^+ region is formed upon which the ohmic contact can be made. R_{chan} can be reduced if the distance between the two contacts is reduced. Modern lithography permits the contacts to be separated by as little as 0.2 μm, although the contact separation is typically on the order of 1 μm. Unfortunately, the electric field between the two contacts increases as the spacing is reduced. If the electric field is increased too much because of the RF power or the dc bias, metal shorts may develop between the contacts leading to device failure. Therefore, limitations on the power handling capability of the diode are normally imposed.

Lastly, consider the diode itself. It has already been shown that the depletion region creates a capacitance called the junction capacitance, C_j. In addition, there is also a junction resistance, R_j, which is in parallel to C_j. R_j accounts for the current flow through the depletion region and can be derived from Equation (3-6) as

$$R_j = \frac{nkT}{qJA} \tag{3-8}$$

where A is the diode area. Therefore, the equivalent circuit for a planar Schottky diode is shown in Figure 3-11(c). Notice that R_j is shown as a variable resistance due to its dependence on J, which in turn is dependent on the applied voltage.

The important figure of merit for Schottky diodes is the forward current cutoff frequency:

$$f_c = \frac{1}{2\pi R_F C_F} \tag{3-9}$$

where R_F is the total series resistance and C_F is the junction capacitance at a slight forward bias. Schottky diodes have been fabricated with cutoff frequencies greater than 1 THz. In general, a diode can be used at frequencies less than $f_c/10$. Therefore, it is desirable to have a small R_j and C_j as well as a small R_{ohm} and R_{chan}.

To minimize R_j, the diode area must be increased, but to minimize C_j, the diode area must be decreased. Furthermore, a parasitic capacitance due to the fringing fields along the edges of the Schottky contact exists. This parasitic capacitance is proportional to the diode periphery and the number of corners on the contact. Since the ratio of the contact periphery to area increases as the area of the contact decreases, it is not practical to reduce C_j solely by decreasing the area.

D. Reliability

Reducing the diode area has been discussed as a method of reducing C_j. Besides the disadvantages of having a small diode area already discussed, there are other disadvantages. First, the fringing fields around the periphery of the diode will be greater and will lead to increased leakage current. Second, the fringing fields can be larger than the electric field predicted by Equation (3-3), especially around the corners of the contact. Therefore, the reverse breakdown voltage will be smaller. Third, the current through the diode is equal to $J*A$. Therefore, the current density must be increased as the area is decreased to maintain reasonable current through the device. If the current density is increased too much, failures due to electromigration may occur. Finally, the increased current density and the reduced junction area may cause the junction temperature to increase. Since GaAs is a relatively poor thermal conductor, thermal-related failure mechanisms may increase as well. To get around these problems, it is better to maximize the area and to minimize C_j by decreasing N_d. Note that reducing N_d requires n^+ regions for the ohmic contacts and increases the risk of ionic contamination failures.

The critical points to remember about the junction as it relates to device reliability are

(1) The sensitivity of its electrical characteristics to the semiconductor doping concentration.

(2) The interface barrier potential.

(3) The junction temperature.

Small changes in any of these parameters can greatly change the junction impedance and therefore the current that flows through the junction. While the circuit designer can control the junction temperature through proper packaging and heat sinking, unfortunately the barrier potential and doping concentration may change unpredictably over the life of the junction—especially at higher operating temperatures or if metal–

semiconductor interactions occur. These failure mechanisms will be fully described in Chapter 4.

Reference

[1] K. A. Christianson, "Contribution of Surface and Interface States to the Reverse Bias Aging of GaAs Schottky Barriers," 1*990 GaAs REL Workshop Digest*, Oct. 7, 1990.

Additional Reading

Goyal, R., *Monolithic Microwave Integrated Circuits: Technology & Design,* Artech House, Norwood, MA, 1989.

Sze, S. M., *Physics of Semiconductor Devices,* John Wiley & Sons, New York, 1981.

Warner, R. M., Jr., and B. L. Grung, *Transistors—Fundamentals for the Integrated-Circuit Engineer*, John Wiley & Sons, New York, 1983.

Yang, E. S., *Fundamentals of Semiconductor Devices,* McGraw-Hill Book Company, New York, 1978.