10 *Diodes*

The device now called a *diode* was invented in the late nineteenth century, but it first saw widespread use in the galena crystal detector of 1907. This device was actually a Schottky diode formed between a metallic cat's whisker and semiconducting lead sulfide (galena). The copper oxide rectifiers and selenium stacks of the vacuum-tube era were also primitive Schottkies. Modern semiconductor diodes emerged from a different line of development that began with germanium point-contact diodes developed for military and computer applications. These were replaced in the mid-1960s by silicon PN-junction diodes similar to those in use today.

Diodes have found a number of applications in modern integrated circuits. Schottky diodes are often used as anti-saturation clamps for the collector-base junctions of NPN transistors. PN junction diodes form part of current mirrors and biasing networks. Junction diodes operated in reverse breakdown can also serve as voltage references and clamping devices. This chapter examines these and other applications of integrated diodes.

10.1 DIODES IN STANDARD BIPOLAR

The standard bipolar process can construct a wide variety of diodes. Of these, the most popular are the diode-connected transistor, the base-emitter Zener, and the Schottky diode. The first two are both variations of the bipolar NPN transistor, while the Schottky diode relies on the formation of a rectifying contact to lightly doped silicon. Not all versions of standard bipolar offer Schottky diodes, because they require the formation of platinum or palladium silicides and the addition of a special masking step to allow contact through the thick-field oxide. This section also discusses several additional types of Zener diodes sometimes available in standard bipolar.

10.1.1. Diode-connected Transistors

An NPN transistor consists of two back-to-back PN junctions, either of which could theoretically serve as a PN junction diode. In practice, the parasitic transistors associated with these junctions render them unsuitable for use as diodes. The collector-base

diode loses most of its current to substrate due to parasitic PNP action. NBL does not stop this parasitic conduction, because the carriers can still flow to the isolation sidewalls. Ringing the tank with deep-N+ contains the minority carriers and minimizes the current loss, but only at the price of greatly enlarging the structure.[1] The collector-base diode also exhibits a relatively high series resistance due to the light doping of the tank.

The base-emitter diode loses the vast majority of its current to the enclosing tank due to parasitic NPN action. Most of the carriers injected by the emitter travel across the base and into the tank. The accumulation of electrons in the collector causes the base-collector junction to forward-bias, turning on the parasitic PNP and diverting current to the substrate.

A very useful type of diode is created by tying the collector and the base of an NPN transistor together (Figure 10.1). The resulting device is often called a *diode-connected transistor*. Most of the current flows through a diode-connected transistor from collector to emitter by means of transistor action. Only a small current flows through the base terminal, so the base resistance has little effect on the device's forward voltage. The forward voltage also remains independent of collector resistance as long as the transistor does not fully saturate. A typical diode-connected transistor can tolerate about 400mV of debiasing at 25°C, or about 200mV at 150°C. If the collector debiasing exceeds these limits, then the diode begins losing current to the substrate due to parasitic PNP action. Diode-connected transistors often incorporate NBL to minimize collector series resistance. Diodes that must conduct more than a few hundred microamps should also contain a deep-N+ sinker. Diodes conducting 10mA or more should be laid out as power devices and should incorporate deep-N+ rings to minimize substrate injection during transients.

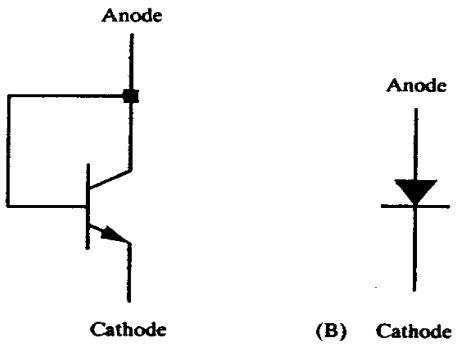

FIGURE 10.1 (A) Schematic and (B) symbol for a diode-connected transistor.

The diode-connected transistor will not suffer any loss of current to the substrate as long as collector debiasing does not exceed the limits given above. It also has much less series resistance than either the base-emitter diode or the base-collector diode. A minimum-size diode-connected transistor typically exhibits no more than 10 to 20Ω of series resistance. Its only serious drawback lies in its relatively low breakdown voltage, which is limited by the V_{EBO} of the NPN transistor to 6 to 8V.

Diode-connected transistors usually employ the CBE configuration rather than the CEB configuration (Figure 8.14). Although the CBE configuration has slightly more collector resistance, it allows first-level metal to connect the collector and base

[1] Some rather elaborate structures offer reasonable performance at the cost of large areas. See B. Murari, "Power Integrated Circuits: Problems, Tradeoffs, and Solutions," *IEEE J. Solid-State Circuits*, Vol. SC-13, #6, p. 307–319.

contacts. Many processes allow merged collector-base contacts, which save additional space (Figure 10.2). In this structure, the emitter diffusion surrounding the collector contact overlaps the base diffusion so that a single contact can touch both.[2] This contact must extend far enough into the base diffusion to account for misalignment and outdiffusion while still allowing sufficient base contact area for conduction. The contact must also extend into the collector far enough to counter misalignment while allowing adequate collector contact. Even with these overlaps, the merged structure is still considerably smaller than a traditional NPN layout.

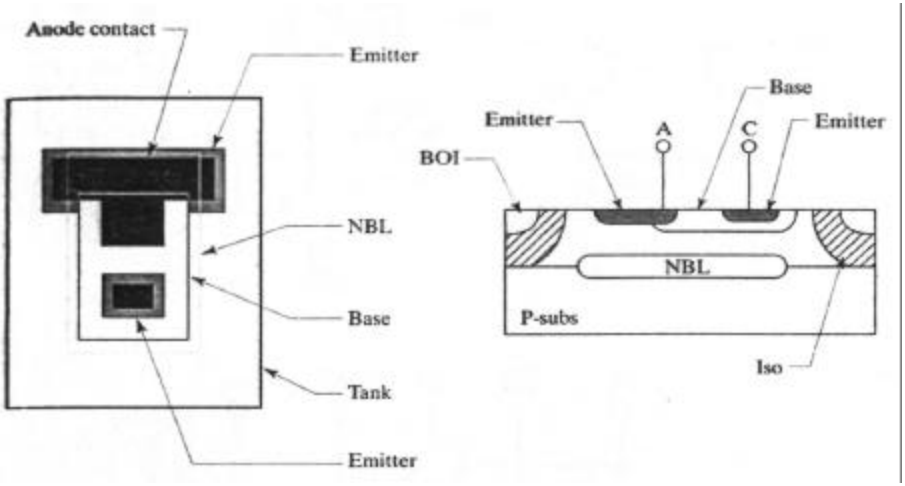

FIGURE 10.2 Layout and cross section of a standard bipolar diode-connected transistor.

The diode-connected transistor can serve as a convenient voltage reference. A base-emitter junction exhibits a forward voltage of about 0.65V at a current density of 1mA/mil^2 and a temperature of 25°C. A typical layout has an emitter junction area of 0.1mil^2 (about 60μm^2) and requires a current of about 100μA to develop a forward voltage of 0.65V. The forward voltage of a diode is relatively insensitive to small fluctuations in current. Even if the current through the diode were to double, the forward voltage would increase by only 18mV. The forward voltage also exhibits a temperature coefficient of about −2mV/°C. A stack of several diodes connected in series can develop larger voltages, but temperature and current variability increase proportionally.

A substrate PNP transistor can also serve as a diode, but the collector current of this device flows directly into the substrate. Currents that are much in excess of 1mA may debias the substrate enough to saturate the transistor. Diode-connected substrate transistors are sometimes used in CMOS processes that cannot fabricate other bipolar components (Section 10.3.1). They rarely see much use in processes that can fabricate isolated bipolar transistors.

Lateral PNP transistors make relatively poor diodes. They require large tanks that not only consume die area, but also contribute unwanted parasitic capacitance. Some portion of the collector current always flows to the substrate regardless of how thoroughly the device has been guard-ringed. The cathode current of a diode-connected lateral PNP is always less than its anode current. This loss prevents the use of lateral PNP transistors in applications in which current matching is critical. Still, diode-connected lateral PNP transistors are occasionally inserted into circuits to balance other PNP base-emitter voltage drops. An NPN transistor would not

[2] "Diodes," *Semiconductor Reliability News*, Vol. III, #7, July 1991, p. 9.

serve as well because its base-emitter voltage does not exactly match that of a PNP. The base and collector of the lateral PNP form the cathode of a diode-connected PNP, while the emitter acts as its anode. The layout sometimes uses a merged cathode contact that is analogous to the merged anode contact of the diode-connected NPN transistor shown in Figure 10.2.

10.1.2. Zener Diodes

A reverse-biased diode conducts very little current until the voltage across it exceeds a certain value. Beyond this point, the current through the diode increases exponentially until it eventually approaches an asymptote defined by the series resistance of the diode (Figure 10.3). The breakdown curve usually shows a fairly definite inflection point or *knee* corresponding to the *breakdown voltage* of the diode. The magnitude of the breakdown voltage depends on the width and curvature of the depletion region of the diode (Section 1.2.4). Carriers can tunnel across a very thin depletion region by a quantum process discovered by Zener.[3] Diodes with breakdown voltages of less than 6V are called *Zener diodes* since they conduct primarily by Zener tunneling. Diodes with breakdown voltages in excess of 6V are properly called *avalanche diodes* because they conduct primarily by avalanche multiplication instead of by tunneling. Designers frequently use the term *Zener diode* to describe all junction diodes operated in reverse breakdown regardless of conduction mechanism.[4]

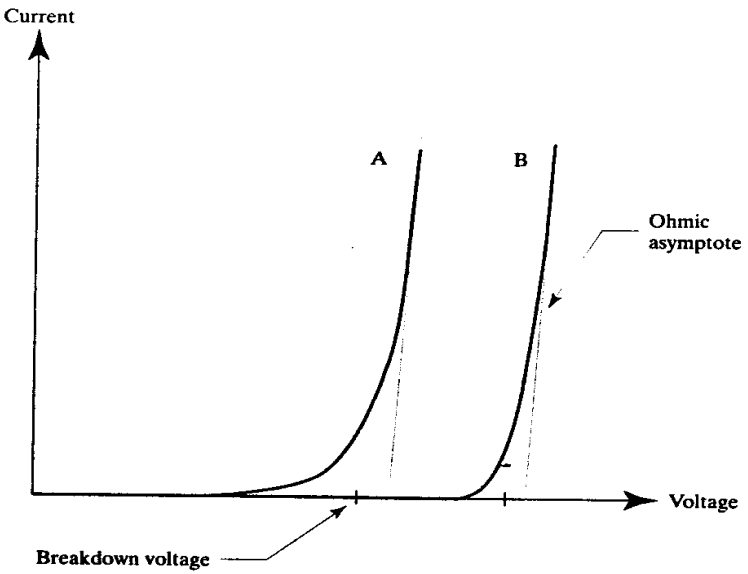

FIGURE 10.3 A comparison of the reverse breakdown characteristics of (A) a Zener diode and (B) an avalanche diode.

Zener tunneling exhibits a negative temperature coefficient, while avalanche breakdown exhibits a positive one. Diodes that break down at less than about 5.6V have negative temperature coefficients that increase in magnitude as the breakdown voltage diminishes. Diodes with breakdowns in excess of 5.6V exhibit increasingly positive temperature coefficients. The familiar emitter-base Zener has a break-

[3] C. Zener, *Proc. Roy. Soc.* A145, London: 1934, p. 523.

[4] Some authors use the term *breakdown diode* to refer to both Zener and avalanche diodes, but the engineering community has not widely adopted this practice.

down voltage of 6 to 8V with a positive temperature coefficient of 2 to 4mV/°C. A 40V base-collector Zener exhibits a much larger temperature coefficient, perhaps 35 to 40mV/°C.

Zener diodes with breakdown voltages of 5 to 6V have very small temperature coefficients. These diodes are sometimes used to construct temperature-independent voltage references, but they have several drawbacks that severely restrict their usefulness. Zener walkout (Section 4.3.1) causes the reference voltage to drift over time unless the device is specifically constructed to ensure subsurface breakdown. Zener references also require a supply voltage of at least 6V. Because of these disadvantages, Zener references have largely been replaced with lower-voltage alternatives such as bandgap references.

Zener diodes with breakdown voltages below 5V exhibit a rather gradual onset of reverse conduction (Figure 10.3A). This *soft breakdown* becomes more pronounced in lower-voltage devices. Diodes with breakdowns below about four volts have extremely soft breakdown characteristics. Designers often blame soft breakdown on leakage, but it is actually an unavoidable characteristic of Zener tunneling. In any event, Zener diodes with breakdown voltages of much less than 5V have little practical application because they do not exhibit well-defined reverse characteristics.

Surface Zener Diodes

The emitter-base junction of an NPN transistor forms a convenient Zener diode. Its breakdown voltage, V_{EBO}, depends on base doping and emitter junction depth. Most standard bipolar processes provide an emitter-base breakdown of about 6.8V. Advanced bipolar and BiCMOS processes often use lightly doped bases that exhibit emitter-base breakdowns of as much as 10V. Breakdown proceeds primarily by avalanche rather than by tunneling, so the temperature coefficient of the breakdown voltage is positive. A typical 6.8V emitter-base Zener exhibits a temperature coefficient of +3 to 4mV/°C.

Emitter-base Zeners have historically exhibited large amounts of process variation and long-term drift. Older bipolar processes used an emitter pilot step to control NPN beta, allowing the use of poorly controlled dopant sources such as boron nitride disks. The resulting variation in base doping and the compensatory changes in emitter junction depth caused V_{EBO} to vary by as much as ±1V. Ion implantation has dramatically improved base doping control, and most modern processes guarantee no more than ±0.25V initial variation. Process improvements have also reduced long-term drift due to Zener walkout, which now rarely exceeds 0.1V.

Emitter-base Zeners use essentially the same layout as NPN transistors (Figure 10.4). The emitter acts as the *cathode* of the Zener and the base acts as the *anode*. The tank serves only to isolate the Zener from the surrounding isolation. It should connect either to the cathode of the Zener or to some equal or higher voltage. The tank must never connect to the anode of the Zener lest the transistor become biased into the reverse-active region. If this occurs, then V_{ECO} breakdown will produce a snapback phenomenon similar to that exhibited by V_{CEO} breakdown (Section 8.1.2). Some designers leave the tank unconnected, but this practice is not recommended because it can amplify leakage currents. The floating tank acts as the base of a parasitic substrate PNP. The leakage current across the tank-isolation junction exceeds that across the smaller base-tank junction, and the difference between these currents forms the base drive of the substrate PNP. This mechanism can cause substantial loss of anode current to the substrate at elevated temperatures. Connecting the tank to the cathode prevents the parasitic PNP from amplifying leakage currents in this manner. The tank contact does not require a deep-N+ sinker because it only conducts

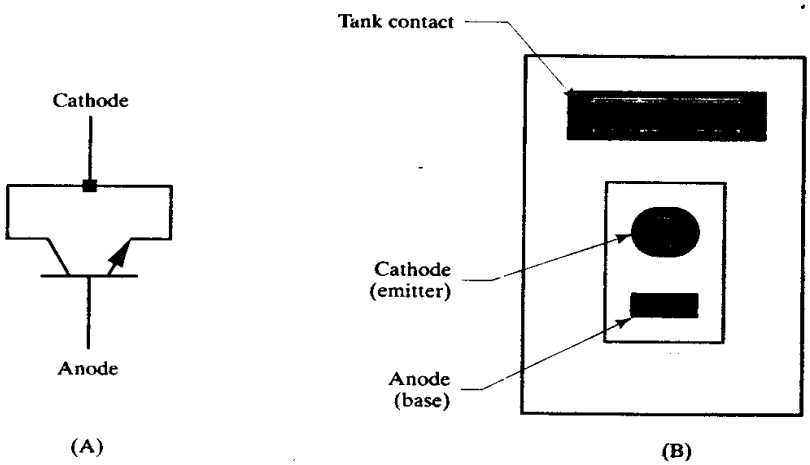

FIGURE 10.4 (A) Schematic and (B) layout of a typical base-emitter Zener diode.

leakage currents. NBL has little effect on a Zener and can therefore be omitted if the designer so chooses.

Emitter-base Zeners usually have circular or oval emitters like the one shown in Figure 10.4. These round geometries are intended to prevent electric field intensification at the corners of the emitter. Not all processes exhibit this phenomenon, but if it occurs it increases the variability of the breakdown voltage. One can determine whether a given process exhibits this effect by examining an avalanching rectangular emitter under a microscope in a darkened room. The avalanching junction will emit a faint white light.[5] If this light appears brightest at the corners, then electric field intensification is occurring at these points, and the device will benefit from the use of round emitters. Many designers routinely use round emitters because, even if they do not provide any benefit, they do no harm.

The base doping profile causes the base-emitter depletion region to narrow near the surface. The intensification of the electric field across the narrowed depletion region ensures that avalanche breakdown occurs at this point. Hot carriers produced by the intense electric field sometimes penetrate into the overlying oxide, where a small fraction becomes trapped. The gradual accumulation of this trapped oxide charge causes *Zener walkout* (Section 4.3.1). Some designers have attempted to suppress surface breakdown by flanging metal over the base-emitter junction to form a field plate. The thickness of the emitter oxide and the relatively low voltages placed across it prevent this field plate from having much effect. The avalanche process continues to occur near the surface and the field plate cannot prevent Zener walkout.

Emitter-base Zeners are relatively fragile devices because they dissipate energy in the relatively small volume of the emitter-base depletion region. The heat generated by this process can damage the junction or the adjacent contacts. Extreme overloads usually induce metal migration that causes permanent short-circuit failures. This mechanism has been employed as a replacement for fuses (Section 5.6.2). Base-emitter Zeners used as voltage references or clamp devices should not conduct more than about 10μA per micron of emitter periphery. For example, a 5×5μm emitter can safely conduct some 200μA. Zeners can tolerate much higher currents for brief periods of time, but long-term operation at elevated currents can cause

[5] "Junction Breakdown Characteristics," *Semiconductor Reliability News*, Vol. II, #1. January 1990, p. 8.

shifts in breakdown voltage due to junction damage. If an application requires more current than a small Zener diode can safely handle, consider using a power transistor to amplify the current conducted by the Zener, since this circuit takes much less area than a large Zener (Figure 13.22).

Some circuits connect the anode of the base-emitter Zener to the substrate. Since the base diffusion of the Zener operates at the same potential as the surrounding isolation, these two diffusions can overlap one another. This practice saves considerable area because it eliminates the large amount of spacing required to isolate the base diffusion from the isolation (Figure 10.5). A tank is placed beneath the emitter to prevent the isolation diffusion from reducing the Zener voltage. Although this tank remains unconnected, the diffusions around it are all biased to the same potential. The parasitic PNP transistor inherent in this structure cannot amplify leakage currents and is therefore harmless.

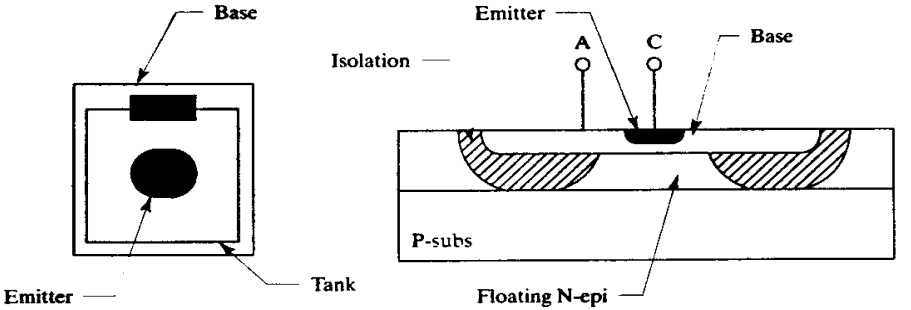

FIGURE 10.5 Layout and cross section of a nonisolated base-emitter Zener diode.

The nonisolated base-emitter Zener is vulnerable to both substrate debiasing and noise coupling. Nonisolated Zeners should never reside near structures that inject substrate currents of more than a few hundred microamps. Conservative designers often avoid using nonisolated Zeners, preferring to accept the extra die area of a conventional base-emitter Zener rather than risk unexpected debiasing or noise coupling.

Buried Zeners

If the avalanche region lies several microns beneath the oxide, hot carriers will scatter off the lattice and lose energy before they can reach the oxide interface. Zeners that avalanche beneath the surface are called *subsurface Zeners*, or more colloquially, *buried Zeners*. The breakdown voltages of buried Zeners remain constant throughout their operational lifetime, making these devices ideal for use as voltage references. This section presents several common varieties of buried Zeners that are compatible with standard bipolar processing.

A buried Zener can be constructed from the emitter and isolation diffusions of certain standard bipolar processes. This diode consists of a plug of P+ isolation covered by an emitter diffusion that overlaps into a surrounding tank (Figure 10.6).[6] The emitter counterdopes the surface of the isolation to form the cathode of the Zener. The anode consists of the portion of the isolation plug beneath the emitter. The N-epi surrounding the isolation plug prevents the emitter sidewall radius from further reducing the already-low breakdown voltage of the structure. The *emitter-in-iso* buried Zener usually exhibits a breakdown voltage of 5 to 6V and a temperature coefficient of about +1mV/°C.

[6] A. B. Grebene, *Bipolar and MOS Analog Integrated Circuit Design* (New York: John Wiley and Sons, 1984), pp. 133–134.

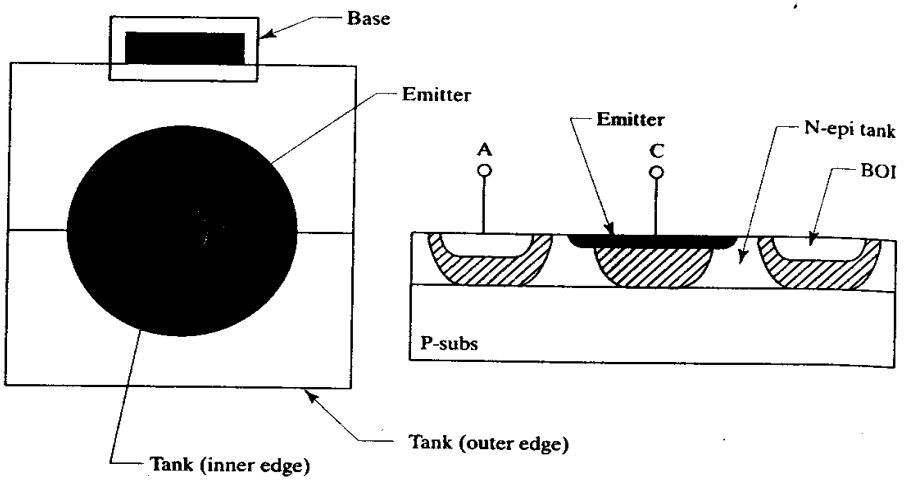

FIGURE 10.6 Layout and cross section of an emitter-in-isolation Zener diode.

Some processes use base over isolation to prevent channel formation. The presence of base in the isolation plug further reduces the breakdown voltage of the emitter-in-iso Zener. Although this may help minimize the residual temperature coefficient of the Zener, it also produces a softer breakdown characteristic. Breakdown voltages of less than 5V become rather indefinite and are therefore unsuited for use as voltage references. If the addition of base to the emitter-in-iso Zener causes its breakdown voltage to drop below 5V, then the designer should omit it from the area around the emitter plug.

The emitter-in-iso Zener shares most of the same drawbacks as the nonisolated base-emitter Zener. Its anode is electrically common to the substrate. This severely limits its range of applications and raises the possibility of substrate debiasing and noise coupling. Emitter-in-iso Zeners do not experience any appreciable long-term drift, but their initial voltage varies considerably because of the heavy deposition and long drive time required to fabricate the P+ isolation. A typical emitter-in-iso Zener exhibits a breakdown voltage of 5.4 ± 0.4V.

Several alternative styles of buried Zeners eliminate the drawbacks of the emitter-in-iso device at the cost of introducing additional processing steps. Figure 10.7 shows a buried Zener that requires the use of a deep-P+ diffusion inserted after isolation and before base.[7] The doping concentration in this deep-P+ diffusion signif-

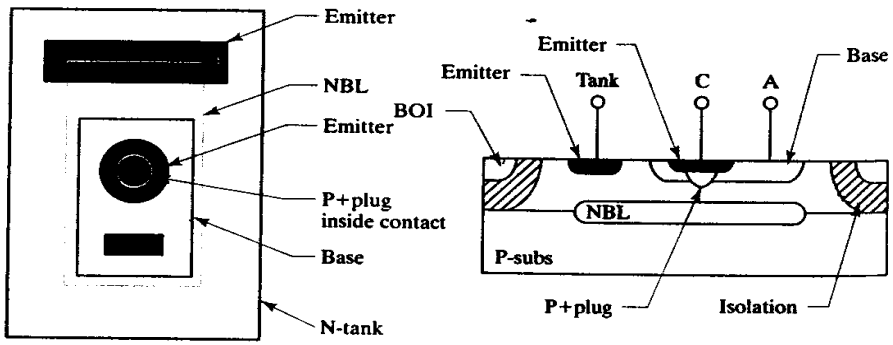

FIGURE 10.7 Buried Zener using a special deep-P+ diffusion in combination with emitter diffusion.

[7] Grebene, p. 134.

icantly exceeds that of the base diffusion but falls short of that of the emitter. The active region of the Zener consists of a plug of deep-P+ diffusion covered by emitter. The emitter diffusion is in turn enclosed by a base diffusion. The anode of the diode consists of the deep-P+ plug and is contacted by means of the surrounding base diffusion. The cathode of the diode consists of the emitter diffusion. The tank of this Zener is usually connected to the cathode to prevent beta multiplication of leakage currents. The structure in Figure 10.7 contains NBL, but it actually plays no role in the operation of the device. It can be omitted as long as the deep-P+ is sufficiently shallow to prevent punchthrough breakdown to the substrate.

The breakdown voltage of this structure can be tailored to suit a specific application by adjusting the profile of the deep-P+ diffusion. A lightly doped diffusion will have a higher breakdown voltage, while a heavily doped diffusion will have a lower one. The breakdown voltage cannot exceed that of the base-emitter junction, nor should it drop so low as to produce a soft breakdown characteristic. In practice, the breakdown voltage ranges from 5 to 6.5V. A typical device has a breakdown voltage of 6.3 ± 0.2V with a temperature coefficient of about +2mV/°C.

Another style of buried Zener substitutes a high-energy implant for the emitter diffusion of the standard base-emitter structure (Figure 10.8).[8] At high implant energies, the peak of the dopant distribution actually lies beneath the surface of the silicon. The high-energy implant therefore creates a shallow N-buried layer. The intersection of this layer with a plug of base diffusion forms a buried Zener diode. The breakdown voltage is set by adjusting the implant energy and dose used to fabricate the implanted NBL.

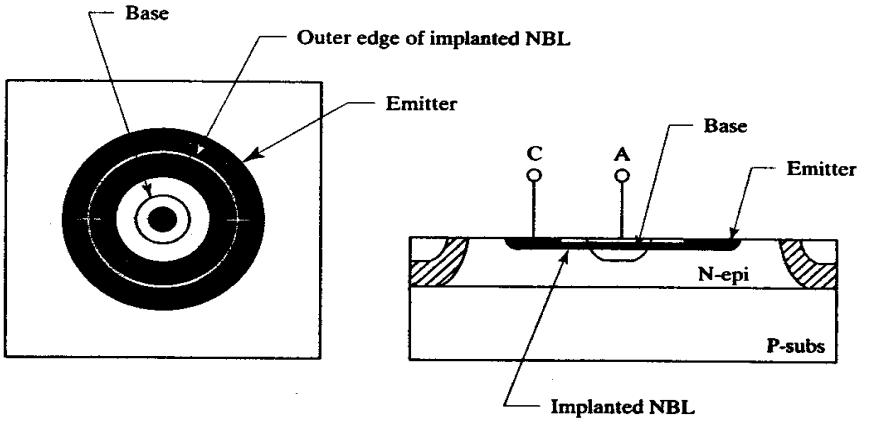

FIGURE 10.8 Buried Zener using an implanted NBL in combination with base diffusion.

All three of the buried Zeners just discussed provide breakdown voltages of 5 to 7V with relatively low temperature coefficients. Certain applications, such as ESD protection, require higher-voltage Zeners. These are usually constructed from stacks of series-connected base-emitter Zeners, possibly supplemented by diode-connected transistors. In addition to helping adjust the voltage of the stack, the negative temperature coefficient of the diode-connected transistors can offset part or all of the temperature coefficient of the Zeners. A few specialized high-voltage Zener structures take advantage of the large breakdown voltages between deep, lightly doped diffusions. Of these, the isolation-NBL Zener deserves special mention. This Zener

[8] Grebene, p. 134.

consists of a plug of isolation diffused into a plate of NBL. The NBL plate is enclosed in a tank contacted by means of a deep-N+ sinker. The voltage of the isolation-NBL Zener can vary by a few volts because it depends on a number of factors, including NBL and isolation doping, isolation drive time, and epi thickness. The isolation-NBL junction lies far beneath the surface and is therefore immune to Zener walkout. The relatively large area of the depletion regions and their depth beneath the surface also allow this structure to safely dissipate several times the power density of a base-emitter Zener.

10.1.3. Schottky Diodes

Schottky diodes depend on the formation of a rectifying Schottky barrier between a conductor and a semiconductor. Both P-type and N-type silicon form rectifying Schottky barriers with numerous metals and metal silicides. The forward voltages of the resulting diodes depend on the composition of the conductor and the polarity of the silicon (Table 10.1). Schottky diodes with forward voltages of less than 0.5V usually exhibit large junction leakages, especially at higher temperatures. This limitation generally prevents the use of P-type silicon in Schottky diodes.

TABLE 10.1 Typical forward voltages of Schottky diodes constructed from selected metals and silicides[9] (25°C, $1\mu A/\mu m^2$).

Material	N-type Silicon	P-type Silicon
Aluminum	0.54V	0.40V
Gold	0.62V	0.16V
Molybdenum	0.50V	0.24V
Palladium silicide (Pd_2Si)	0.57V	
Platinum silicide (PtSi)	0.66V	
Titanium silicide ($TiSi_2$)	0.42V	

The concentration of acceptors or donors in the silicon determines whether a given Schottky barrier exhibits rectifying or Ohmic behavior. Dopant concentrations exceeding 10^{17} atoms/cm^3 reduce the width of the depletion region to the point at which carriers can successfully surmount it by quantum tunneling. The resulting tunneling current increases almost exponentially with doping.[10] When the tunneling current exceeds the forward conduction current, the Schottky barrier begins to resemble a resistor. Practical Schottky diodes normally employ surface doping concentrations of no more than 10^{16} atoms/cm^3 to minimize tunneling effects.

Many applications require Schottky diodes with forward voltages significantly lower than those of PN junction diodes. Aluminum would appear to be an ideal material for constructing such diodes. Its forward voltage lies comfortably below that of a PN junction diode, yet not so low as to cause excessive junction leakage. Unfortunately, sintering can drastically alter the properties of an aluminum-silicon Schottky. The aluminum dissolves a small amount of silicon during sintering. As the wafer cools, some of the dissolved silicon redeposits at the metal-semiconductor interface as an aluminum-doped

[9] These values are derived from the barrier potential ϕ_B, assuming current density of $1\mu A/\mu m^2$ and a Richardson constant of $120 A/cm^2/K^2$. Under these circumstances, the forward voltage is 180mV less than ϕ_B. Barrier potentials from S. M. Sze. *Physics of Semiconductor Devices*, 2nd ed. (New York: John Wiley and Sons, 1981), pp. 290–291.

[10] Actually, theory predicts that the tunneling current varies exponentially with respect to the square root of the dopant concentration and linearly with respect to the applied voltage; See W. R. Runyan and K. E. Bean, *Semiconductor Integrated Circuit Processing Technology* (Reading, MA: Addison-Wesley, 1994), p. 524.

P-type semiconductor. This deposit constricts the Schottky contact area and, in extreme cases, may entirely cover the contact opening. This mechanism can cause the forward voltage of a Schottky to increase by several hundred millivolts.[11]

Few other metals exhibit the properties required to construct practical Schottky diodes. Most are difficult to sinter and exhibit excessive forward voltage variation. Modern integrated Schottky diodes employ the silicides of certain noble metals, most notably platinum and palladium. These *noble silicides* provide extremely stable and repeatable forward voltages lying in the desired range of 0.5 to 0.7V. The inability of noble silicides to withstand the temperatures required for source/drain annealing limits their application in CMOS processes. The forward voltages of the *refractory silicides* (such as titanium silicide) are rarely sufficient to prevent leakage, so processes that use these silicides rarely offer Schottky diodes.

Most processes indiscriminately silicide all contact openings. Those lying over heavily doped silicon become Ohmic contacts, while those residing over lightly doped N-type silicon become Schottky diodes. These processes generally prohibit the opening of contacts over lightly doped P-type silicon, because the resulting Schottky barrier has a contact resistance that is too high to function as an Ohmic contact and a forward voltage that is too low to function as a Schottky diode.

Figure 10.9 shows the layout and cross section of a Schottky diode constructed in a standard bipolar process. The metal system consists of a sandwich of platinum silicide, refractory barrier metal, and copper-doped aluminum. The Schottky barrier forms between the platinum silicide and the lightly doped N-type epi. In order to reach the epi, the Schottky contact must penetrate the thick-field oxide. If this contact opening were etched simultaneously with the base and emitter contacts, the latter would suffer severe overetching before the former cleared. Most standard bipolar processes include an additional etching step to thin the oxide over the Schottky contact openings prior to the regular contact oxide removal. The process extension required to form Schottky contacts therefore consists of a single masking step and a single oxide removal.

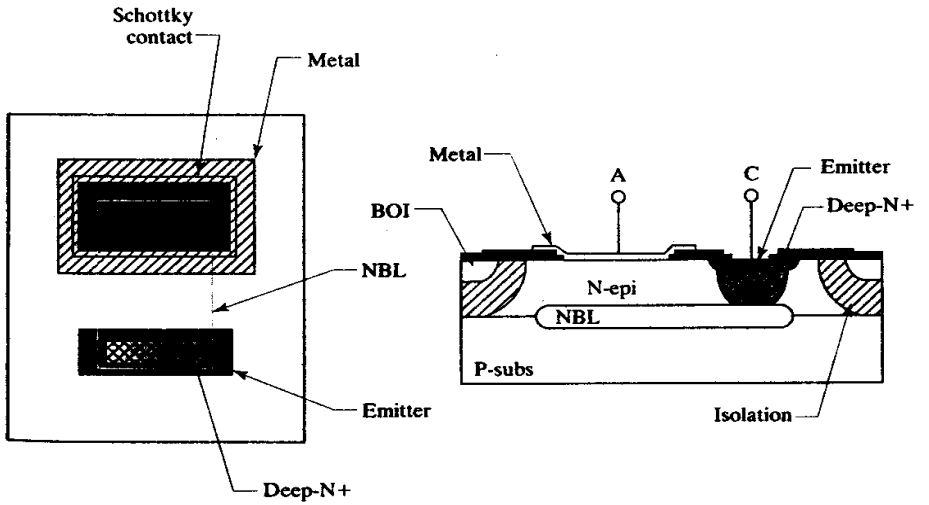

FIGURE 10.9 Layout and cross section of a field-plated Schottky diode.

[11] M. Mori, "Resistance Increase in Small-Area Si-Doped Al-n-Si Contacts," *IEEE Trans. on Electron Devices*, Vol. ED-30, #2, 1983, pp. 81–86.

The cathode of the Schottky includes NBL and deep-N+ to minimize its series resistance. A small plug of deep-N+ suffices to extract the cathode current as long as this does not exceed a milliamp or two. High-current Schottky diodes usually employ rings of deep-N+ around the periphery of their tanks to further reduce the series resistance. Large diodes can occupy empty spaces between other components, because the shape of the Schottky contact has little effect on its performance.

The planar breakdown voltage of a Schottky diode usually exceeds the V_{CEO} rating of the corresponding NPN transistor by at least a factor of two. Practical Schottky diodes obtain only a fraction of this theoretical breakdown voltage because the sharp edges of the Schottky contact greatly intensify the electric field. An unprotected Schottky usually begins to avalanche at a reverse bias of only a few volts. Several techniques have been developed that diminish the electric field intensity at the edges of the Schottky contact. The simplest of these consists of flanging the metallization over the Schottky contact to form a field plate. The full reverse-bias voltage appears between the field plate and the underlying silicon. The resulting vertical electrical field repels electrons from the surface of the silicon and makes it appear to be more lightly doped. This delays the onset of avalanche and increases the reverse-bias voltage rating by a few volts. The field plate need only overlap the contact by 3 to 5μm. This relatively simple and compact arrangement suffices for many low-voltage applications.

Figure 10.10 shows an alternative style of Schottky diode that can withstand much higher reverse voltages. This structure encloses the edge of the Schottky contact within a thin strip of base diffusion called a *field relief guard ring*.[12] The presence of the guard ring completely eliminates the lateral electric field at the edge of the Schottky contact and raises the breakdown voltage of the structure to equal the V_{CBO} of the base diffusion. These field-relief guard rings are completely unrelated to the minority carrier guard rings of Section 4.4.2.

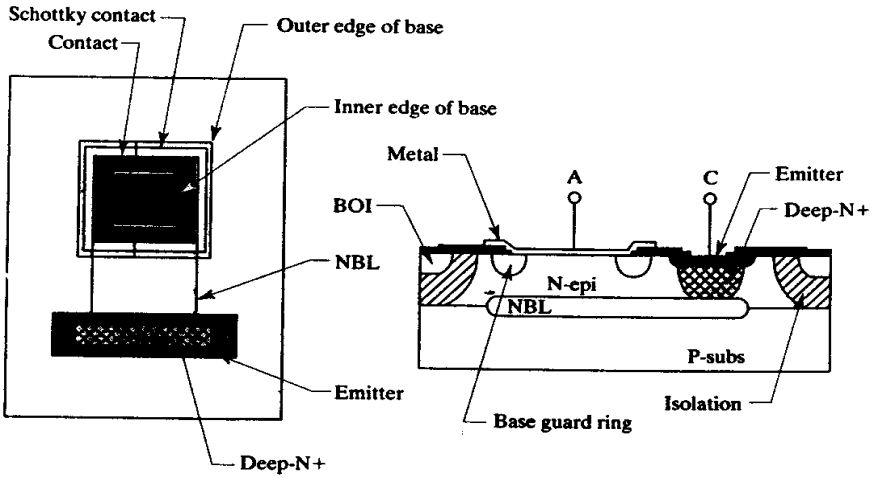

FIGURE 10.10 Layout and cross section of a base guard-ring Schottky diode.

The inclusion of a field-relief guard ring inevitably enlarges the Schottky diode. The spacing between the tank and the Schottky contact must increase to avoid punchthrough between the guard ring and the isolation. Outdiffusion also constricts

[12] M. P. Lepselter and S. M. Sze, "Silicon Schottky Barrier Diode with Near-Ideal I-V Characteristics." *Bell Sys. Tech. J.*, Vol. 47, #2 1968, p. 195–208.

the contact opening by several microns on all sides. These considerations affect the area of small Schottky diodes much more severely than they affect the area of large ones. Many designers routinely add guard rings to large Schottkies because they render the breakdown characteristics of the diode much more predictable and repeatable. On the other hand, these same designers frequently use field plates on small Schottkies to conserve area.

Schottky diodes are often used as antisaturation clamps for NPN transistors (Section 8.1.4). An antisaturation diode can occupy the same tank as the transistor it protects. A field-plated diode is easily created by extending the base contact out into the surrounding tank (Figure 10.11A). A guard-ringed diode requires the addition of a narrow strip of base around the periphery of the Schottky contact opening (Figure 10.11B). Minimum-size devices usually forego the addition of the guard ring to conserve area. Devices that must tolerate higher collector-base voltages may require guard rings, as may those that cannot tolerate any base-collector leakage.

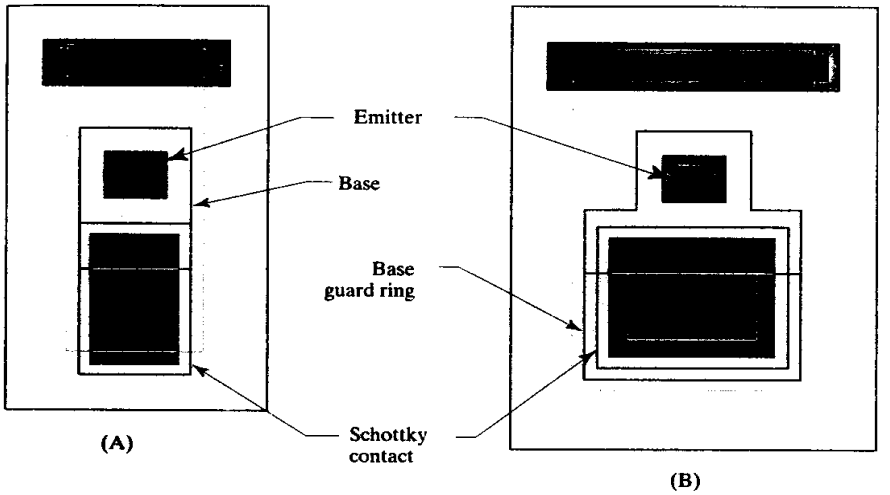

FIGURE 10.11 Schottky-clamped NPN transistors using (A) field plates and (B) base guard rings.

In order for a Schottky diode to serve as an antisaturation clamp, it must have a lower forward voltage than the base-collector junction of the protected transistor. The size of the required contact opening depends on three factors: the materials used to fabricate the Schottky, the maximum operating temperature, and the current it must conduct. Palladium silicide Schottky diodes make excellent antisaturation clamps because they have relatively small forward voltages. Platinum silicide Schottky diodes have larger forward voltages and therefore require careful attention to contact areas. Devices that must operate at high temperatures require larger contact areas because the Schottky forward voltage has a smaller temperature coefficient than the base-collector forward voltage. This phenomenon prevents platinum silicide Schottky antisaturation clamps from operating at temperatures much higher than 150°C. The same effect can cause guard-ringed Schottky diodes to begin to inject minority carriers to the substrate at temperatures in excess of 130 to 140°C. Junction leakage imposes similar temperature limitations on palladium silicide Schottky diodes.

The area of the Schottky contact required to protect a transistor from saturation must be determined by empirical measurement. In order to perform this measurement, current is forced through the base-emitter junction of a transistor whose collector remains unconnected. The base drive is increased until substrate injection exceeds a predetermined percentage of the base drive—perhaps 5%.

This measurement should be performed at the maximum operating junction temperature. The amount of base drive that the Schottky clamp can successfully absorb scales with the ratio of Schottky area divided by base-collector junction area. Based on a single experiment, one can determine approximately how much Schottky contact area is required to protect any given transistor. Outdiffusion may also affect the Schottky contact area of a guard-ringed device. Assuming the Schottky contact is rectangular, then the effective Schottky area, A_S, can be computed from the drawn dimensions of the base opening X_D and Y_D using the following equation

$$A_s = (X_D - 2\delta)(Y_D - 2\delta) \qquad [10.1]$$

where δ is a correction factor that accounts for base outdiffusion, photolithographic size adjusts, and so forth. Since forward current scales linearly with Schottky area, the correction factor can be determined by measuring the currents required to obtain a specified forward voltage for two Schottky diodes of different dimensions.

Schottky diodes also see occasional use as power devices. High current densities tend to activate parasitic mechanisms that would otherwise remain dormant. Figure 10.12 shows the parasitics associated with a typical Schottky diode. D_1 represents the Schottky diode itself, R_1 models the resistance of the NBL and deep-N+, and R_2 models the resistance of the N-epi beneath the Schottky contact. The epi contributes more resistance to a Schottky diode than to an NPN transistor because current in the Schottky must cross the entire thickness of the epi. This resistance is not necessarily undesirable because it acts as distributed ballasting and prevents thermal runaway under all but the most extreme conditions. Power Schottky diodes can therefore consist of a single contact opening of any desired shape.

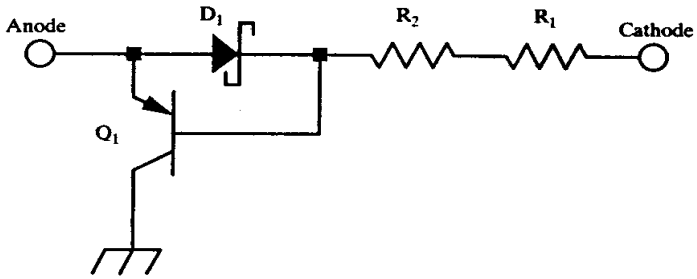

FIGURE 10.12 A simplified parasitic model of a Schottky diode.

Transistor Q_P in Figure 10.12 models minority carrier injection within the Schottky diode. Although Schottky diodes are majority carrier devices, they do experience some minority carrier injection at high current. Schottky field-relief guard rings will inject minority carriers into the cathode tank when the voltage across the diode exceeds the forward voltage of the junction between the guard ring and the N-epi. Field-plated devices lack this parasitic PN junction, but the Schottky barrier itself emits small numbers of minority carriers at high current densities. Unless these minority carriers are blocked, they will flow across the N-epi into the substrate. The presence of NBL in the Schottky helps minimize substrate injection, but power Schottky diodes should also contain a continuous ring of deep-N+ around their perimeter. This ring serves as both a cathode contact and as a minority carrier guard ring.

10.2 DIODES IN CMOS AND BICMOS PROCESSES

Every semiconductor process can produce at least one type of diode. Analog BiCMOS processes can fabricate most of the diodes available in standard bipolar. Pure CMOS processes offer fewer options and therefore deserve closer consideration.

Although one could theoretically construct PN junction diodes using any of the three junctions available in a CMOS process, only one can be biased into conduction under normal operating conditions. In an N-well CMOS process, the anodes of the NSD/P-epi diode and the N-well/P-epi diode both connect to the substrate. These diodes can only be forward-biased by pulling their cathode below the substrate. Not only do these diodes require a negative power supply, but they also pose a latchup hazard due to minority carrier injection into the substrate. The PSD/N-well junction does not exhibit these problems, but it contains a parasitic PNP that diverts a significant fraction of the diode current to the substrate. The PSD/N-well diode also has a large internal series resistance (Section 10.3.1).

Zener diodes can be constructed using either the PSD/N-well junction or the NSD/P-epi junction. The parasitic PNP transistor inherent in the PSD/N-well diode does not conduct as long as the junction remains reverse-biased. Similarly, the NSD/P-epi junction does not inject minority carriers into the substrate as long as it remains reverse-biased. Although these Zeners do function, their breakdown voltages always exceed the operating voltage of the process, and they find application only as ESD protection devices and gate protection clamp diodes.

Figure 10.13 shows a typical layout for a PSD/N-well Zener diode. Avalanche breakdown occurs across the depletion region surrounding the PSD implant. The high sheet resistance of the well tends to focus conduction at the edges of the PSD facing the NSD contact. The effective area of conduction can be increased by using long narrow stripes of PSD interdigitated with strips of NSD (Figure 10.13). Even with these improvements, the current-handling capability of the PSD/N-well Zener is substantially inferior to that of a base-emitter Zener.

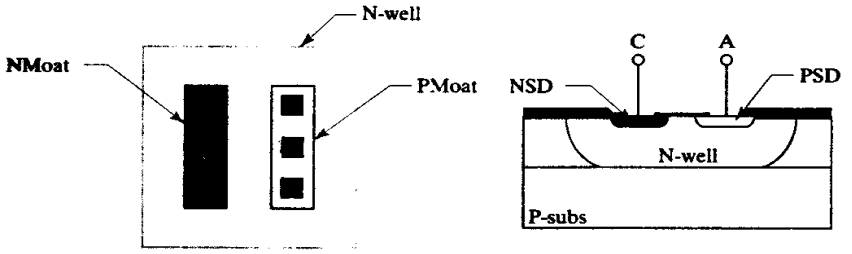

FIGURE 10.13 Layout and cross section of a PSD/N-well Zener diode.

Figure 10.14 shows a typical Schottky diode formed in a CMOS process between a noble silicide and the N-well. The moat geometry coded across the contact opening ensures that the contact resides over thin oxide. A ring of PSD diffusion placed around the edge of the contact opening serves as a field-relief guard ring. The reverse breakdown voltage of the Schottky is limited by the avalanche voltage of the PSD/N-well junction rather than by the planar breakdown voltage of the Schottky. A simple field plate can replace the PSD guard ring, but since field-plated Schottky diodes exhibit larger leakages than guard-ringed Schottkies, most designers opt to use guard rings whenever possible.

Schottky diodes fabricated in CMOS processes have relatively high series resistances due to the absence of NBL and deep-N+. This resistance can be minimized by elongating the Schottky contact and by surrounding it on all sides with cathode contacts. Larger Schottky diodes may employ interdigitated anode and cathode contacts. These measures cannot reduce the series resistance as effectively as NBL and deep-N+, so the resulting Schottky diodes are useful only for relatively low-current applications.

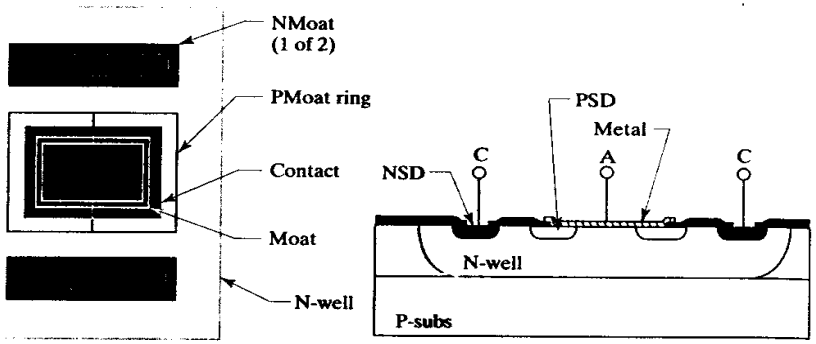

FIGURE 10.14 Layout and cross section of a PSD guard-ring Schottky diode in a CMOS process.

One other type of CMOS diode deserves mention. Some processes fabricate both P-type and N-type gate polysilicon. A polysilicon diode could reside over active circuitry and would therefore save considerable die area. Unfortunately, poly diodes exhibit much higher leakage currents than ordinary diffused diodes due to the presence of defects along the grain boundaries intersecting the junction. The properties of poly diodes also depend on annealing conditions, because the grain structure of the polysilicon changes during high-temperature processing. These concerns have prevented widespread utilization of poly diodes.

The selection of diodes available in analog BiCMOS closely parallels that offered by standard bipolar. Every BiCMOS process offers some variant of the diode-connected transistor, and most can also fabricate Zener diodes analogous to the base-emitter Zener of standard bipolar. The BiCMOS Zener often exhibits a higher breakdown voltage (7 to 10V) due to the use of a more lightly doped base region than standard bipolar. The shallowness of the diffusions in modern BiCMOS processes reduces the conduction volume in these Zeners, making them somewhat more fragile than their standard bipolar counterparts.

Analog BiCMOS processes that employ platinum or palladium silicides can fabricate Schottky diodes. The structure of these diodes resembles that of their standard bipolar counterparts. A moat geometry coded around the anode contact takes the place of the Schottky contact geometry (Figure 10.15). This substitution allows Schottky diodes to be constructed without requiring any additional masking steps. Many BiCMOS designers use these devices as replacements for diode-connected transistors.

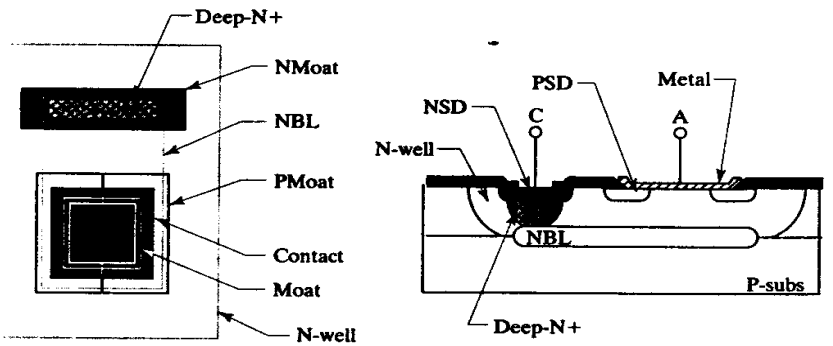

FIGURE 10.15 Layout and cross section of a PSD guard-ring Schottky diode in a BiCMOS process.

10.3 MATCHING DIODES

The devices discussed in this chapter fall into three broad categories: PN junction diodes, Zener diodes, and Schottky diodes. Different types of diodes do not match one another because they obey different principles of operation. Diodes of the same type can match, but only if they are properly constructed. This section briefly discusses the advantages and disadvantages of each type of matched diode and presents some guidelines to aid the designer in constructing matched diodes.

10.3.1. Matching PN Junction Diodes

Most junction diodes used in bipolar and BiCMOS processes are actually diode-connected transistors. As such, their layouts closely resemble those of conventional bipolar transistors. The only unique feature found in diode-connected transistors is the merged collector-base contact (Figure 10.2). The techniques used to match diode-connected transistors are otherwise identical to those used for other bipolar transistors (Sections 9.2 and 9.3).

Pure CMOS processes do not support the construction of diode-connected transistors, but they can still produce PN junction diodes. An N-well CMOS process can construct a PSD/N-well diode, and a P-well process can construct an NSD/P-well diode. Both of these devices contain parasitic vertical bipolar transistors that divert a significant fraction of their forward current to the substrate. The beta of these bipolar transistors can range from less than 0.1 to more than 10. A typical device has a beta of 2. The resulting base current generates a voltage drop across the well resistance that adds to the forward voltage of the diode. Any variation in beta produces a corresponding variation in the voltage drop across the well resistance. The matching of junction diodes therefore depends on the matching of parasitic bipolar betas and well resistances as well as forward voltages.

The layout of a CMOS PN junction diode is identical to that of a CMOS substrate transistor (Section 8.3.3). This device may be viewed either as a PN junction diode containing a parasitic bipolar transistor, or as a bipolar transistor having so low a gain that it resembles a diode. Either way, one must maximize beta and minimize well resistance to obtain any semblance of matching. These factors have such a strong impact that a device with a small junction area-to-periphery ratio will actually exhibit better matching than one that has a large area-to-periphery ratio. The best layout consists of an array of minimum-width junctions interdigitated with well contacts. Matched devices should be interdigitated with one another to produce a common-centroid array (Section 7.2.6). Figure 10.16 shows a pair of matched PSD/N-well diodes.

Many CMOS processes use silicided moats (*clad moats*), but the emitter of a bipolar transistor should not be silicided because this reduces its beta. The matching of PN junction diodes actually improves when they incorporate high-gain parasitic bipolar transistors, because the high gain reduces the currents that must flow through the well resistance. If a silicide block mask is available, then the designer should code this layer around the fingers of the diodes that form its PN junctions (Figure 10.16).

The matching of PN junction diodes also improves at low current densities, because the voltage drop across the well resistance decreases. The current density can be decreased either by increasing the area of the devices or by operating them at lower currents. CMOS diodes typically operate at current densities of 1 to $10\mu A/mil^2$ of junction area (4 to $40nA/\mu m^2$).

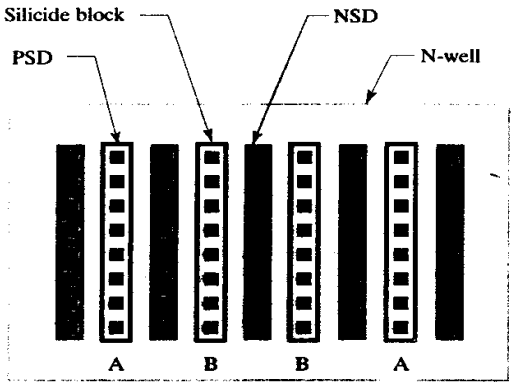

FIGURE 10.16 Layout of a pair of matched PSD/N-well diodes. The anodes of the two diodes are marked A and B.

Regardless of the care taken in their construction, PSD/N-well and NSD/P-well diodes will always exhibit residual mismatches of several millivolts because of the large emitter periphery-to-area ratios and the influence of well resistance and beta variation. Diode-connected transistors and Schottky diodes will almost certainly provide better matching.

10.3.2. Matching Zener Diodes

Zener diodes are difficult to match because their breakdown voltages depend so strongly on electric field intensity. Any curvature in the junction geometry intensifies the electric field and produces a localized reduction in breakdown voltage. The portion of the junction that has the largest curvature conducts the majority of the current and therefore sets the breakdown voltage of the device. Localized breakdown degrades matching because it reduces the effective area of the junction and magnifies the effects of outdiffusion. Matched Zeners should employ circular junction geometries to avoid introducing unnecessary corner curvature. Unfortunately, even circular junctions seldom break down uniformly. Linewidth variations produce minute irregularities that have larger curvatures than the remainder of the junction sidewall. The resulting variations in conduction are often visible under a microscope in a darkened room. The faint glow of the avalanching junction usually appears at only a few spots around the perimeter of the junction. Almost all of the current flows through these spots, which represent only a small fraction of the junction periphery. Devices that have identical layouts often show radically different patterns of conduction. These variations highlight the essentially random distribution of defects and linewidth variations.

Large devices generally exhibit a more uniform distribution of defects. Higher current densities also minimize variability by increasing the voltage drop across the lightly doped diffusion adjacent to the junction. The resulting voltage drop provides ballasting and helps distribute conduction across a larger area, but it also represents another source of variability.

Matched Zeners usually employ large circular geometries to minimize random variations and to eliminate edge effects. Both the anode and the cathode contacts should be circularly symmetric, or nearly so. If necessary, the amount of resistive ballasting in the device can be increased by spacing the contacts further away from the junction. Figure 10.17 shows a cross-coupled pair of matched base-emitter Zeners constructed following these guidelines. The shape of the anode contacts resembles a four-leaf clover, or *quatrefoil,* to allow leads to connect the cathodes to one another without stacking vias over contacts. The four individual Zeners occupy a common tank to minimize the separation between them. The tank isolates the

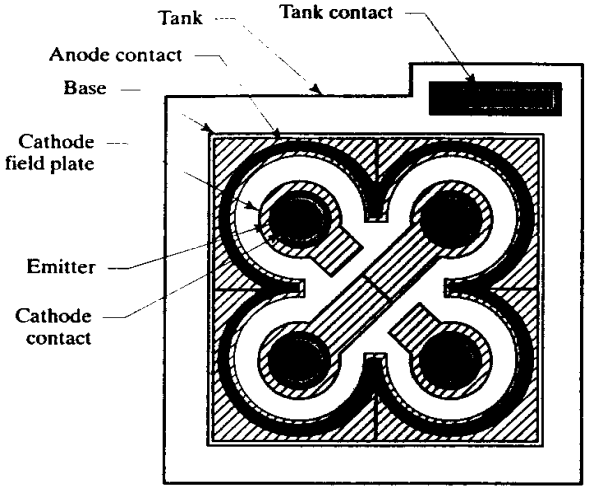

FIGURE 10.17 Quatrefoil layout of cross-coupled base-emitter Zeners. This layout assumes the use of double-level metal.

Zeners from the substrate, but it does not include NBL and deep-N+ because it does not have to conduct any significant current.

Even the elaborate quatrefoil layout of Figure 10.17 may not provide precise matching between surface Zeners, because these devices are susceptible to Zener walkout. The magnitude of this walkout depends on the total charge that has passed through the device, the magnitude and direction of the electric fields through the overlying oxide, and the concentration of mobile ions in the oxide. The layout in Figure 10.17 flanges the emitter metallization over the junction. This flange functions as a field plate that ensures that each Zener sees exactly the same voltage across its junction. The overlap of the field plate over the base-emitter junction should account for misalignment, outdiffusion, and fringing fields. An overlap of 5 to 8μm usually suffices. The field plate cannot stop Zener walkout, but it may help minimize its variability. Buried Zeners provide much better matching because they do not exhibit Zener walkout.

10.3.3. Matching Schottky Diodes

Like Zeners, Schottky diodes are inherently difficult to match. The characteristics of the Schottky barrier depend on several factors, including metal composition, silicon doping, edge effects, annealing conditions, and the presence or absence of surface contaminants. Most of these factors are difficult to control, so Schottky diodes usually exhibit larger mismatches than diode-connected transistors.

Matched Schottky diodes should always incorporate diffused guard rings rather than field plates because field-plated structures often exhibit leakages that could interfere with low-current matching. The contact opening should have a large area-to-periphery ratio to minimize mismatches due to linewidth variations. If possible, the diodes should incorporate NBL and deep-N+ to minimize the portion of the cathode resistance not directly associated with the Schottky contact. If NBL and deep-N+ are used, then several matched Schottky diodes can reside in the same tank or well. If deep-N+ and NBL are not used, then each Schottky should occupy its own well or tank, all of which should have exactly the same dimensions to ensure that the cathode resistances of the diodes match. Ratioed Schottky diodes should always consist of arrays of identical unit contacts, in much the same way (and for much the same reasons) as ratioed NPN transistors use arrays of identical unit emitters. Because Schottky diodes are very sensitive to thermal gradients, they should always use interdigitated or cross-coupled layouts similar to those employed for bipolar transistors.

Schottky diodes will not reliably match PN junction diodes or bipolar transistors. The voltage differential between a diode forward voltage and a Schottky forward voltage may vary by a few percent because of variations in surface conditions, anneal times, and other factors. Even the voltage differential between two Schottky diodes operated at different current densities may depend on processing conditions due to the nonideality factor in the diode equation, which typically plays a larger role in Schottky diodes than it does in junction diodes and bipolar transistors.

10.4 SUMMARY

Every semiconductor process offers one or more types of diodes. Standard bipolar and analog BiCMOS processes can fabricate excellent diode-connected NPN transistors that exhibit mismatches of only a few millivolts and can handle large amounts of current. These transistors can also function as emitter-base Zener diodes. Pure CMOS processes offer fewer options, but they may still be able to construct Schottky diodes if their metallization system includes a noble silicide.

10.5 EXERCISES

Refer to Appendix C for layout rules and process specifications.

10.1. Lay out a standard-bipolar diode-connected transistor using a minimum-size emitter. Compare the area of this device with the area of a minimum-size NPN transistor that does not contain deep-N+. Additional rules for constructing the collector-base contact of the diode are as follows:

1. **CONT** extends into **EMIT** 4μm
2. **CONT** overhang **EMIT** 6μm
3. **BASE** overhang **EMIT** 2μm

10.2. Lay out a minimum-size emitter-in-isolation Zener using standard bipolar rules. The emitter should overlap the isolation plug by 18μm.

10.3. What is the approximate temperature coefficient of a series combination of a 6.8V emitter-base Zener and two diode-connected NPN transistors?

10.4. Construct a standard bipolar field-plated Schottky diode with an area of 200μm^2. Overlap the field plate over the contact by 4μm. Include a deep-N+ sinker along one end of the device. Include all necessary metallization.

10.5. Construct a standard bipolar base guard-ringed Schottky diode having an effective area of 200μm^2. Assume that the correction factor δ equals 2.0μm. The layout rules for the base guard ring are as follows:

1. **BASE** overhang **CONT** 4μm
2. **BASE** extends into **CONT** 4μm

10.6. Construct a standard bipolar power Schottky diode with an effective area of 15000μm^2. Include a base guard ring constructed according to the rules given in Exercise 10.5. Ring the cathode tank with deep-N+ and provide as much metallization as possible, leaving space for a 12μm-wide cathode lead. The anode metallization should be at least 10μm wide at all points and should exit the device through a 12μm-wide anode lead opposite the cathode lead.

10.7. Why does the analog BiCMOS process of Appendix C not support Schottky diodes? What modification would allow their construction?

10.8. Lay out a pair of CMOS PSD/N-well diodes for optimal matching. The diodes should have drawn areas of 60 and 120μm^2, respectively. Draw silicide block (SBLOCK) around the anodes so that SBLOCK overlaps PMOAT by 1μm.

10.9. Lay out a quatrefoil Zener structure consisting of four individual diodes sharing a common anode, using standard bipolar rules. Use a diameter of 12μm for each of the four emitters. Include all necessary metallization, showing how the cathode lead exits from the device and how the tank contact is connected.