Document Type: Prentice Hall
Author: Djafar K. Mynbaev and Lowell L. Scheiner
Book: Fiber-Optic Communications Technology
Copyright: 2001
ISBN: 0-13-962069-9
NI Supported: No
Publish Date: Sep 6, 2006

Photodiodes

Overview

If LEDs and laser diodes convert an electrical signal into light, the function of a photodiode is just the opposite: to convert light into an electrical signal. Thus, the principle of operation of a photodiode (PD) can be explained simply as operating in a manner exactly the opposite to the way an LED works. Indeed, the discussion of a PD's operation involves the same elements: energy bands and a p-n junction. This is why rereading the appropriate sections in Chapter 9 would be a good idea if you need to refresh your memory on how an LED works.

Table of Contents

p-n Photodiodes: How They Work

From the standpoint of energy bands

You will recall that in semiconductors we deal with conduction and valence energy bands—two bundles of energy levels—separated by a forbidden region, an energy gap (E_g). The conduction band has higher energy than the valence band.

Electrons at the valence band are bonded and cannot move; thus, no current flows through the material. Electrons at the conduction band are free and when a small voltage is applied, they move, constituting current. In other words, *to induce material to conduct current, one needs to populate the conduction band with electrons.* But one obstacle stands in the way: the energy gap. The value of E_g determines the conductive (resistive) properties of the material. Good conductors have no gap between the valence and conduction bands, good insulators have a big energy gap, and semiconductors have a gap somewhere in between. Indeed, diamond—a good insulator—has an energy gap around 6 eV, while silicon (Si) and germanium (Ge)—the most popular semiconductors—have gaps of 1.17 eV and 0.775 eV, respectively. (See Table 9.1.)

When a photon with energy $E_p = hf = hc/\lambda \geq E_g$ strikes the material, the photon is absorbed and its energy acquired by an electron. Thus, the electron is excited at the conduction band and is now able to move. This is how light power—a number of photons times a photon's energy per unit of time—is converted into electrical current. This explanation is visualized in Figure 11.1(a).

If we apply external voltage—bias—to this semiconductor, we make electrons flow in a much more pronounced manner, thus increasing the efficiency of the light-to-current conversion.

From the standpoint of a *p-n* junction

When a photon strikes a depletion region, its energy separates an electron from its hole, as Figure 11.1(b) shows. (Remember, electrons and holes have recombined at the *p-n* junction, thus creating a depletion region.) The separated electron and hole are attracted by the positive and negative potentials of the depletion voltage, respectively.

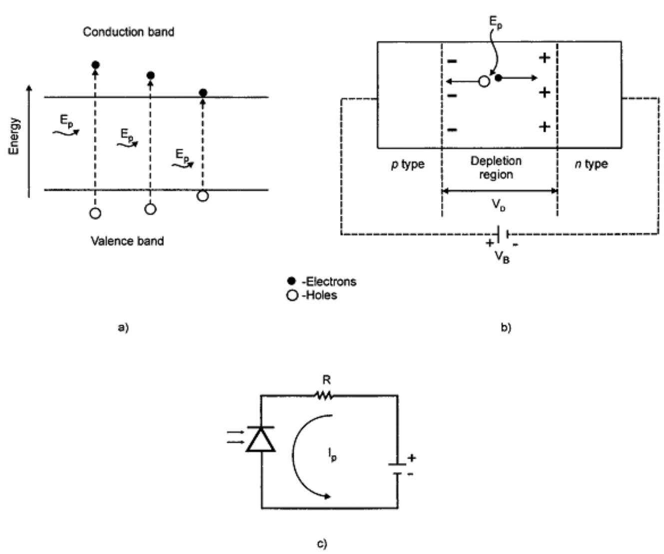

Figure 11.1 A *p-n* photodiode—the principle of operation: (a) Energy-band diagram; (b) *p-n* junction; (c) electrical circuit.

Thus, a flow of charge carriers—current—is generated. Applying external voltage (reverse bias) enhances the flow of electrons and holes. Observe closely how the external battery is connected in Figure 11.1(b) to create *reverse bias.*

Let's summarize what we have discussed thus far: *External photons—that is, light—strikes the semiconductor and separates the electrons and holes. The flow of these free charge carriers* *produces current. External voltage (reverse bias) enhances this effect.* The electrical circuit of a photodiode is shown in Figure 11.1(c).

It is important to compare Figure 11.1 with Figures 9.2 and 9.3 to see the similarities and differences between the operations of a light source and a photodiode.

Input-output characteristic

The input for a photodiode is light power (*P*); the output is current, which is usually called *photocurrent* (I_p) because it is caused by light. It follows from the principle of operation that the more photons that strike the active area of a PD, the more charge carriers will be created; that is, the greater will be the photocurrent. Thus, I_p is proportional to *P*:

$$I_p = R\,P, \qquad (11.1)$$

where *R* is constant. This relationship is shown in Figure 11.2(a).

The slope of this graph is one of the major PD parameters and is called *responsivity, R (A/W).* It is defined by the following formula:

$$R(A/W) = I_p/P \qquad (11.1a)$$

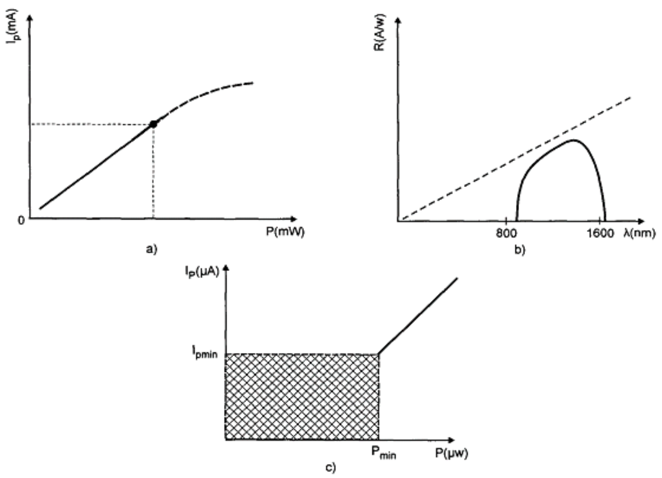

Figure 11.2 Responsivity of a photodiode: (a) Input-output characteristic; (b) responsivity vs. wavelength; (c) dark-current sensitivity.

Typical values of *R* range from 0.5 A/W to 1.0 A/W. This characteristic shows how efficiently a photodiode does its main job—converting light into an electrical signal. Since the value of *R is* provided by photodiode manufacturers, you can calculate the output photocurrent with the given light-input power.

How far does this graph keep its linearity? It seems reasonable to expect that for a high level of light power—that is, when there are a tremendous number of photons per unit of time striking the PD—all available electron-hole pairs will be involved in producing photocurrent; therefore, one can expect to see a *saturation* effect. This is shown by the dotted line in Figure 11.2(a). Obviously, for a saturation region, Formula 11.1 cannot be applied.

Example 11.1.1

Problem:

The responsivity of a photodiode is 0.85 A/W and the input-power saturation is 1.5 mW. What is the photocurrent if the incident light power is (1) 1 mW? (2) 2 mW?

Solution:

1. For input power 1 mW, we can apply Formula 11.1 and get

$$I_p = R\,P = 0.85 \text{ mA}$$

2. For input power 2 mW, Formula 11.1 is not valid, so we cannot find the photocurrent value.

Power Relationship

We usually evaluate any communications device by looking for its response to the amplitude and the frequency of the input signal. The first parameter (the device's response to amplitude) describes the power input–output relationship, while the second (the response to the frequency of the input signal) tells us about the device's bandwidth. Thus, the photocurrent recounts all the power relationships in a photodiode.

Responsivity versus wavelength

Responsivity actually represents the power input–output characteristic of a photodiode, as Formula 11.1 states. You may wonder whether responsivity depends on the operating wavelength. The following simple derivation provides the answer. Responsivity, by definition, is equal to I_p/P. But photocurrent is the number of electrons, N_e, flowing per unit time, that is,

$$I_p = N_e/t \qquad (11.2)$$

On the other hand, light power is light energy per unit of time, where light energy is equal to the energy of a photon (E_p) times the number of photons (N_p). Thus, we can write:

$$P = (N_p E_p)/t \quad (11.3)$$

Substitute $E_p = hc/\lambda$ and divide I_p by P:

$$R = I_p/P = (N_e/N_p)(\lambda/hc) \quad (11.4)$$

The ratio of the number of produced electrons, N_e, to the number of falling photons, N_p, shows how efficiently the semiconductor material converts light into current. This ratio is called the *quantum efficiency of a photodiode*, η. (Sound familiar? It should, because we introduced and discussed this concept and a similar term in Chapters 9 and 10.) Thus,

$$\eta = N_e/N_p \quad (11.5)$$

The quantum efficiency of a regular communications photodiode ranges from 50% to almost 100%.

If we recall that the product $h \times c$ is the constant and is equal to 1248 (eV•nm), then Formula 11.4 takes the following simple form:

$$R(\text{A/W}) = (\eta/1248)\lambda(\text{nm}) \quad (11.6)$$

Thus, theoretically, the graph "Responsivity vs. Wavelength" should be a straight line, as the dotted line in Figure 11.2(b) shows. The slope of this line is equal to η/1248 (when λ is expressed in nm).

The question that comes up at this point is, why is responsivity proportional to wave-length? If you closely examine the course of our derivation of Formula 11.6, you'll see that responsivity is inversely proportional to light power; the latter, in turn, is proportional to the number of photons and the energy of an individual photon, and this energy is inversely proportional to the wavelength. Thus, the longer the wavelength, the greater the number of photons needed to provide a certain amount of light power. Of course, increasing the number of photons also generates more electrons to produce more current. To sum all this up in a nutshell, then, *the longer the wavelength, the greater the amount of current produced from the same amount of light power.* If you now look at the definition of responsivity given in Formula 11.1, you will understand why $R \sim \lambda$.

Example 11.1.2

Problem:

What is the responsivity of an InGaAs photodiode if its quantum efficiency is equal to 70%?

Solution:

We can find R by using Formula 11.6 but we need to know the wavelength. Table 9.1 shows that the energy gap of InGaAs is equal to 0.75 eV, which corresponds to a wavelength of 1664 nm. Thus, from Formula 11.6 we find:

$$R = (\eta/1248)\lambda = 0.933 \text{ A/W}$$

This is a good number to obtain (each milliwatt of light power results in almost one milliampere of photocurrent) but, from a practical standpoint, it will turn out to be a little lower.

The real course of curve $R = f(\lambda)$ is shown in Figure 11.2(b) as a solid line. As you can see, in reality this graph is very far from the expected straight line. In fact, the graph shows *short* and *long cutoff wavelengths.* Let's discuss the reasons for this discrepancy.

As you can see from Figure 11.1(a) there is a cutoff wavelength, λ_c, determined by the energy gap, E_g, so that $E_g = hc/\lambda_c$. For wavelengths longer than λ_c, the energy of the photons is less than E_g; consequently, those photons will travel through this material without interaction. In other words, *for a given semiconductor material—that is, for a given energy gap—the photodiode can detect only wavelengths* $\lambda < \lambda_c = hc/E_g$. Looking at Table 9.1, you can appreciate what values of wavelength we are discussing. The most popular materials used in photodetectors are Si with $\lambda_c \sim 1100$ nm and InGaAs with $\lambda_c \sim 1700$ nm. Si PDs are used in the first transparent window (around 850 nm), while InGaAs PDs are used in the second (around 1300 nm) and third (around 1550 nm) transparent windows and even higher. Thus, the cutoff wavelength, λ_c, determines the longest wavelength a PD can detect. This is why responsivity goes to zero at the longer wavelengths in Figure 11.2(b).

Why responsivity depends on wavelength can be further clarified by using the following approach [1]: Light falling on the active area of a photodiode is partially absorbed and partially transmitted. Assuming the common exponential dependence of absorption, we can write:

$$P_{abs} = P_{in}(1 - \exp[-\alpha_{abs} w]), \quad (11.7)$$

where P_{in} and P_{abs} are incident and absorbed power, respectively, α_{abs} is the *absorption coefficient,* and w is the width (thickness) of the PD's active regions. When the absorption coefficient goes to zero, $P_{abs} \to 0$; when α_{abs} goes to infinity, $P_{abs} \to P_{in}$. In the case of a photodiode, each *absorbed* photon creates an electron; hence, the quantum efficiency is given by

$$\eta = P_{abs}/P_{in} = 1 - \exp(-\alpha_{abs} w) \quad (11.8)$$

Thus, quantum efficiency is not the constant, as was previously assumed, but the variable. As one can read from Formula 11.8, the bigger the product ($\alpha_{abs} w$), the closer η is to its maximum. It follows from the physics of absorption (see Figure 2.10) that the absorption coefficient is the function of wavelength; the graph exhibiting this dependency is shown in Figure 11.3. All these considerations elucidate the cause of the long-wavelength cutoff.

For wavelengths much *shorter* than the semiconductor's bandgap, photons will strike electrons at the valence band far from the energy-gap edge and the probability of exciting these electrons at the conduction band is very low. This is the main cause of the short-wavelength cutoff.

If you summarize all these considerations, you'll understand the reason for the discrepancy between the linear and the real graphs in Figure 11.2(b).

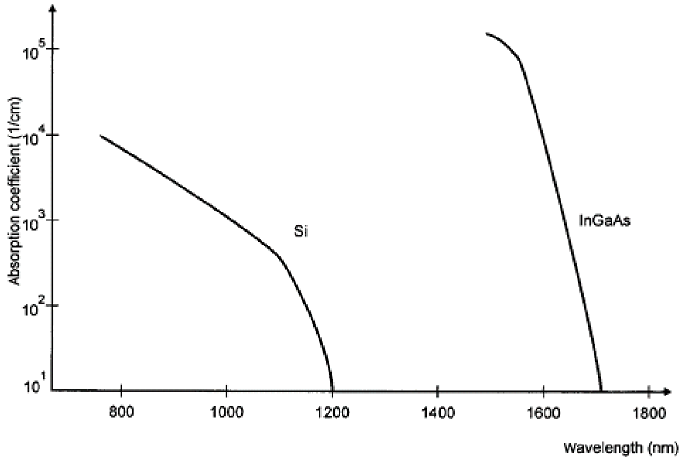

Figure 11.3 Absorption coefficients as a function of wavelength.

Example 11.1.3

Problem:

What width (thickness) of a depletion region of an InGaAs PD do we need to make its quantum efficiency 70%?

Solution:

The solution to this problem is, obviously, given by Formula 11.8. Solving Formula 11.8 with respect to *w*, we find:

$$w = [\ln(1-\eta)]/(-\alpha_{abs}) \qquad (11.8a)$$

To compute *w*, we need to know α_{abs}. We can use Figure 11.3, where, for InGaAs, α_{abs} is approximately 1×10^5 (1/cm) at the 1600-nm operating wavelength. Thus, we find *w* is equal to about 12 μm for InGaAs.

It is apparent that for a given absorption coefficient, *a wider (thicker) depletion region results in a higher quantum efficiency.* And this is an extremely important point. We need to make the depletion region wider to increase the quantum efficiency. This statement simply follows from the principle of operation of a PD. Indeed, the wider the depletion region, the greater the probability that most of the incident photons will fall here and thus the greater the likelihood they will be absorbed to create photocurrent.

(Note that InGaAs is much more efficient than Si as a material for a photodiode. One reason is that InGaAs is a direct-bandgap semiconductor, while Si is an indirect-bandgap material. Refer to the discussion of direct- and indirect-bandgap materials in Section 10.1.)

Photovoltaic and photoconductive modes of operation

Let's return to Figure 11.1. Observe that a photodiode can produce current without bias voltage because light conveys the external energy necessary to excite electrons at the conduction band (or to separate electrons from holes, if you prefer) and the depletion voltage (V_D) makes them flow. This mode of operation is called *photovoltaic*. It is how solar panels convert sunlight into electrical power. (Shaped like wings, these large panels, which contain a huge number of photocells, are wrapped around space satellites, with the photocells producing the photocurrent necessary to supply the satellite with electric power.) If external voltage is applied, the photodiode operates in the *photoconductive* mode. These two terms—*photovoltaic* and *photoconductive*—define the meaning of both operations: Without biasing, a photodiode works as the source of an electrical signal; with biasing, it's a good conductor of current originated by incident light. But remember: *A photodiode is actually a current source, with or without bias*.

Advantages of reverse biasing

For all practical purposes, we always use biasing because it dramatically improves the response of a photodiode. Without incident light, the depletion region of a photodiode does not contain free charge carriers (all electrons and holes are recombined, which is why we have a depletion region), whereas the *n* and *p* regions of a semiconductor have mobile charge carriers that are ready to flow. Hence, nearly all the bias voltage drops across the depletion region because this zone doesn't conduct. As soon as the incident photon creates electron-hole pairs, this voltage (or electric forces, to put it another way) helps to separate these free charge carriers and quickly removes them from the depletion region, thus generating photocurrent. This is the first—and the major—advantage of using reverse biasing.

What happens if the incident photon does not strike a depletion region but, rather, the *n* or *p* regions of a semiconductor? This can also create a free charge carrier, but the electric forces in these regions are weak so they will remove the electrons and holes there very slowly. Thus, a photogenerated electron-hole pair is separated by reverse voltage quickly and efficiently in the depletion region, but this separation occurs very slowly and inefficiently in the *p* or *n* regions because of the weakness there of electric forces. This is the second advantage of using reverse biasing. Incidentally, the photocurrent created in the depletion region is called *drift current.* Photocurrent created in *n* or *p* regions is called *diffusion* current.

A puzzling thought may cross your mind at this point: "If electrons and holes created by the incident photons and separated by reverse voltage have to drift through the depletion region before they reach the wire to flow to the battery, why don't they recombine again and radiate a photon?" Good question. Theoretically, electrons and holes can recombine again but, in reality, the loss of charge carriers due to secondary recombination is negligibly low. This is because the reverse voltage sweeps them from the depletion region faster than they can recombine again. In other words, the separation time of these carriers due to applied voltage is much less than their recombination lifetime. Thus, we have the third advantage of using reverse biasing.

The last, but not the least, advantage of reverse biasing is its ability to eliminate what's called *dark current.* Without incident light, some free charges in the depletion region can be created mostly by external thermal energy (temperature). The flow of these charges creates dark current, I_d. In other words, *dark current is current generated by a photodiode without light.* Clearly, dark current is a detrimental phenomenon because it eventually determines the minimum light power that can be detected, that is, a photodiode's *sensitivity.* How does reverse biasing help here? Since all voltage is applied across the depletion region, any free charge carriers that are occasionally created without light will be swept away by the reverse-bias voltage. This means that reverse biasing controls dark current.

So, from a practical standpoint, reverse biasing improves a photodiode's linearity, increases its speed and efficiency of operation, and reduces its dark current. All these advantages will be clarified in the course of this chapter.

Dark-current sensitivity

Sensitivity is the key parameter determining the quality of a photodiode. As noted above, sensitivity refers to the minimum light power that a given photodiode can detect. It is measured in watts (in microwatts, actually) or in dBm, which is more common.

Here we're discussing only sensitivity determined by dark current, which is depicted in Figure 11.2(c). As you can see from this figure, there is an area of uncertainty around zero-input power in the I_p-*P* graph. This is because some current flows through a PD's circuit, but we don't know whether it is dark current or photocurrent. Thus, until some minimal light power (P_{min}) truly generates photocurrent, we cannot rely on the output of a photodiode. As an example, it is easy to calculate that for *R = 1* A/W, dark-current sensitivity is 5 nW for I_d = 5 nA.

The value of I_d = 5 nA at room temperature is typical for modern photodiodes. It is apparent from this discussion that dark current increases with temperature but it is still not more than 50 nA, typically, at $T \leq 70$°C. Compare this value with the value for photocurrent: If *R = 1* A/W and input light power is 0.1 μW, then
I_p = 100 nA. Thus, dark current is of concern at this typical level of photodetection.

Dark-current sensitivity is the major concern with photodetectors used in measuring devices such as power meters. For communication PDs, it is not a major issue. A more general means by which to evaluate a photodiode's sensitivity is noise description. We develop this approach later in this chapter.

Power digest

A *p-n* photodiode converts light power into electric current. The efficiency of this conversion (1) diminishes at the air-semiconductor interface, where light is reflected, (2) decreases where photogenerated electrons and holes undergo a secondary recombination, and (3) increases within the active region, where light is better absorbed. Applying an antireflecting coating over the surface of the photodiode and using an angled fiber tip, we can resolve the reflection problem.

A widening depletion (active) region is the solution to two other problems. For instance, where power consideration is a major factor, we need a wide depletion region, a place where photons are absorbed, in order to achieve high quantum efficiency, which, in turn, provides high responsivity. But the width of a depletion region in a *p-n* junction photodiode is determined by the reverse voltage ($w \sim \sqrt{V}$) because the higher the reverse voltage, the more depleted the region around the *p-n* junction becomes. It might look, then, as though we need to apply high reverse-bias voltage to enhance the power response of a photodetector. But before jumping to any conclusion about the level of the reverse-bias voltage needed, we have to consider bandwidth.

Bandwidth

Bandwidth, in terms of our current discussion, can be defined as the *maximum frequency, or bit rate, that a photodiode can detect without making essential errors*. (Again, strictly speaking, the term *bandwidth* [Hz] is applied only to analog signals; for digital transmission, we have the term *bit rate* [bit/s]. This said, it has become common practice today to use the term *bandwidth* to encompass both analog and digital technologies.)

There are two basic mechanisms restricting bandwidth in a photodiode. The first restriction stems from the fact that charge carriers created by a photon need some time to be collected. This time is often called *transit time,* τ_{tr}. If we denote the maximum drift velocity of the charge carriers as v_{sat}, then for a depletion region with thickness *w*, transit time can be estimated as:

$$\tau_{tr} = w/v_{sat}\,, \tag{11.9}$$

where v_{sat} is saturation velocity. With typical values of $w \sim 10$ μm and $v_{sat} \sim 10^5$ m/s, we can compute , $\tau_{tr} \sim 100$ ps.

The second restriction on bandwidth derives from the inherent capacitance of a *p-n* photodiode (C_{in}). Indeed, a *p-n* junction can be considered as two charged plates isolated by a depletion region. This is the classical model of a capacitor. Hence, inherent capacitance is equal to

$$C_{in} = (\epsilon A)/w, \tag{11.10}$$

where ϵ is the permitivity of a semiconductor and *A* is the active area (the photosensitive area) of the photodiode. This capacitance is parallel to the output of the photodiode.

To better understand this discussion, consider the equivalent circuit of a *p-n* photodiode. (See Figure 11.4 [2], [3].) Here the diode stands for an ideal diode operation; the current source (I_p) represents the flow of the photogenerated carriers; R_j and R_S correspond to the junction (shunt) and series resistance of a photodiode, respectively, and they form internal resistance R_{in}; C_{in} is defined by Formula 11.10; and R_L is the load resistance.

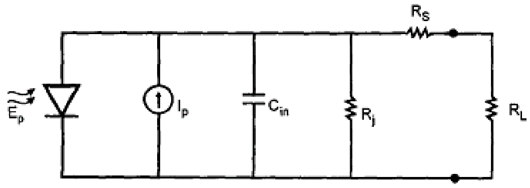

Figure 11.4 Equivalent circuit of a *p-n* photodiode. ***(Adapted from Diode Lasers & Instruments Guide, Melles Griot, Boulder, Colo., April 1998. Used with permission.)***

The junction (shunt) resistance (*R*j) is the resistance of a photodiode's depletion region; by its very nature, this resistance is extremely high (from units to tens of MΩ). However, manufacturers can control the value of *R*j Resistance of the *p* and *n* layers and of the electrical contacts acts as a series resistance between the active region and the load circuit of a photodiode. We denote this resistance as R_S. The series resistance is very small, ranging from 5 to 10 Ω. So, from the point of view of the RC circuit, it is the load resistance (R_L) that determines the timing properties of a photodiode.

Let's return to the time constant (τ_{RC}) induced by a capacitor. It follows from the above consideration that

$$\tau_{RC} = (R_s + R_L)C_{in} \approx R_L C_{in} \tag{11.11}$$

Typical values of C_{in} are from 1 to 2 pF; hence, for R_L = 50 Ω, τ_{RC} is in the range of 50 to 100 ps. Thus, the bandwidth of a photodiode is given by the following formula:

$$BW_{PD} = 1/[2\pi(\tau_{tr} + \tau_{RC})] \qquad 11.12$$

Example 11.1.4

Problem:

A p-n photodiode has τ_{tr} = 100 ps and τ_{RC} = 100 ps.

a. What is its bandwidth?

b. Can we increase the bandwidth of a *p-n* photodiode by varying the thickness of the depletion region?

c. Can we increase the bandwidth by varying its load resistance?

d. What is the role of the active area of a photodiode?

Solution:

a. It comes directly from Formula 11.12 that BW = 0.796 Gbit/s.

b. We need to decrease both τ_{tr} and τ_{RC} . To decrease τ_{tr}, we have to increase the thickness of the depletion region, *w*. (We can do almost nothing about drift velocity.) To decrease τ_{RC} , we need to decrease *w*, as you can see from Formulas 11.11 and 11.10. To clarify the point, let's rewrite Formula 11.12 in explicit form with respect to *w*:

$$BW_{PD} = 1/\{2\pi[(w/v_{sat}) + R_L(\epsilon A/w)]\}, \tag{11.13}$$

where $R_{in} = R_S + R_L \approx R_L$, as discussed above. If we rewrite Formula 11.13 in a different form,

$$BW_{PD} = (v_{sat}/2\pi)[w/(w^2 + v_{sat}R_L\epsilon A)], \tag{11.13a}$$

we see that the thickness of the depletion region (*w*) appears simultaneously in the numerator and the denominator of a bandwidth formula. Thus, one has to find a compromise for the value of *w* to achieve the optimal value of the bandwidth. Toward this end, let's take the derivative $\partial BW_{PD}/\partial w$:

$$\partial BW_{PD}/\partial w = (v_{sat}/2\pi)[(-w^2 + v_{sat}R_L\epsilon A)/(w^2 + v_{sat}R_L\epsilon A)^2] \tag{11.13b}$$

Thus, the optimal thickness found from the condition derivative $\partial BW_{PD}/w = 0$ is given by

$$w_{opt} = \sqrt{(v_{sat}R_L\epsilon A)} \tag{11.13c}$$

All parameters under the square root, except R_L, are predetermined by the semiconductor material and the fabrication process of a *p-n* photodiode. Therefore, in a *p-n* photodiode we can control the thickness of the depletion region (*w*) only by reverse-bias voltage. But the value of reverse-bias voltage is determined mostly by the receiver package. (You don't want to have several different voltages in a small receiver chip, do you?) The fact is we don't have very much freedom when it comes to increasing the bandwidth of a *p-n* photodiode.

c. Load resistance (R_L) determines τ_{RC} , as Formula 11.11 shows. This is why R_L appears in Formula 11.13c. It seems, then, that we need to decrease the value of this resistance to increase the bandwidth. But we cannot do so because, first, R_L must be at least ten times more than R_s (see Figure 11.4) and, secondly, to minimize noise we need to increase the load resistance. (This topic is discussed in Sections 11.3 and 11.4.)

d. The active area (*A*) of a photodiode plays two roles. From the power standpoint, it's better to have this area large because it allows a PD to gather more light. (It is common practice to specify *A* through the term *active-area diameter*, D_A.) Ultra-sensitive photodiodes for special (noncommunications) applications can have a huge active area. For example, large-area PDs can have a D_A from 1 to 16 mm [4].

From the bandwidth standpoint, the active area must be small, as Formula 11.13 says. Indeed, the larger the active area, the higher the internal capacitance, which results in a large *R-C* time constant. (See Formulas 11.10, 11.11, and 11.12 above.) To fabricate very wide bandwidth photodiodes, manufacturers have to decrease the active-area diameters. To achieve a bandwidth up to 50 GHz, the manufacturer restricts the active-area diameter to 10 μm [5]. It's interesting to observe how a PD's capacitance changes with a change in the active-area diameter. For example, an InGaAs PIN photodiode exhibits the following relationships between active-area diameter and diode capacitance [4]:

Active-area diameter, D_A (μm)	60	80	100	150	300
Diode capacitance, *C*(pF)	0.2	0.75	1.5	2	7

You can easily draw your own conclusion as to how the bandwidth of this photodiode changes in response to a rise in active-area diameter.

And so once again we confront a trade-off between power and bandwidth requirements. But don't forget: We need to couple light from an optical fiber into the active area of a photodiode. Therefore, we cannot make D_A any arbitrary small value. In practice, you will find photodiodes used in fiber-optic communications technology with active-area diameters ranging from units of μm to hundreds of μm.

The above example shows that bandwidth is at its maximum when the thickness of a depletion region (*w*) is optimal. On the other hand, the above discussion shows that the power efficiency is proportional to *w*. In other words, *there is a trade-off between power and bandwidth efficiencies in a* p-n *photodiode.* We have to remember this fact because it shows that a *p-n* photodiode is not a very efficient device for communication purposes.

There is a third mechanism restricting the bandwidth of a *p-n* PD. Since the *p* and *n* regions are wider than the depletion region, many photons strike those regions, creating a diffusion photocurrent, as noted in the discussion of the advantages of reverse biasing. But diffusion flow is very slow compared with the drift flow of charge carriers. Therefore, the electric-output pulse of a *p-n* photodiode will be much wider than the optical-input pulse and its tail will be determined by the diffusion photocurrent. This is illustrated in Figure 11.5, where all the causes of the widening of the electric-output pulse are shown qualitatively.

Bandwidth and *p-n* photodiode digest

Increasing the bandwidth efficiency of a *p-n* photodiode requires a wide depletion region to reduce the diffusion current. Hence, it looks as though we need to increase the reverse bias because this voltage determines the width of a depletion region. Indeed, this is exactly what we need to do to increase the power efficiency. On the other hand, however, taking into account the transit and RC time constants, one has to find the appropriate reverse-biasing voltage to optimize the bandwidth of the photodiode. We cannot choose this voltage arbitrarily. For one thing, we need to remember that a photodiode is a part of a receiver, where low-voltage electronics is used.

The key point is simply this: *We need to increase the width of a depletion region without manipulating unnecessarily the value of the reverse-bias voltage.* The solution to this dilemma is a *p-i-n* photodiode.

p-i-n Photodiodes

The basic structure of a *p-i-n* PD is shown in Figure 11.6. The major feature of this photodiode is that *it consists of a thick, lightly doped* intrinsic *layer sandwiched between thin p and* n *regions.* The word *intrinsic*, in semiconductor-industry parlance, means “natural,” “undoped.” Thus, we now have the full meaning of the letters *p-i-n*: positive-intrinsic-negative.

There are two major types of *p-i-n* photodiodes: front-illuminated (Figure 11.6[a]) and rear-illuminated (Figure 11.6[b]). In a front-illuminated PD, light enters the hole through the top contact. To reduce backreflection of the incident light, the active surface is covered by an antireflection coating. Then light passes through the thin *p* region and generates electron-hole pairs in the thick intrinsic layer.

In a rear-illuminated PD, light enters the active region through a heavily doped *n*+ layer. This layer is transparent to the incident light because its energy gap is larger than the energy of incident photons. All other processes are similar to those that take place in the front-illuminated PD.

To list the advantages of a *p-i-n* PD, we need simply to recall the drawbacks of a *p-n* pho-todiode. The first and the major feature of a *p-i-n* PD is that its intrinsic layer is its depletion layer, where the absorption of photons occurs. Since the intrinsic layer is naturally thick, most of the

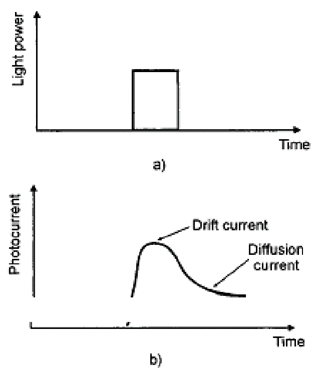

Figure 11.5 Input and output pulses of a photodiode: (a) Optical input pulse; (b) electric-output pulse.

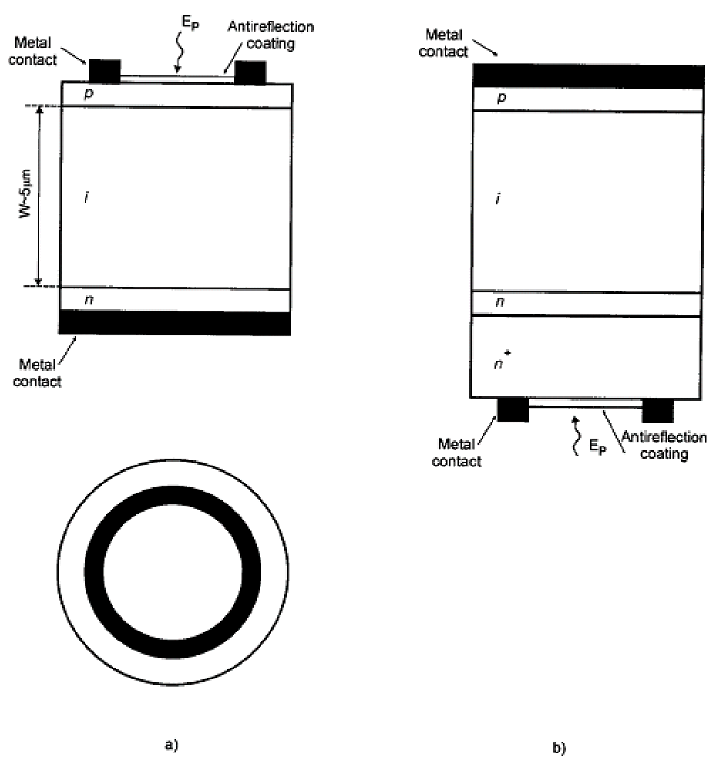

Figure 11.6 *p-i-n* photodiode: (a) Front-illuminated PD; (b) rear-illuminated PD.

incident photons enter this layer and generate electron-hole pairs. This action results in the high quantum efficiency of this device. In addition, there is no need for tinkering with reverse voltage to increase the width of an absorbing layer. This is why both the power and bandwidth efficiencies are high, which is the major advantage of a *p-i-n* photodiode.

All other advantages are simply the consequences of the structure of a *p-i-n* PD. For example, since the intrinsic layer contains almost no free charge carriers, the electric field across this layer is large. The result is the efficient separation of electrons and holes generated by the incident photons. In addition, this field (reverse biasing, remember) decreases the dark current by sweeping away all thermally generated charge carriers. What’s more, the diffusion current in a *p-i-n* PD is very small because the *p* and *n* layers are extremely thin compared with the intrinsic layer. Furthermore, incident photons are much more likely to enter the intrinsic (active) region than the *p* or *n* regions. The result is an increase in the bandwidth efficiency of the photodiode. One other advantage of a *p-i-n* PD is that the reverse-biasing voltage is small (usually, 5 V) because the thickness of the depletion region is controlled by the thickness of the intrinsic layer, not by reverse voltage.

There is one problem, however, that the *p-i-n* structure cannot resolve: the width of the intrinsic layer. Widening this layer will result in an increase in power efficiency but a decrease in bandwidth efficiency because of a rise in transit time $\tau_{tr} = w/v_{sat}$. (See Formula 11.9.) Hence, some compromise must be found and the key to this compromise is given in Figure 11.3. A photodiode made from Si has to have a wider intrinsic layer because the absorption coefficient (α_{abs}) at its operating wavelength (near 850 nm) is a little more than 10^3 1/cm, while InGaAs PDs have $\alpha_{abs} \sim 10^5$ 1/cm at λ = 1600 nm and, therefore, can have a smaller *w*. In practice, Si *p-i-n* photodiodes are fabricated with the width of an intrinsic layer on the order of 40 µm, while InGaAs PDs have a *w* around 4 µm. This results in different bandwidths for Si and InGaAs photodiodes.

Example 11.1.4

Problem:

What are the bandwidths of Si and InGaAs *p-i-n* photodiodes?

Solution:

The solution to this problem is obviously given by Formula 11.12 or Formula 11.13 but we need to know several specific parameters to compute the numbers. Instead of delving into reference sources in search of these parameters, we are better off observing that $\tau_{tr} \sim w$ and $\tau_{RC} \sim 1/w$. Hence, for a wide intrinsic layer, which is exactly the case here, we can assume $\tau_{tr} \gg \tau_{RC}$.

With this assumption, Formulas 11.12 and 11.13 take the following form:

$$BW_{PD} = 1/(2\pi\tau_{tr}) = 1/[2\pi(w/v_{sat})] \qquad (11.14)$$

In Example 11.1.3 we used $v_{sat} \sim 10^5$ m/s. Thus, taking *w* = 40 µm for Si PD and *w* = 4 µm for InGaAs PD, we obtain τ_{tr} (Si) = 400 ps and τ_{tr} (InGaAs) = 40 ps, which results in:

$$BW_{Si} = 0.398 \text{ Gbit/s}$$
$$BW_{InGaAs} = 3.98 \text{ Gbit/s}$$

From Figure 11.3 and this example, one can see the areas of photodiode applications: A silicon PD is used in the first transparent window (near 850 nm), where relatively low-speed networks operate, while an InGaAs PD is suited for the second and third windows (near 1300 nm and 1550 nm, respectively), where high-speed networks operate.

Further improvement in the efficiency of a *p-i-n* operation can be accomplished by fabricating the photodiode in a double heterostructure, one similar to that used in LEDs and LDs. In fact, if you make the *n* and *p* regions of the diode shown in Figure 11.4 from InP and the intrinsic layer from InGaAs, your goal is achieved: All photons at operating wavelength will pass the *n* and *p* layers without interaction. This dramatically improves the quantum efficiency and eliminates the diffusion-current problem.

A p-i-n *photodiode is the most commonly employed light detector in today's fiber-optic communications systems because of its ease in fabrication, high reliability, low noise, low voltage, and relatively high bandwidth.* Many efforts have been made to improve its characteristics, in particular the bandwidth, where up to 110 GHz has been achieved [1].

Avalanche Photodiodes (APDs)

What do we ultimately want from a photodiode? One of its major parameters is sensitivity—the minimum light power a PD can detect. This parameter determines the length of a fiber-optic link imposed by a power limitation. Indeed, refer to Section 8.4, where the power budget of a fiber communications link is considered. The more sensitive the photodiode, the longer the link a designer can afford to have with given losses. It seems that the remedy for this problem is quite apparent: Use an amplifier to magnify the photocurrent produced by the photodiode. In fact, a receiver circuit always includes an amplifier, as will be explained in the following sections. But an amplifier, as is true of any electronic circuit, introduces its own noise, thus reducing sensitivity. So, you are probably thinking, if only we were able to amplify photocurrent without an external amplifier and, therefore, without the noise associated with this circuitry! Well, we can; in fact, this is why the *avalanche photodiode (APD)* was invented.

Power consideration

The basic mechanism of an avalanche PD is as follows: A special *p-i-n* structure of a photodiode is used. Incident photons generate primary electrons and holes, as they do in a regular *p-i-n* PD. Relatively high (around 20 V) reverse voltage is applied to the photodiode. This voltage accelerates photogenerated electrons and holes, which thereupon acquire high energy. These electrons and holes strike neutral atoms and separate other bonded electrons and holes. These secondary carriers gain enough energy to ionize other carriers, causing a so-called avalanche process of creating new carriers. Thus, one photon eventually generates many charge carriers, which means this photodiode internally amplifies photocurrent. This is equivalent to saying the APD's quantum efficiency is more than 1 (typically, it is from 10 to 100).

Referring to an energy-band diagram, one can say that primary photogenerated electrons strike bonded electrons at the valence band and cause them to rise to the conduction band; holes are left at the valence band to complete the picture of how secondary carriers are generated.

The process of creating many secondary carriers is called *impact ionization.* The basic diagram of an APD is shown in Figure 11.7.

Photons pass through the heavily doped p^+ region and enter the intrinsic layer, where they produce electron-hole pairs. Reverse voltage separates photogenerated electrons and holes and moves them toward the *pn*+ junction, where a high electric field (on the order of 10^5 V/cm) exists. This electric field accelerates the charge carriers, resulting in impact ionization.

The major advantage of an avalanche photodiode over a *p-i-n* PD is clear from the physics of its operation: The quantum efficiency of the APD is *M* times larger than that of a *p-i-n* PD. (*M is* called the *multiplication, or gain, factor.*) Indeed, an APD produces *M* charge carriers in response to one photon. Thus, referring to Formula 11.6, one can write:

$$R_{APD} = MR_{p\text{-}i\text{-}n} = M(\eta/1248)\lambda \qquad (11.15)$$

(Don't forget that the wavelength has to be measured in nm.)

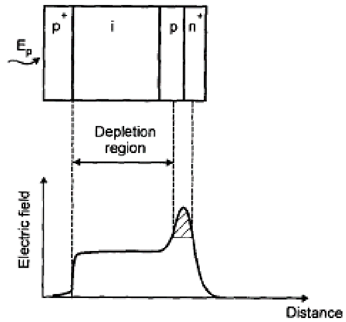

Figure 11.7 Avalanche photodiode (APD).

Even though, theoretically, both electrons and holes can be involved in the ionization process, from a practical standpoint, an APD works better when only one type (usually electrons) is used. Thus, in the above explanation you can always substitute the word *electrons* where you see *charge carriers.*

The multiplication factor (*M*) depends on the accelerating voltage, the thickness of the gain region, and the ratio of electrons to holes participating in the ionization process. This implies that you can control the gain of an APD by varying the reverse voltage. *M* values range from 10 to 500.

What we have to keep in mind is that since the ionization process is essentially random, so too is the multiplication factor. Thus, when concerned with the value of *M*, one works with an average number. It also follows from the physics of the avalanche process that it is noisy. However, this doesn't nullify the major advantage of an APD: internal amplification of photocurrent without the noise associated with external electronic circuitry.

APD bandwidth

Consideration of the bandwidth of an avalanche photodiode requires a different approach from that used when considering the bandwidth of a *p-i-n* PD. Since an APD introduces amplification, the most universal characteristic of such a device is the *gain-bandwidth product: M* × BW. For a typical APD, the gain-bandwidth product can be evaluated by [1]

$$M \times \mathrm{BW} = 1/(2\pi\,\tau_e), \tag{11.16}$$

where *M is* zero-frequency gain and τ_e is effective transit time equal to

$$\tau_e = k_A \tau_{tr} \tag{11.17}$$

Here, τ_{tr} is transit time defined by Formula 11.9 and k_A *is the ratio of holes to electrons involved in the ionization process.* (Strictly speaking, k_A is the ratio of the impact-ionization coefficients of holes and electrons but, in a sense, it can be considered as simply the ratio of electrons to holes.) The ratio k_A depends on the semiconductor material and it is in the range of 0.03 for Si, 0.8 for Ge, and 0.6 for InGaAs. The assumption τ_{tr} » τ_{RC} is used in Formula 11.16.

To complicate the matter, the gain (the multiplication factor) of an APD depends on frequency [1]

$$M(\omega) = M/\sqrt{(1+(\omega\tau_e M)^2}, \tag{11.18}$$

where ω stands for radian frequency and *M is M*(0), as above.

The gain-bandwidth product is around 500 GHz for an Si APD and 120 GHz for an InGaAs APD. Since an Si APD has a gain as high as 500, its bandwidth is not more than 1 GHz, while an InGaAs APD has a typical gain of about 40, which yields a 3-GHz bandwidth.

As was the case for a *p-i-n* photodiode, we can conclude that Si APDs are useful for a moderate-speed (up to 1 GHz) fiber-optic network, which usually operates at 850 nm, while InGaAs APDs can be used in higher-speed fiber links (up to 3 GHz), which usually operate at 1300 nm and 1550 nm.

Overall, Si APDs demonstrate good performance characteristics from both power and bandwidth standpoints but they are inherently restricted to the use of the first transparent window, which is, for the most part, out of the commercial-application realm. InGaAs APDs—the devices for today's 1300-nm and 1550-nm wavelengths—have much worse power and bandwidth parameters. A major effort has gone into improving the characteristics of long-wavelength APDs. This has resulted in an increase in the gain-bandwidth product up to 150 GHz with a gain (*M*) of 10 [1].

Let's compare an APD's bandwidth of 15 GHz and gain of 10 with a *p-i-n* PD's bandwidth of 5 GHz and gain of 1. It is easy to conclude that an APD is at least 10 times more sensitive than a *p-i-n* PD with comparable bandwidth, which implies a 10-times-longer fiber-optic span between a transmitter and a receiver. But this advantage almost vanishes if you recall that an APD requires relatively high reverse voltage. From a practical standpoint, this means an increase in power consumption, implying less freedom for miniaturization of a receiver unit and, therefore, longer transmission lines with increasing noise and parasitic capacitance, not to mention the need for a separate power supply that is not compatible with other power units used in electronic circuits. So, when choosing a photodetector for your fiber-optic communications system, use a systems approach and take into account all the advantages and shortcomings of every type of device.

In conclusion, we have summarized in Table 11.1 the typical characteristics of *p-i-n* and avalanche photodiodes. These numbers give you a general idea of what today's PDs look like.

MSM Photodetectors

An MSM (metal-semiconductor-metal) is another type of photodetector used in fiber-optic communications. This is not a *p-n* junction diode; however, its basic mechanism of light-current con-version is still the same: Photons generate electron-hole pairs whose flow makes current. The basic structure of an MSM photodetector is shown in Figure 11.8.

A set of flat metal contacts is deposited on the surface of a semiconductor. These contacts are called fingers and they are biased alternately so that a relatively high electric field exists between the fingers. Photons strike the semiconductor material between the fingers and create electron-hole pairs, which are separated by the electric field; thus, electric current is created.

Since both electrodes and a photosensitive region are fabricated on the same side of the semiconductor, this structure is called *planar.* The advantage of this photodetector, as compared with both types of photodiodes, is that a planar structure results in low capacitance and, consequently, in higher bandwidth. Indeed, the MSM photodetector promises to work at 300 GHz. Ease of fabrication is another advantage of a planar structure.

Table 11.1 Typical Characteristics of *p-i-n* and avalanche photodiodes

Parameter	Symbol	Unit	Type	Material: Si	Material: Ge	Material: InGaAs
Wavelength	λ	nm		0.4–1.1	0.8–1.8	1.0–1.7
Responsivity	R	A/W	*p-i-n*	0.4–0.45	0.8–0.87	0.5–0.95
Quantum efficiency	η	%	*p-i-n*	75–90	50–55	60–70
APD gain	M	—	APD	—	50–200	10–40
Dark current	I_d	nA	*p-i-n*	1–10	50–500	1–20
			APD	0.1–1	50–500	1–5
Bandwidth	BW	GHz	*p-i-n*	0.125–1.4	0–0.0015	0.0025–40
			APD	—	1.5	1.5–3.5
Bit rate	BR	Gbit/s	*p-i-n*	0.01	—	0.1555–53
			APD	—	—	2.5–4
Reverse voltage*	V	V	*p-i-n*	50–100	6–10	5–6
			APD	200–250	20–40	20–30
k-factor	k_A	—	APD	0.02–0.05	0.7–1.0	0.5–0.7

*Note: The reverse voltages listed here reflect only the orders of magnitude. Actual numbers found in the field vary widely. For example, you will find InGaAs *p-i-n* PDs with reverse bias up to 30 V, Si *p-i-n* PDs with reverse bias less than 30 V, and high-speed InGaAs APDs with reverse bias around 50 V.

Sources: Govind Agrawal, *Fiber-Optic Communication Systems,* 2d ed., New York: John Wiley & Sons, 1997, and *Lightwave 1999 Worldwide Directory of Fiber-Optic Communications Products and Services,* March 31, 1999, pp. 62–66.

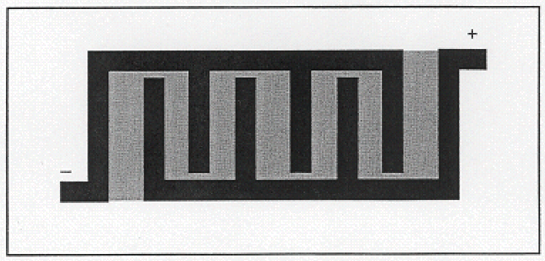

Figure 11.8 MSM photodetector. Metal contact is black and active area is shaded.

There is, however, a drawback to an MSM photodetector: its relatively low responsivity, which ranges from 0.4 to 0.7 A/W. What's more, an essential area of the semiconductor material is taken up by metal contacts, thereby reducing the active area of the device.

As you can well imagine, then, MSM photodetectors are an area of intensive research and development today, so we can expect a number of these devices to be competing commercially in the near future.