5 PHOTODETECTORS AND IMAGE SENSORS

5.1 PRINCIPLE OF THE *pn* JUNCTION PHOTODIODE

A. Basic Principles

Photodetectors convert an incident radiation to an electrical signal such as a voltage or current. In many photodetectors such as photoconductors and photodiodes this conversion is typically achieved by the creation of **free electron hole pairs** (EHPs) by the absorption of photons, that is, the creation of electrons in the conduction band (CB) and holes in the valence band (VB). In some devices, such as pyroelectric detectors, the energy conversion involves the generation of heat that increases the temperature of the device, which changes its polarization and hence its relative permittivity. We will first consider *pn* junction–based photodiode-type devices as these devices are small and have high speed and good sensitivity for use in various optoelectronics applications, the most important of which is in optical communications.

Figure 5.1 (a) shows the simplified structure of a typical *pn* junction **photodiode** that has a p^+n type of junction, that is, the acceptor concentration N_a in the *p*-side is much greater than the donor concentration N_d in the *n*-side. The illuminated side has a window, defined by an annular electrode, to allow photons to enter the device. There is also an **antireflection (AR) coating**, typically Si_3N_4, to reduce light reflections. The p^+-side is generally very thin (less than a micron) and is usually formed by planar diffusion into an *n*-type epitaxial layer. Figure 5.1 (b) shows the net space charge distribution across the p^+n junction. These charges are in the **depletion region**, or in the **space charge layer** (SCL), and represent the exposed negatively charged acceptors in the p^+-side and exposed positively charged donors in the *n*-side. The depletion region extends almost entirely into the lightly doped *n*-side and, at most, it is a few microns.

The photodiode is normally reverse biased. The applied reverse bias V_r drops across the highly resistive depletion layer width W and makes the voltage across W equal to $V_o + V_r$ where V_o is the built-in voltage. Normally V_r (*e.g.*, 5–20 V) is much larger than V_o (less than 1 V) and the voltage across the SCL is essentially V_r. The field E in the SCL is found by the integration of the net space charge density ρ_{net} in Figure 5.1 (b) across W subject to a voltage difference of $V_o + V_r$, *i.e.,* V_r. The field only exists in the depletion region and is not uniform. It varies across the depletion region as shown in Figure 5.1 (c), where it is maximum at the junction and

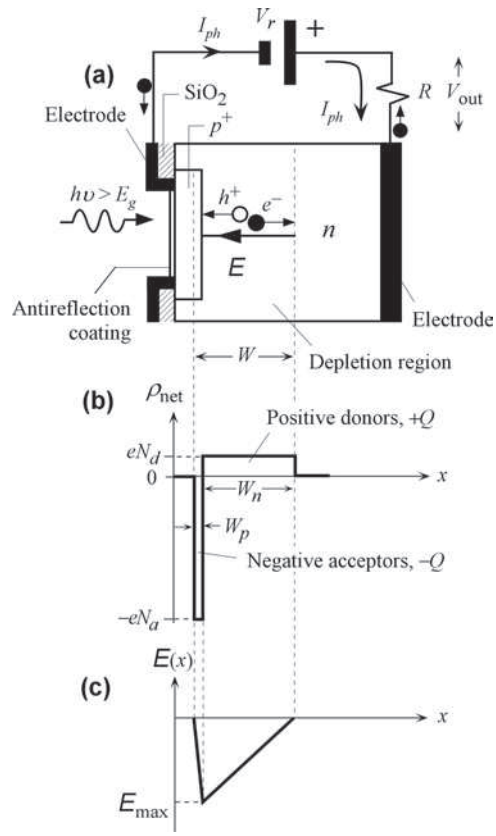


FIGURE 5.1 (a) A schematic diagram of a reverse biased *pn* junction photodiode. (b) Net space charge across the diode in the depletion region. N_d and N_a are the donor and acceptor concentrations in the *p*- and *n*-sides. (c) The field in the depletion region.

penetrates into the *n*-side. The regions outside the SCL are the **neutral regions** in which there are majority carriers.

When a photon with an energy greater than the bandgap E_g is incident, it becomes absorbed to **photogenerate** a free EHP, that is, an electron in the CB and a hole in the VB. Usually the energy of the photon is such that photogeneration takes place in the depletion layer. The field E in the depletion layer then separates the EHP and drifts them in opposite directions until they reach the neutral regions, as illustrated in Figure 5.1 (a). Drifting carriers generate a current, called the **photocurrent** I_{ph}, in the external circuit that provides the electrical signal. The photocurrent lasts for the duration it takes for the electron and hole to cross the SCL (width, W) and reach the neutral regions. As the electron drifts in the SCL toward the neutral *n*-side, there is an electron that has come out from *n*-side and is flowing around the external circuit toward the battery's positive terminal. Similarly, as the hole drifts in the SCL toward the *p*-side, there is an electron flowing in the external circuit from the battery's negative terminal to the *p*-side. When the drifting hole reaches the neutral p^+-region, it recombines with an electron entering the p^+-side from the external circuit. Similarly, when the drifting electron in the SCL reaches the neutral *n*-side, it has replenished the electron that had left the *n*-side for the battery. The photocurrent I_{ph} depends on the number of EHPs photogenerated and the drift velocities of the carriers while they are transiting the depletion layer. Since the field is not uniform and the absorption of photons occurs over a distance that depends on the wavelength, the time dependence of the photocurrent signal cannot be determined in a simple fashion.

As indicated in Figure 5.1 (b), the depletion layer width W_p in the p-side is very narrow whereas that on the n-side, W_n, is much wider because the amount of total charge in the SCL on the p-side must be the same as that on the n-side, that is, $N_a W_p = N_d W_n$. Thus, for the p^+n photodiode $W_p \ll W_n$ and $W \approx W_n$. It should be mentioned that the photocurrent in the external circuit is due to the flow of electrons only even though there are both electrons and holes drifting within the device. Suppose that there are N number of EHPs photogenerated. If we were to integrate the photocurrent I_{ph} to calculate how much charge has flowed, we would find an amount of charge that is due to the total number of photogenerated electrons (eN) and not due to both electrons and holes ($2eN$).

B. Energy Band Diagrams and Photodetection Modes

To obtain a better understanding of the *pn* junction detector, we need to consider the energy band diagram. Figure 5.2 (a) shows the *pn* junction reverse biased by V_r, which increases the built-in voltage to $V_o + V_r$. The field in the SCL also increases. The Fermi levels E_{Fn} and E_{Fp} on the n- and p-sides are separated by V_r. The potential hill, that is, the change in E_c from the E_c on the n-side to that on the p-side in the SCL is very steep due to the large field in the SCL. The absorption of a photon in the SCL creates an EHP. The electron and hole become separated and drifted by the field. The drift corresponds to the electron rolling down the energy hill (along E_c) toward the n-side, whereas the hole rolls down the energy hill toward the p-side. (Remember that the hole energy increases in the downward direction.) The drift creates a photocurrent I_{ph}, which is the quantity detected in the external circuit. It lasts for the duration of the drift of the electron and hole.

Photogeneration within a diffusion length to the SCL would also generate a photocurrent as illustrated in Figure 5.2 (b). An EHP is created in the p-side within the diffusion length L_e of an electron in this p-side. The electron can only reach the SCL by diffusion. The photogenerated electron diffuses (by "random walk") to the SCL where the internal field then drifts the electron over to the n-side. The drift creates the photocurrent. As the electron drifts in the SCL toward the neutral n-side, there is an electron flowing in the external circuit toward the battery's positive terminal; this electron came out from the n-side. The photogenerated hole in the p-side is neutralized by the flow of an electron from the external circuit into the p-side. The photocurrent due to photogeneration in the neutral region is weaker than that due to photogeneration in the depletion region; in the latter, the field separates and drifts the carriers immediately. Photodetector designs prefer the photogeneration process to take place in the depletion region, which is the reason for keeping the p^+-layer as thin as possible. The photodetection in which there is an applied reverse bias V_r across the *pn* junction represents photodetection with the **photodiode reverse biased**.

Certain photodiodes are operated in the **photovoltaic mode**. There is no external battery connected to the photodiode. First, consider the *pn* junction shorted as in Figure 5.2 (c). Photogeneration in the SCL will again create an electron and a hole and these will now be driven by the built-in field E_o. The potential energy (PE) hill for the electron, change in E_c from the n- to the p-side, is no longer as steep as it was in the reverse biased *pn* junction. Nonetheless, there is a hill, and the electron will roll down the energy hill toward the n-side, that is, drift in the SCL. Similarly, the hole will roll down its own energy hill and drift in the opposite direction. As the electron and hole drift, they generate a photocurrent I_{ph} in the external circuit (remember that the current must be continuous). Notice that I_{ph} is, as before, a current that flows toward the n-side but now it is flowing in a short circuit. The situation is not far different than in Figure 5.1 (a) except there is no reverse bias V_r. If the photoexcitation is a very short pulse of light, the photocurrent magnitude is less than that in the reverse biased case (the field is lower and hence carriers

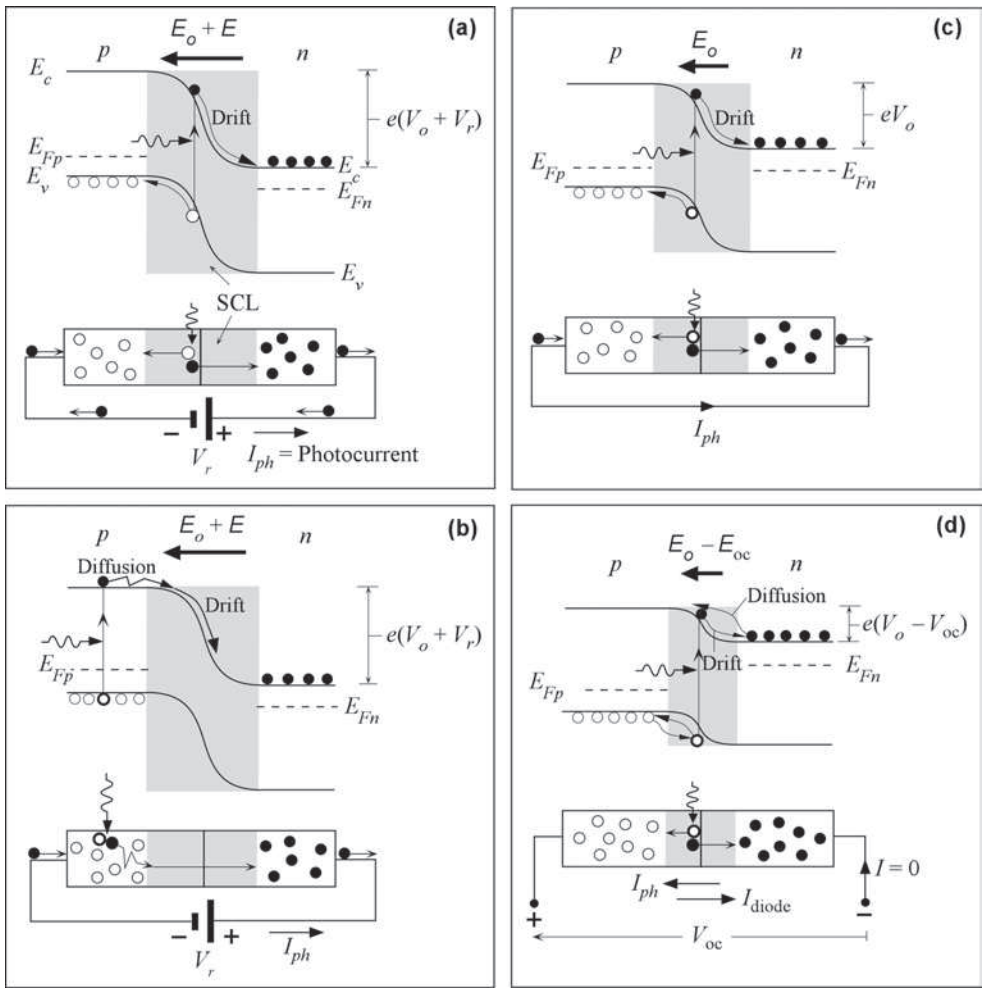


FIGURE 5.2 (a) A reverse biased *pn* junction. Photogeneration inside the SCL generates an electron and a hole. Both fall in their respective energy hills (electron along E_c and hole along E_v), *i.e.,* they drift, and cause a photocurrent I_{ph} in the external circuit. (b) Photogeneration occurs in the neutral region. The electron has to diffuse to the depletion layer and then roll down the energy hill, *i.e.,* drift across the SCL. (c) A shorted *pn* junction. The photogenerated electron and hole in the SCL roll down their energy hills, *i.e.,* drift across the SCL, and cause a current I_{ph} in the external circuit. (d) The *pn* junction in open circuit. The photogenerated electron and hole roll down their energy hills (drift) but there is a voltage V_{oc} across the diode that causes them to diffuse back so that the net current is zero. Note that E_c and E_v are only shown in (a), and are self-apperent in (b)–(d).

drift more slowly) and the time duration of the photocurrent is longer (drift takes longer across the SCL). The total collected charge is the same, if we neglect recombination losses.

If we leave the photodiode in an open circuit, we would observe an open circuit voltage V_{oc}. The photogeneration as before creates an electron and a hole that separate and drift in the SCL as shown in Figure 5.2 (d). Extra electrons are added to the *n*-side and extra holes are added to the *p*-side. As photogeneration continues, accumulated electrons on the *n*-side neutralize some of the positive donor charges in the SCL close to the neutral *n*-side. Similarly, the accumulated holes on the *p*-side neutralize some of the negative acceptor charges in the SCL next to the

neutral *p*-side. First, note that the accumulation of electrons on the *n*-side and holes on the *p*-side essentially upsets the equilibrium in such a way that the extra negative charge on the *n*-side and positive charge on the *p*-side give rise to a voltage V_{oc} across the diode. We can also understand this by noting that the reduction in positive and negative donor and acceptor charges in the SCL causes the field to decrease from E_o to, say, $E_o - E_{oc}$, where E_{oc} is the field that has resulted from the effects of photogeneration in an open circuit *pn* junction. The built-in voltage must therefore also decrease from V_o to $V_o - V_{oc}$. The voltage V_{oc} appears as a forward bias across the junction. Such forward bias would inject electrons from the *n*- to the *p*-side and holes from the *p*- to the *n*-side as in normal diode operation, and hence result in a diode current I_{diode}. The current I_{diode} is actually in the opposite direction to the photocurrent I_{ph} and its magnitude is such that the total current I is zero, as it must be in an open circuit. Thus, the result is an appearance of an open circuit voltage V_{oc} across the *pn* junction. It is apparent that the **photovoltaic operation** in Figure 5.2 (c) and (d) represents the short and open circuit configurations with no external battery, and also represents the principle of operation of a **solar cell**. Normally, a load (R) would be connected to the terminals of the photodiode (or the solar cell).

C. Current–Voltage Convention and Modes of Operation

We need to establish a sign convention to be able to analyze photodiode circuits. The voltage across a *pn* junction is stated in terms of the *p*-side with respect to *n*-side and is shown as V in Figure 5.3 (a). The current direction I is taken as the current flowing out from the *n*-side or entering the *p*-side as indicated in Figure 5.3 (a). If the applied voltage is a reverse bias of 5 V, and for convenience we write it as $V_r = 5$ V, as shown in Figure 5.3 (b), then $V = -V_r = -5$ V. If we illuminate the *pn* junction, then there will be a photocurrent I_{ph} flowing from the positive battery terminal to the *n*-side as shown in Figure 5.2 (a) and (b). I_{ph} is also shown in Figure 5.3 (b),

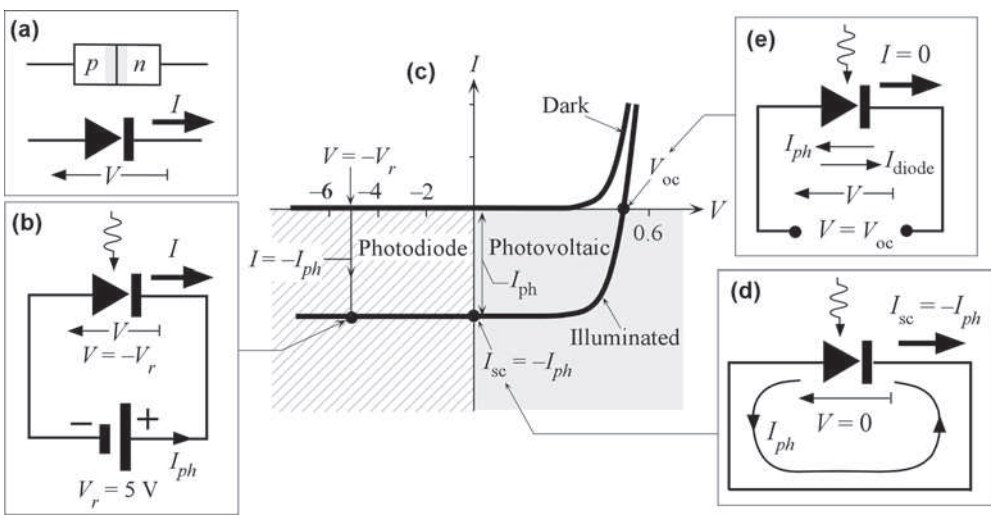


FIGURE 5.3 (a) The sign convention for the voltage V and current I for a *pn* junction. (b) If the *pn* junction is reverse biased by $V_r = 5$ V, then $V = -V_r = -5$ V. Under illumination, the *pn* junction current $I = -I_{ph}$ and is negative. (c) The *I*–*V* characteristics of a *pn* junction in the dark and under illumination. (d) A short circuit *pn* junction under illumination. The voltage $V = 0$ but there is a short circuit current so that $I = I_{sc} = -I_{ph}$. (e) An open circuit *pn* junction under illumination generates an open circuit voltage V_{oc}.

where I_{ph} is positive because it represents the actual current flow. However, the conventional diode current I is negative and written as $I = -I_{ph}$.

Figure 5.3 (c) shows the *I–V* characteristics of an ideal *pn* junction in the dark and also under illumination. In the dark, the *I–V* characteristic follows the usual diode equation, *i.e.,*

pn junction equation

$$I = I_o\left[\exp\left(eV/\eta k_B T\right) - 1\right] \tag{5.1.1}$$

where I and V are the diode current and voltage as in Figure 5.3 (a), I_o is the reverse saturation current, and η is the ideality factor.[2] For a reverse biased *pn* junction, $I = -I_o$ and is very small, which has been neglected in the dark characteristics in Figure 5.3 (c). We know from Figure 5.2 (c) that, in a short circuit, the photocurrent is I_{ph}, and it is in the opposite direction to the conventional *pn* junction current, that is, $I = I_{sc} = -I_{ph}$, where I_{sc} is the short circuit diode current. This short circuit situation is indicated in Figure 5.3 (d). We can sketch the *I–V* characteristics of the illuminated *pn* junction by simply shifting the dark *I–V* curve down until the current on the negative *I*-axis becomes $I = -I_{ph}$ as shown in Figure 5.3 (c). Notice that the *pn* junction *I–V* curve cuts the *V*-axis at $V = V_{oc}$. This is the **open circuit voltage** that would be generated when the *pn* junction is illuminated as shown in Figure 5.3 (e) in an open circuit, without any current in the external circuit. However, within the diode itself there are two currents that cancel each other exactly; one due to photogeneration (I_{ph}), and the other due to the diode current (I_{diode}) caused by V_{oc}, which acts like a forward bias across the junction.

The regions of the *pn* junction *I–V* characteristics that are bound by the positive *V* and negative *I* axes represent a **photovoltaic mode of operation**, shown as the gray region in Figure 5.3 (c). There is no applied bias and the light generates a photocurrent and a voltage across the device. When the *pn* junction is shorted, the current is I_{ph}, and when it is in open circuit, the voltage is V_{oc}. If there is a load of resistance R connected to the *pn* junction, we need to draw a load-line construction to find the operation (see Section 5.14). The region bound by the negative *V*-axis and the negative *I*-axis represents a reverse biased **photodiode mode of operation**; this is the most common mode of operation for the detection of light. However, it is not the only means of photodetection because there are several photovoltaic-type *pn* junction diodes used in radiation detection; two good examples are InAs and InSb infrared *pn* junction diodes. Although these are *pn* junctions, they are normally not biased due to the large dark current they have.

It should be emphasized that the dark current in the reverse biased *pn* junction was neglected in Figure 5.3 (c), and I_{ph} was assumed to be much greater than the magnitude of this reverse dark current. This assumption is not always true, especially under weak light conditions and in *pn* junctions made from narrow bandgap semiconductors that have high I_o.

5.2 SHOCKLEY–RAMO THEOREM AND EXTERNAL PHOTOCURRENT

Consider a semiconductor material with a negligible dark conductivity that is electroded and biased as shown in Figure 5.4 (a). The electrodes do not inject carriers but allow excess carriers in the sample to leave and become collected by the battery (they are termed *noninjecting electrodes*).[3] The field E in the sample is uniform and it is V/L. We will later see that this situation

[2] The symbol η is also used later for the external quantum efficiency. One should be able to tell the difference from usage and content. Some books use n but this is used for the electron concentration.

[3] As will be apparent later on, the situation in Figure 5.4 (a) also represents a reverse biased *pin* photodiode, one of the most common detector elements in optoelectronics.

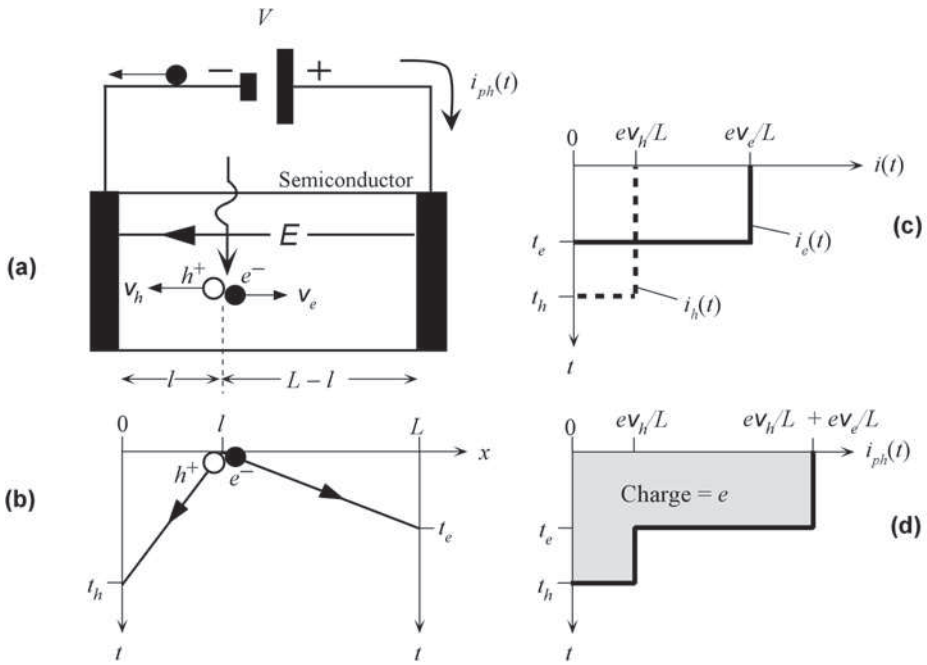


FIGURE 5.4 (a) An electron–hole pair is photogenerated at $x = l$. The electron and the hole drift in opposite directions with drift velocities v_h and v_e. (b) The electron arrives at time $t_e = (L - l)/v_e$ and the hole arrives at time $t_h = l/v_h$. (c) As the electron and hole drift, each generates an external photocurrent shown as $i_e(t)$ and $i_h(t)$. (d) The total photocurrent is the sum of hole and electron.

is almost identical to the intrinsic region of a reverse biased *pin* photodiode. Suppose that a single photon is absorbed at a position $x = l$ from the left electrode and instantly creates an electron hole pair. The electron and the hole drift in opposite directions with respective drift velocities $v_e = \mu_e E$ and $v_h = \mu_h E$, where μ_e and μ_h are the electron and hole drift mobilities, respectively. The **transit time** of a carrier is the time it takes for a carrier to drift from its generation point to the collecting electrode. The electron and hole transit times t_e and t_h, respectively, are marked on the t vs. x diagram in Figure 5.4 (b), where

Electron and hole transit times

$$t_e = \frac{L - l}{v_e} \quad \text{and} \quad t_h = \frac{l}{v_h} \tag{5.2.1}$$

Consider first only the drifting electron. Suppose that the external photocurrent due to the motion of this electron is $i_e(t)$. The electron is acted on by the force eE of the electric field. When it moves a distance dx, work must be done by the external circuit. In time dt, the electron drifts a distance dx and does an amount of work $eEdx$ which is provided by the battery in time dt as $Vi_e(t)dt$. Thus,

$$\text{Work done} = eEdx = Vi_e(t)dt$$

Using $E = V/L$ and $v_e = dx/dt$ we can find the electron photocurrent

Electron photocurrent

$$i_e(t) = \frac{ev_e}{L}; \quad t < t_e \tag{5.2.2}$$

It is apparent that this current continues to flow as long as the electron is drifting (has a velocity v_e) in the sample. It lasts for a duration t_e at the end of which the electron reaches the battery. Thus, although the electron has been photogenerated instantaneously, the external photocurrent is *not* instantaneous and has a time spread. Figure 5.4 (c) shows the electron photocurrent $i_e(t)$.

We can apply similar arguments to the drifting hole as well, which will generate a hole photocurrent $i_h(t)$ in the external circuit given, as in Figure 5.4 (c), by

Hole photocurrent

$$i_h(t) = \frac{ev_h}{L}; \quad t < t_h \tag{5.2.3}$$

The total external current, called the transient or **instantaneous photocurrent**, $i_{ph}(t)$ will be the sum of $i_e(t)$ and $i_h(t)$, as shown in Figure 5.4 (d).

If we integrate the external photocurrent $i_{ph}(t)$ to evaluate the collected charge $Q_{\text{collected}}$, we would find

$$Q_{\text{collected}} = \int_0^{t_e} i_e(t) + \int_0^{t_h} i_h(t) = e \tag{5.2.4}$$

This result can be verified by evaluating the area under the $i_{ph}(t)$ curve in Figure 5.4 (d). Thus, the collected charge is not $2e$ but just one electron, as shown by the area in Figure 5.4 (d). Equations (5.2.2)–(5.2.4) constitute the **Shockley–Ramo theorem**.[4] In general, if a charge q is being drifted with a velocity $v_d(t)$ by a field between two biased electrodes separated by L, then this motion of q generates an external current given by

Shockley–Ramo theorem

$$i(t) = \frac{qv_d(t)}{L}; \quad t < t_{\text{transit}} \tag{5.2.5}$$

The total external current is the sum of all currents of the type in Eq. (5.2.5) from all drifting charges between the electrodes, and the sign of q is negative for a drifting negative charge (electron).

5.3 ABSORPTION COEFFICIENT AND PHOTODETECTOR MATERIALS

The photon absorption process for photogeneration, that is, the creation of EHPs, requires the photon energy to be at least equal to the bandgap energy E_g of the semiconductor material to excite an electron from the valence band to the conduction band. The **upper cutoff wavelength** (or the threshold wavelength) λ_g for photogenerative absorption is therefore determined by the bandgap energy E_g of the semiconductor so that $h(c/\lambda_g) = E_g$ or

Cutoff wavelength and bandgap

$$\lambda_g(\mu\text{m}) = \frac{1.24}{E_g(\text{eV})} \tag{5.3.1}$$

For example, for Si, $E_g = 1.12\,\text{eV}$ and λ_g is $1.11\,\mu\text{m}$, whereas for Ge $E_g = 0.66\,\text{eV}$ and the corresponding $\lambda_g = 1.87\,\mu\text{m}$. It is clear that Si photodiodes cannot be used in optical communications at $1.3\,\mu\text{m}$ and $1.55\,\mu\text{m}$, whereas Ge photodiodes are commercially available for use

[4] Equation (5.2.5) is a simplified version of the more general treatment that examines the induced current on an electrode due to the motion of an electron. Its origins lie in tube-electronics in which engineers were interested in calculating how much current would flow into various electrodes of a vacuum tube as the electrons in the tube drifted. See W. Shockley, *J. Appl. Phys.*, *9*, 635, 1938, and S. Ramo, *Proc. IRE 27*, 584, 1939.

TABLE 5.1 Bandgap energy E_g at 300 K, cutoff wavelength λ_g, and type of bandgap (D = Direct and I = Indirect) for some photodetector materials

Semiconductor	E_g (eV)	λ_g (μm)	Type
InP	1.35	0.91	D
$GaAs_{0.88}Sb_{0.12}$	1.15	1.08	D
Si	1.12	1.11	I
$In_{0.7}Ga_{0.3}As_{0.64}P_{0.36}$	0.89	1.4	D
$In_{0.53}Ga_{0.47}As$	0.75	1.65	D
Ge	0.66	1.87	I
InAs	0.35	3.5	D
InSb	0.18	7	D

at these wavelengths. Table 5.1 lists some typical bandgap energies and the corresponding cutoff wavelengths of various photodiode semiconductor materials.

Incident photons with wavelengths shorter than λ_g become absorbed as they travel in the semiconductor and the light intensity, which is proportional to the number of photons, decays exponentially with distance into the semiconductor. The light intensity I at a distance x from the semiconductor surface is given by

Light intensity and absorption coefficient

$$I(x) = I_o \exp(-\alpha x) \tag{5.3.2}$$

where I_o is the intensity of the incident radiation and α is the **absorption coefficient** that depends on the photon energy or wavelength λ. Absorption coefficient α is a material property. Most of the photon absorption (63%) occurs over a distance $1/\alpha$, and $1/\alpha$ is called the **penetration** or **absorption depth** δ. Figure 5.5 shows the α vs. λ characteristics of various semiconductors where it is apparent that the behavior of α with the wavelength λ depends on the semiconductor material.

In **direct bandgap** semiconductors such as III–V semiconductors (*e.g.*, GaAs, InAs, InP, GaP) and in many of their alloys (*e.g.*, InGaAs, GaAsSb) the photon absorption process is a direct

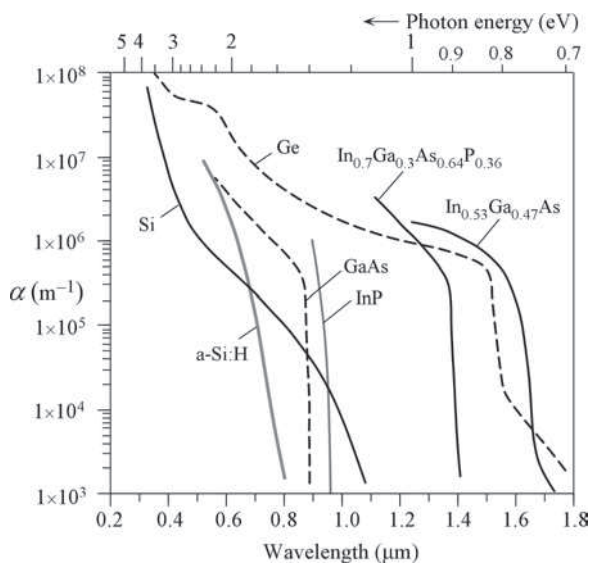


FIGURE 5.5 Absorption coefficient (α) vs. wavelength (λ) for various semiconductors. (Data selectively collected and combined from various sources.)

process which requires no assistance from lattice vibrations. The photon is absorbed and the electron is excited directly from the valence band to the conduction band without a change in its k-vector (or its crystal momentum $\hbar k$) inasmuch as the photon momentum is very small. The change in the electron momentum from the valence to the conduction band $\hbar k_{CB} - \hbar k_{VB} =$ photon momentum ≈ 0. This process corresponds to a vertical transition on the E-k diagram, that is, electron energy (E) vs. electron momentum ($\hbar k$) in the crystal in Figure 5.6 (a). A direct transition on the E-k diagram is a vertical transition from an initial energy E and wave vector k in the VB to a final energy E' and wave vector k' in the CB where $k' = k$, as illustrated in Figure 5.6 (a). The energy $(E' - E_c)$ is the kinetic energy $(\hbar k)^2/(2m_e^*)$ of the electron with an effective mass m_e^*, and $(E_v - E)$ is the kinetic energy $(\hbar k)^2/(2m_h^*)$ of the hole left behind in the VB, where m_h^* is the hole effective mass. The ratio of the kinetic energies of the photogenerated electron and hole depends inversely on the ratio of their effective masses. (The absorption coefficient α is derived from the quantum mechanical transition probability from E to E'.) The absorption coefficient α of these direct bandgap semiconductors rises sharply with decreasing wavelength from λ_g as apparent for GaAs and InP in Figure 5.5.

In **indirect bandgap** semiconductors such as Si and Ge, the electron at the top of the VB has a crystal momentum $\hbar k_{VB}$ and at the bottom of the CB, its momentum is $\hbar k_{CB}$, which is very different than $\hbar k_{VB}$ as illustrated in Figure 5.6 (b). The electron therefore cannot simply absorb a photon of energy E_g to get excited to the bottom of the CB inasmuch as this will violate the conservation of momentum. The photon absorption for photon energies near E_g requires the absorption and emission of lattice vibrations, that is, **phonons**,[5] during the photon absorption process, as shown in Figure 5.6 (b). If K is the wave vector of a lattice wave (lattice vibrations travel in the crystal), then $\hbar K$ represents the momentum associated with such a lattice vibration, that is, $\hbar K$ is a **phonon momentum**. When an electron in the valence band is excited to the conduction band there is a change in its momentum in the crystal and this change in the momentum cannot be supplied by the momentum of the incident photon, which is very small. Thus, the momentum difference must be balanced by a phonon momentum:

$$\hbar k_{CB} - \hbar k_{VB} = \text{phonon momentum} = \hbar K$$

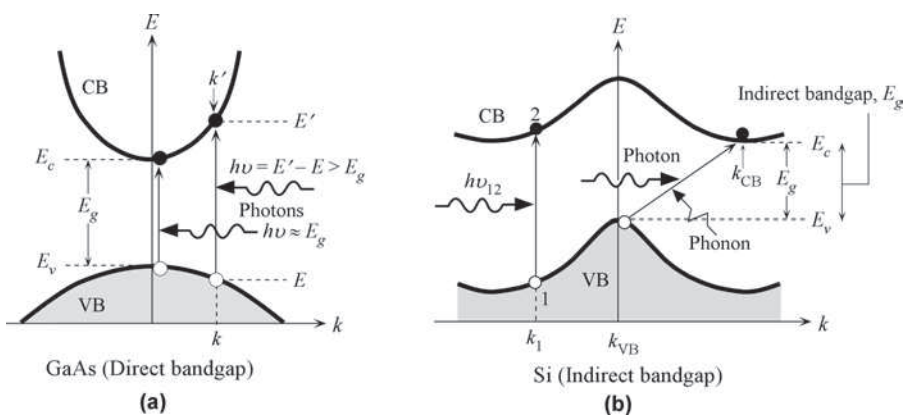


FIGURE 5.6 (a) Photon absorption in a direct bandgap semiconductor. (b) Photon absorption in an indirect bandgap semiconductor (VB, valence band; CB, conduction band).

[5] As much as an electromagnetic radiation is quantized in terms of photons, lattice vibrations in the crystal are quantized in terms of phonons.

The absorption process is said to be **indirect** as it depends on lattice vibrations which in turn depend on the temperature. Since the interaction of a photon with a valence electron needs a third body, a lattice vibration, the probability of photon absorption is not as high as in a direct transition. Furthermore, the cutoff wavelength is not as sharp as for direct bandgap semiconductors. During the absorption process, a phonon may be absorbed or emitted. The example in Figure 5.6 (b) shows the absorption of a phonon during the transition of the electron from E_v to E_c. If ϑ is the frequency of the lattice vibrations then the phonon energy is $h\vartheta$. The photon energy is $h\upsilon$ where υ is the photon frequency. Conservation of energy requires that

$$h\upsilon = E_g \pm h\vartheta$$

Thus, the onset of absorption does not exactly coincide with E_g, but typically it is very close to E_g inasmuch as $h\vartheta$ is small (<0.1 eV). The absorption coefficient initially rises slowly with decreasing wavelength from about λ_g, as apparent in Figure 5.5 for Ge and Si.

At sufficiently high photon energies, direct transitions eventually become possible, as shown by the transition 1 to 2 in Figure 5.6 (b). The 1-to-2 transition involves the absorption of a photon with energy $h\upsilon_{12}$. If the gradient of the *E-k* curve in the CB and VB are the same, then the transition probability is high.[6] The 1-to-2 transition in Figure 5.6 (b) was chosen for this reason. If we examine α vs. λ for Si in Figure 5.5, the sharp increase for wavelengths smaller than $\sim 0.5\,\mu$m is due to these direct transitions.

The choice of material for a photodiode must be such that the photon energies are greater than E_g. Further, at the wavelength of radiation, the absorption occurs over a depth covering the depletion layer so that the photogenerated EHPs can be separated and drifted by the field, and collected at the electrodes. If the absorption coefficient is too large, then absorption will occur very near the surface of the p^+-layer which is outside the depletion layer. First, the absence of a field means that the photogenerated electron can only make it to the depletion layer to cross to the *n*-side by diffusion. Second, photogeneration near the surface invariably leads to rapid recombination due to surface defects that act as recombination centers. On the other hand, if the absorption coefficient is too small, only a small portion of the photons will be absorbed in the depletion region and only a limited number of EHPs can be photogenerated.

5.4 QUANTUM EFFICIENCY AND RESPONSIVITY

Not all the incident photons on a photodiode are absorbed to create *free* electron and hole pairs (EHPs) that can be collected, *i.e.*, give rise to a photocurrent. The efficiency of the conversion process of received photons to charge carriers that can be collected is measured by the **external** or **device quantum efficiency (QE) η_e of the detector** defined as[7]

$$\eta_e = \frac{\text{Number of collected electrons at detector terminals}}{\text{Number of incident photons}} \qquad \textbf{(5.4.1)}$$

External quantum efficiency, QE

The measured photocurrent I_{ph} in the external circuit is due to the flow of electrons per second to the terminals of the photodiode, even though both electrons and holes would be drifting

[6] The reason is that the gradient dE/dk represents the group velocity of the electron, and photon-emitting or absorbing transitions need this group velocity to remain unchanged.

[7] "Free" implies carriers that can be collected, *i.e.,* electrons in the conduction band and holes in the valence band. Remember that the collection of one photogenerated free EHP is equivalent to the collection one electron externally.

inside the detector number of electrons collected per second is I_{ph}/e. If P_o is the incident optical power then the number of photons arriving per second is $P_o/h\upsilon$. The QE η_e can therefore also be defined by

External quantum efficiency, QE

$$\eta_e = \frac{I_{ph}/e}{P_o/h\upsilon} \tag{5.4.2}$$

Not all of the absorbed photons may photogenerate free EHPs that can be collected. Some EHPs may disappear by recombination without contributing to the photocurrent or become immediately trapped. Further, if the semiconductor length is comparable with the penetration depth $(1/\alpha)$, then not all the photons will be absorbed. There will also be photons lost by reflection at the air–semiconductor interface. The device QE is therefore always less than unity. It depends on the absorption coefficient α of the semiconductor at the wavelength of interest and on the structure of the device. QE can be increased by reducing the reflections at the semiconductor surface, increasing absorption within the depletion layer and preventing the recombination or trapping of carriers before they are collected. The QE defined in Eq. (5.4.1) is for the whole device and therefore it is known as the **external** QE. In contrast, **the internal quantum efficiency** is the number of free EHPs photogenerated per *absorbed photon* and is typically quite high for many devices. The QE definition in Eq. (5.4.1) incorporates the internal quantum efficiency as it applies to the whole device. The term QE alone normally implies external QE.

The **responsivity** R of a photodiode characterizes its performance in terms of the photocurrent generated (I_{ph}) per incident optical power (P_o) at a given wavelength

Responsivity

$$R = \frac{\text{Photocurrent (A)}}{\text{Incident optical power (W)}} = \frac{I_{ph}}{P_o} \tag{5.4.3}$$

From the definition of QE, it is clear that

Responsivity and QE

$$R = \eta_e \frac{e}{h\upsilon} = \eta_e \frac{e\lambda}{hc} \tag{5.4.4}$$

In Eq. (5.4.4), η_e depends on the wavelength. The responsivity therefore clearly depends on the wavelength. R is also called the **spectral responsivity** or **radiant sensitivity**. The R vs. λ characteristic represents the spectral response of the photodiode and is generally provided by the manufacturer. Ideally, with a quantum efficiency of 100% ($\eta_e = 1$), R should increase with λ up to λ_g, as indicated in Figure 5.7. In practice, QE limits the responsivity to lie below the ideal photodiode line, with upper and lower wavelength limits as shown for a typical Si photodiode in Figure 5.7. The QE of a well-designed Si photodiode in the wavelength range 700–900 nm can be as high as 90–95%. Suppose that we draw a line through the origin that is a tangent to the R vs. λ curve at X as in Figure 5.7. We can show from Eq. (5.4.4) that point X represents operation under maximum QE; at the wavelength λ_1, the QE is maximum and the responsivity is R_1 (see Example 5.4.2).

The shape of the responsivity R vs. λ curve in Figure 5.7 depends on a number of factors, the most important being the device structure, the absorption coefficient α of the semiconductor and the QE. Figure 5.8 shows a typical *pn* junction photodiode in which the *p*-side is heavily doped. The neutral *p*- and *n*-regions are marked by ℓ_p and ℓ_n, respectively. The reverse bias produces a large internal field E. The SCL has a width W, primarily on the *n*-side, and ℓ_p is much narrower than W. There are essentially three types of photogeneration processes that can contribute to the observed photocurrent.

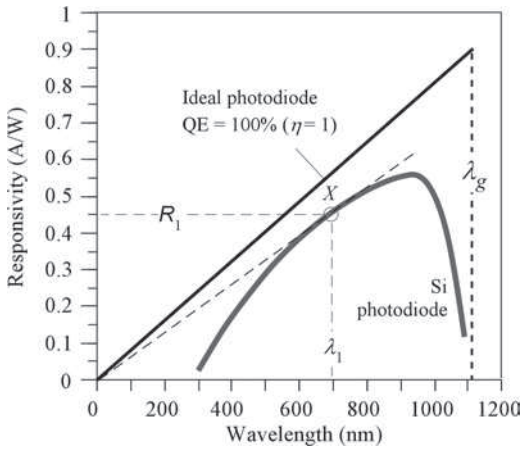


FIGURE 5.7 Responsivity (R) vs. wavelength (λ) for an ideal photodiode with QE = 100% ($\eta_e = 1$) and for a typical inexpensive commercial Si photodiode. (The exact shape of the responsivity curve depends on the device structure.) The line through the origin that is a tangent to the responsivity curve at X identifies operation at λ_1 with maximum QE.

Consider "short wavelengths," that is, the absorption depth $1/\alpha$ is less than the width ℓ_p of the neutral region. The photogeneration takes place mainly within the neutral p-region, *i.e.,* within ℓ_p. If the photogeneration occurs within the **minority carrier diffusion length** L_e to the SCL boundary, as illustrated in Figure 5.8, then the electron can diffuse and reach the SCL, and then be drifted by the field E across the SCL over to the n-side, where it becomes collected. The photogenerated hole on the p-side is a majority carrier and is neutralized by the electron brought

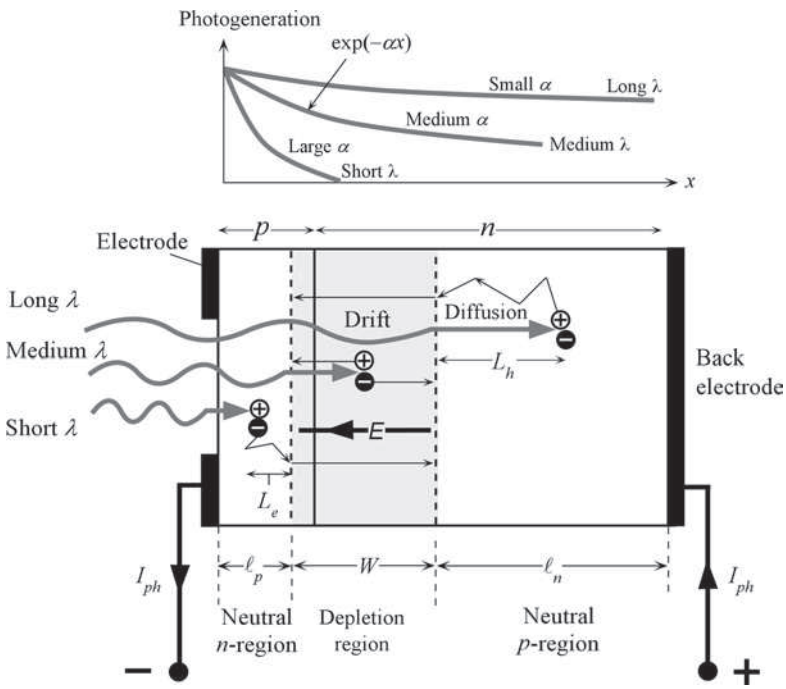
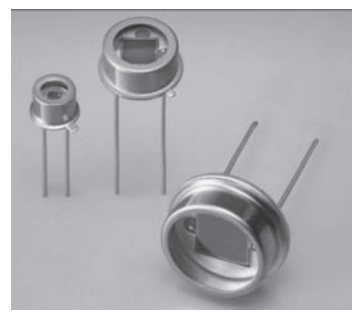


FIGURE 5.8 Different contributions to the photocurrent I_{ph}. Photogeneration profiles corresponding to short, medium, and long wavelengths are also shown.

in by the photocurrent from the negative terminal. The external photocurrent lasts until the drifting electron reaches the neutral *n*-region. (Although we considered the generation of one EHP, photogeneration is an ongoing process.)

Consider next "medium wavelengths," *i.e.,* the absorption depth $1/\alpha$ is comparable to $\ell_p + W$. Significant photogeneration now takes place within the SCL, especially since ℓ_p is narrower than W. The field E in the SCL separates the EHP and drifts them in opposite directions, giving rise to a photocurrent. For "long wavelengths" for which the absorption depth is longer than $\ell_p + W$, as shown in Figure 5.8, only those holes photogenerated in the neutral *n*-side within the hole diffusion length L_h to the SCL can be collected. Thus, the useful photogeneration that contributes to the photocurrent takes place over $L_e + W + L_h$.

Although in this highly simplified discussion we only considered the interplay between the absorption coefficient (which depends on λ) and the device structure, the internal QE can also be important. The QE may not be uniform through the device. For very short wavelength light, the absorption depth is so small that the photogeneration occurs very close to the surface and outside the diffusion length L_e to the SCL. In Figure 5.8, these EHPs would be generated in $\ell_p - L_e$, which behaves like a "dead-zone"; the carriers essentially disappear by recombination. Short wavelength limitation of some of the *pn* junction photodiodes often arises from this dead zone. If ℓ_p is made shorter than L_e to overcome this problem, some carriers are still lost because they easily diffuse to the surface and disappear by recombination (surface is within the diffusion length L_e).



Si photodiodes of various sizes (S1336 series). *(Courtesy of Hamamatsu.)*

EXAMPLE 5.4.1 Quantum efficiency and responsivity

Consider the photodiode shown in Figure 5.7. What is the QE at peak responsivity? What is the QE at 450 nm (blue)? If the photosensitive device area is 1 mm^2, what would be the light intensity corresponding to a photocurrent of 10 nA at the peak responsivity?

Solution

The peak responsibility in Figure 5.7 occurs at about $\lambda \approx 940$ nm where $R \approx 0.56\ A\ W^{-1}$. Thus, from Eq. (5.4.4), that is, $R = \eta_e e\lambda/hc$, we have

$$0.56\ A\ W^{-1} = \eta_e \frac{(1.6 \times 10^{-19}\ C)(940 \times 10^{-9}\ m)}{(6.63 \times 10^{-34}\ J\ s)(3 \times 10^{8}\ m\ s^{-1})}, \quad i.e., \quad \eta_e = 0.74 \quad \text{or} \quad 74\%$$

We can repeat the calculation for $\lambda = 450$ nm, where $R \approx 0.24\ A\ W^{-1}$, which gives $\eta_e = 0.66$ or 66%.

From the definition of responsivity, $R = I_{ph}/P_o$, we have $0.56\,A\,W^{-1} = (10 \times 10^{-9}\,A)/P_o$, *i.e.,* $P_o = 1.8 \times 10^{-8}$ W or 18 nW. Since the area is 1 mm^2 the intensity must be 18 nW mm^{-2}.

EXAMPLE 5.4.2 Maximum quantum efficiency

Show that a photodiode has maximum QE when

$$\frac{dR}{d\lambda} = \frac{R}{\lambda} \tag{5.4.5}$$

Maximum quantum efficiency

that is, when the tangent X at λ_1 in Figure 5.7 passes through the origin ($R = 0$, $\lambda = 0$). Hence, determine the wavelength where the QE is maximum for the Si photodiode in Figure 5.7.

Solution

From Eq. (5.4.4) the QE is given by

$$\eta_e = \frac{hcR(\lambda)}{e\lambda} \tag{5.4.6}$$

where $R(\lambda)$ depends on λ and there is also λ in the denominator. We can differentiate Eq. (5.4.6) with respect to λ and then set it to zero to find the maximum point X. Thus

$$\frac{d\eta_e}{d\lambda} = \frac{hc}{e\lambda}\frac{dR}{d\lambda} - \frac{hcR}{e}\left(\frac{1}{\lambda^2}\right) = 0$$

which leads to Eq. (5.4.5). Equation (5.4.5) represents a line through the origin that is a tangent to the R vs. λ curve. This tangential point is X in Figure 5.7, where $\lambda_1 = 700$ nm and $R_1 = 0.45\,\text{A W}^{-1}$. Then, using Eq. (5.4.6), the maximum QE is

$$\eta_e = (6.626 \times 10^{-34}\,\text{J s})(3 \times 10^{8}\,\text{m s}^{-1})(0.45\,\text{A W}^{-1})/(1.6 \times 10^{-19}\,\text{C})(700 \times 10^{-9}\,\text{m})$$
$$= 0.80 \quad \text{or} \quad 80\%$$

5.5 THE *pin* PHOTODIODE

The simple *pn* junction photodiode (Figure 5.1) has two major drawbacks. Its junction or depletion layer capacitance is not sufficiently small to allow photodetection at high modulation frequencies. This is an *RC* time constant limitation. Second, its depletion layer is at most a few microns. This means that at long wavelengths where the penetration depth is greater than the depletion layer width, the majority of photons are absorbed outside the depletion layer where there is no field to separate and drift the EHPs. The QE is correspondingly low at these long wavelengths. These problems are substantially reduced in the ***pin* (*p*–intrinsic–*n*-type) photodiode**.[8]

The *pin* refers to a semiconductor device that has the structure p^+–intrinsic–n^+ as schematically illustrated in the idealized structure in Figure 5.9 (a). The intrinsic layer has much smaller doping than both p^+- and n^+-regions and it is much wider than these regions, typically 5–50 μm depending on the particular application. In the idealized ***pin* photodiode**, we can take, for simplicity, the *i*-Si region to be truly intrinsic. When the structure is first formed, holes diffuse from the p^+-side and electrons from n^+-side into the *i*-Si layer where they recombine and disappear. This leaves behind a thin layer of exposed negatively charged acceptor ions in the p^+-side and a thin layer of exposed positively charged donor ions in the n^+-side as shown in Figure 5.9 (b).

[8] The *pin* photodiode was invented by Jun-ichi Nishizawa and his research group at Tohoku University in Japan in 1952. Among Professor Nishizawa's many distinguished awards is "The Order of Cultural Merits" from the Japanese Emperor. He is currently the President Emeritus of Tokyo Metropolitan University.

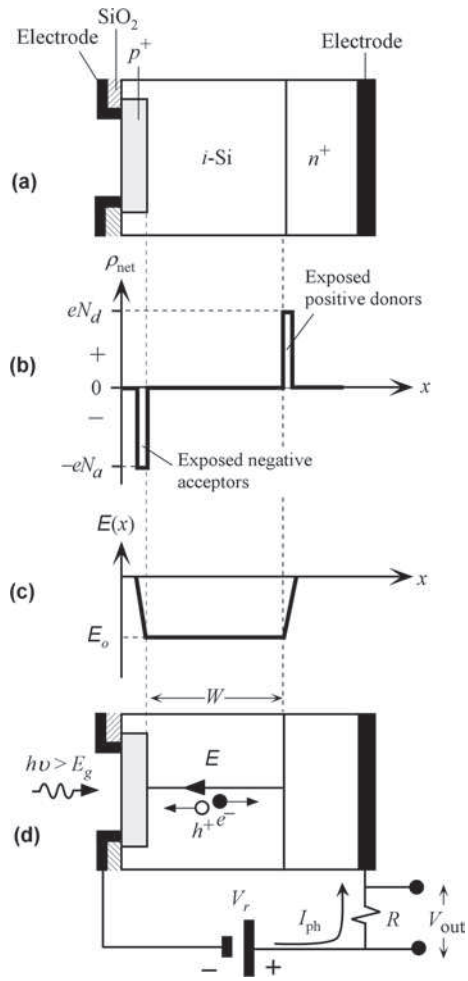


FIGURE 5.9 (a) The schematic structure of an idealized *pin* photodiode. (b) The net space charge density across the photodiode. (c) The built-in field across the diode. (d) The *pin* photodiode reverse biased for photodetection.

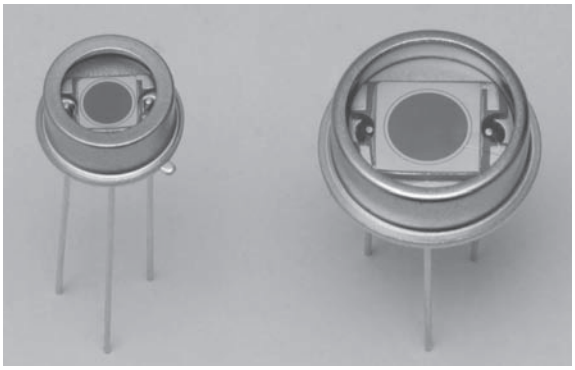
The two charges are separated by the *i*-Si layer of width W. There is a uniform built-in field E_o in *i*-Si layer from the exposed positive ions to exposed negative ions as illustrated in Figure 5.9 (c). In contrast, the built-in field in the depletion layer of a *pn* junction is not uniform. With no applied bias, equilibrium is maintained by the built-in field E_o which prevents further diffusion of majority carriers into the *i*-Si layer.

The separation of two very thin layers of negative and positive charges by a fixed distance, width W of the *i*-Si, is almost the same as that in a parallel plate capacitor. The **junction or depletion layer capacitance** C_{dep} of the *pin* diode is given by

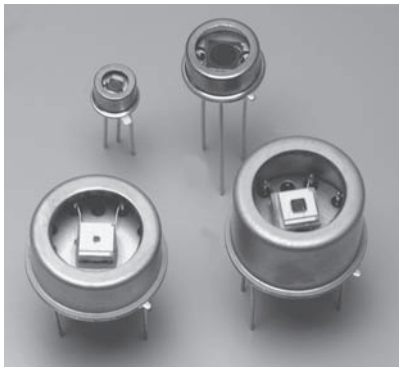
Junction capacitance of pin

$$C_{\text{dep}} = \frac{\varepsilon_o \varepsilon_r A}{W} \tag{5.5.1}$$

where A is the cross-sectional area and $\varepsilon_o\varepsilon_r$ is the permittivity of the semiconductor (Si), respectively, with ε_r being the relative permittivity. Further, since the width W of the *i*-Si layer is fixed by the structure, the junction capacitance C_{dep} does not depend on the applied voltage in contrast



Si *pin* photodiodes. Radiation receiving active device area has a diameter of 3 mm for the left and 5 mm for the right *pin* photodiode. The case diameters are about 8.1 mm and 12.4 mm. *(Courtesy of Hamamatsu.)*



InGaAs *pin* photodiodes. Active device areas have diameters from 0.3 mm (smallest) to 3 mm (largest). The bottom two are thermoelectric cooled, and have a case diameter of 15 mm. *(Courtesy of Hamamatsu.)*

to that of the *pn* junction. C_{dep} is typically of the order of a picofarad in fast *pin* photodiodes so that with a 50-Ω resistor, the RC_{dep} time constant is about 50 ps.

When a reverse bias voltage V_r is applied across the *pin* device, as shown in Figure 5.9 (d), it drops almost entirely across the width of *i*-Si layer. The depletion layer widths of the thin sheets of acceptor and donor charges in the p^+- and n^+-sides are negligible compared with W. The reverse bias V_r increases the built-in voltage to $V_o + V_r$. The field E in the *i*-Si layer is still uniform and increases to

$$E = E_o + \frac{V_r}{W} \approx \frac{V_r}{W} \quad (V_r \gg V_o) \tag{5.5.2}$$

Biased pin

The *pin* structure is designed so that photon absorption occurs over the *i*-Si layer. The photogenerated EHPs in the *i*-Si layer are then separated by the field E and drifted toward the n^+- and p^+-sides, respectively, as illustrated in Figure 5.9 (d). While the photogenerated carriers are drifting through the *i*-Si layer they give rise to an external photocurrent which is detected as a voltage across a small sampling resistor R in Figure 5.9 (d). The **response time** of the *pin* photodiode is determined by the transit times of the photogenerated carriers across the width W of the *i*-Si layer. Increasing W allows more photons to be absorbed, which increases the QE but it slows down the speed of response as carrier transit times become longer. For a charge carrier that is photogenerated at the edge on the *i*-Si layer, the transit time or drift time t_{drift} across the *i*-Si layer is

$$t_{\text{drift}} = \frac{W}{v_d} \tag{5.5.3}$$

Transit time

where v_d is its drift velocity. To reduce the drift time, that is, to increase the speed of response, we have to increase v_d and therefore increase the applied field E. At high fields v_d does not follow the expected $\mu_d E$ behavior, where μ_d is the drift mobility, but instead tends to saturate at v_{sat}, which is of the order of 10^5 m s^{-1} at fields greater than 10^6 V m^{-1} in the case of Si. Figure 5.10 shows the variation of the drift velocity of electrons and holes with the field in Si. The $v_d = \mu_d E$ behavior is only observed at low fields. At high fields, both electron and hole drift velocities saturate. For an *i*-Si layer of width 10 μm, with carriers drifting at saturation velocities, the drift time

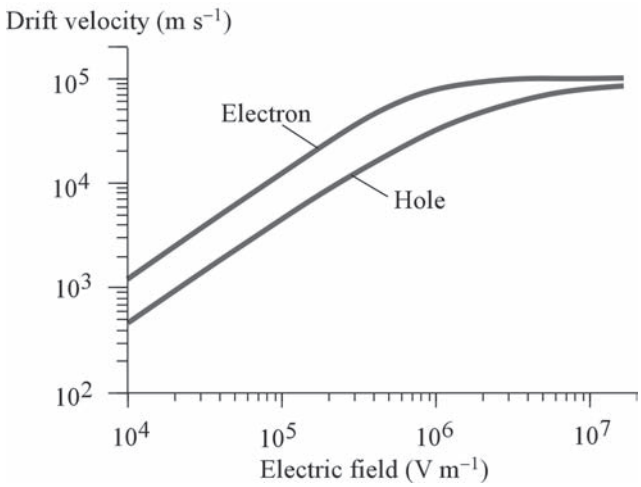


FIGURE 5.10 Drift velocity vs. electric field for holes and electrons in Si.

is about 0.1 ns which is longer than typical RC_{dep} time constants. The speed of *pin* photodiodes are invariably limited by the transit time of photogenerated carriers across the *i*-Si layer.

The *pin* photodiode structure shown in Figure 5.9 is, of course, idealized. In reality, the *i*-Si layer will have some small doping. For example, if the sandwiched layer is lightly *n*-type doped, it is labeled as a υ-layer and the structure is $p^+\upsilon n^+$. The sandwiched υ-layer becomes a depletion layer with a small concentration of exposed positive donors. The field then is not entirely uniform across the photodiode. The field is maximum at the $p^+\upsilon$ junction and then decreases slightly across υ-Si to reach the n^+-side. As an approximation we can still consider the υ-Si layer as an *i*-Si layer.

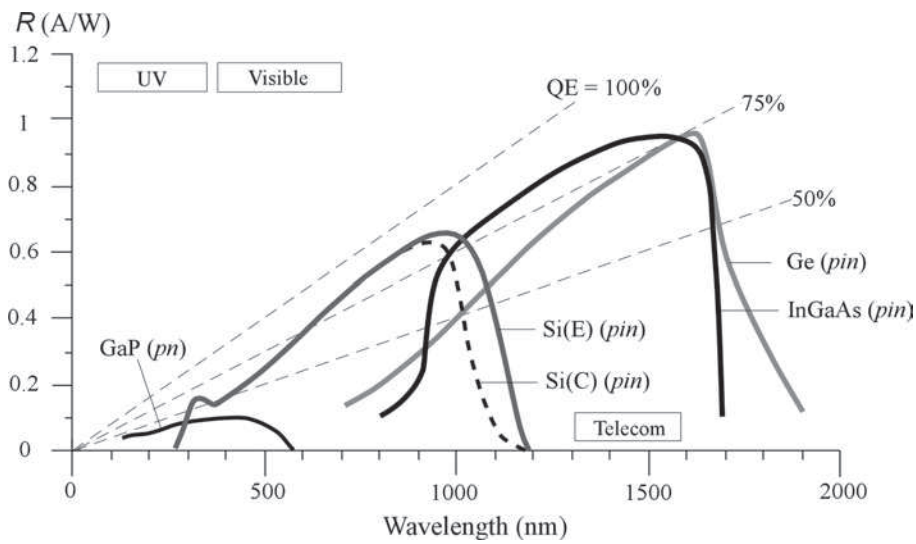


FIGURE 5.11 The responsivity of Si, InGaAs, and Ge *pin*-type photodiodes. The *pn* junction GaP detector is used for UV detection. GaP (Thorlabs, FGAP71), Si(E), IR enhanced Si (Hamamatsu S11499), Si(C), conventional Si with UV enhancement, InGaAs (Hamamatsu, G8376), and Ge (Thorlabs, FDG03). The dashed lines represent the responsivity due to QE = 100%, 75%, and 50%.

As mentioned above, a distinct advantage of the *pin* photodiode is that it allows a wider spectral range to be absorbed in the SCL in which the photogeneration takes place. Consequently, the responsivity R is generally better than the simple *pn* junction photodiode and can be controlled by adjusting the width of the *i*-layer. Both Si and InGaAs *pin* photodiodes are widely available in the market, covering a range of wavelength from around 300 nm to 1700 nm. Ge *pin* photodiodes are also available but have higher dark currents, and usually have to be cooled. Figure 5.11 shows the responsivity of Ge, Si, and InGaAs *pin,* and GaP *pn* junction detectors from the UV (150 nm) to the optical communications channels. The dashed lines show the responsivity corresponding to certain quantum efficiencies, QE = 50%, 75%, and 100%. Notice that both Si and InGaAs *pin* photodiodes operate with quantum efficiencies, above 75%, and are the work horses for numerous photodetection tasks. The GaP *pn* junction is used for UV detection, though there are Si photodiodes available that can also operate in the UV. Further, notice also that the InGaAs photodiodes can be used in both 1550 nm and 1310 nm optical systems.

EXAMPLE 5.5.1 Operation and speed of a *pin* photodiode

A Si *pin* photodiode has an *i*-Si layer of width 20 μm. The p^+-layer on the illumination side is very thin (0.1 μm). The *pin* is reverse biased by a voltage of 100 V and then illuminated with a very short optical pulse of wavelength 900 nm. What is the duration of the photocurrent if absorption occurs over the whole *i*-Si layer?

Solution

From Figure 5.5, the absorption coefficient at 900 nm is $\sim 3 \times 10^4\ \text{m}^{-1}$ so that the absorption depth is ~33 μm. We can assume that absorption, and hence photogeneration, occurs over the entire width W of the *i*-Si layer. The field in the *i*-Si layer is

$$E \approx V_r/W = (100\ \text{V})/(20 \times 10^{-6}\ \text{m}) = 5 \times 10^6\ \text{V m}^{-1}$$

At this field the electron drift velocity v_e is very near its saturation at $10^5\ \text{m s}^{-1}$, whereas the hole drift velocity v_h is about $7 \times 10^4\ \text{m s}^{-1}$ as shown in Figure 5.10. Holes are slightly slower than the electrons. The transit time t_h of holes across the *i*-Si layer is

$$t_h = W/v_h = (20 \times 10^{-6}\ \text{m})/(7 \times 10^4\ \text{m s}^{-1}) = 2.86 \times 10^{-10}\ \text{s} \quad \text{or} \quad 0.29\ \text{ns}$$

This is the response time of the *pin* as determined by the transit time of the slowest carriers, holes, across the *i*-Si layer. To improve the response time, the width of the *i*-Si layer has to be narrowed but this decreases the quantity of photons absorbed and hence reduces the responsivity. There is therefore a trade-off between speed and responsivity.

EXAMPLE 5.5.2 Photocarrier diffusion in a *pin* photodiode

A reverse biased *pin* photodiode is illuminated with a short wavelength light pulse that is absorbed very near the surface as shown in Figure 5.12. The photogenerated electron has to diffuse to the depletion region, where it is swept into the *i*-layer and drifted across by the field in this region. What is the speed of response of this photodiode if the *i*-Si layer is 20 μm and the p^+-layer is 1 μm and the applied voltage is 60 V? The diffusion coefficient (D_e) of electrons in the heavily doped p^+-region is approximately $3 \times 10^{-4}\ \text{m}^2\ \text{s}^{-1}$.

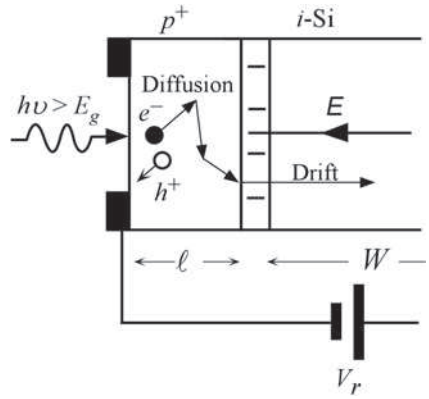


FIGURE 5.12 A reverse biased *pin* photodiode is illuminated with a short wavelength light pulse that is absorbed very near the surface. The photogenerated electron has to diffuse to the depletion region where it is swept into the *i*-layer and drifted across.

Solution

There is no electric field in the p^+-side outside the depletion region as shown in Figure 5.12. The photogenerated electrons have to make it across to the n^+-side to give rise to a photocurrent. In the p^+-side, the electrons move by diffusion. In time t, an electron, on average, diffuses a distance ℓ given by[9]

$$\ell = (2D_e t)^{1/2}$$

The *diffusion time* t_{diff} is the time it takes for an electron to diffuse across the p^+-side (of length ℓ) to reach the depletion layer and is given by

$$t_{\text{diff}} = \ell^2/(2D_e) = (1 \times 10^{-6}\,\text{m})^2/\left[2(3 \times 10^{-4}\,\text{m}^2\,\text{s}^{-1})\right] = 1.67 \times 10^{-9}\,\text{s} \quad \text{or} \quad 1.67\,\text{ns}$$

On the other hand, once the electron reaches the depletion region, it becomes drifted across the width W of the *i*-Si layer at the saturation drift velocity since the electric field here is $E = V_r/W = 60\,\text{V}/20\,\mu\text{m} = 3 \times 10^6\,\text{V m}^{-1}$; and at this field the electron drift velocity v_e saturates at $10^5\,\text{m s}^{-1}$. The *drift time* across the *i*-Si layer is

$$t_{\text{drift}} = W/v_e = (20 \times 10^{-6}\,\text{m})/(1 \times 10^5\,\text{m s}^{-1}) = 2.0 \times 10^{-10}\,\text{s} \quad \text{or} \quad 0.2\,\text{ns}$$

Thus, the response time of the *pin* to a pulse of short wavelength radiation that is absorbed near the surface is very roughly $t_{\text{diff}} + t_{\text{drift}}$ or 1.87 ns. Notice that the diffusion of the electron is much slower than its drift. Further, in a proper analysis, we have to consider the diffusion and drift of many carriers, and we have to average $(t_{\text{diff}} + t_{\text{drift}})$ for all the electrons.

EXAMPLE 5.5.3 Responsivity of a *pin* photodiode

A Si *pin* photodiode has an active light-receiving area of diameter 0.4 mm. When radiation of wavelength 700 nm (red light) and intensity $0.1\,\text{mW cm}^{-2}$ is incident, it generates a photocurrent of 56.6 nA. What is the responsivity and external QE of the photodiode at 700 nm?

Solution

The incident light intensity $I = 0.1\,\text{mW cm}^{-2}$ means that the incident power for conversion is

$$P_o = AI = \left[\pi(0.02\,\text{cm})^2\right](0.1 \times 10^{-3}\,\text{W cm}^{-2}) = 1.26 \times 10^{-7}\,\text{W} \quad \text{or} \quad 0.126\,\mu\text{W}$$

[9] See, for example, S. O. Kasap, *Principles of Electronic Materials and Devices*, 3rd Edition (McGraw-Hill, 2006), Chs. 1 and 5 for the derivation of the root mean square diffusion distance.

The responsivity is

$$R = I_{ph}/P_o = (56.6 \times 10^{-9}\,\text{A})/(1.26 \times 10^{-7}\,\text{W}) = 0.45\,\text{A W}^{-1}$$

The QE can be found from

$$\eta_e = R\frac{hc}{e\lambda} = (0.45\,\text{A W}^{-1})\frac{(6.62 \times 10^{-34}\,\text{J s})(3 \times 10^{8}\,\text{m s}^{-1})}{(1.6 \times 10^{-19}\,\text{C})(700 \times 10^{-9}\,\text{m})} = 0.80 = 80\%$$

EXAMPLE 5.5.4 Steady state photocurrent in the *pin* photodiode

Consider a *pin* photodiode that is reverse biased and illuminated, as in Figure 5.9, and operating under steady state conditions. Assume that the photogeneration takes place inside the depletion layer of width W, and the neutral p-side is very narrow. If the incident optical power on the semiconductor is $P_o(0)$, then $TP_o(0)$ will be transmitted, where T is the transmission coefficient. At a distance x from the surface, the optical power $P_o(x) = TP_o(0)\exp(-\alpha x)$. In a small volume δx at x, the absorbed radiation power (by the definition of α) is $\alpha P_o(x)\delta x$, and the number of photons absorbed per second is $\alpha P_o(x)\delta x/h\nu$. Of these absorbed photons, only a fraction η_i will photogenerate EHPs, where η_i is the **internal quantum efficiency** IQE. Thus, $\eta_i\alpha P_o(x)\delta x/h\nu$ number of EHPs will be generated per second. We assume these will drift through the depletion region and thereby contribute to the photocurrent. The current contribution δI_{ph} from absorption and photogeneration at x within the SCL will thus be

$$\delta I_{ph} = \frac{e\eta_i\alpha P_o(x)\delta x}{h\nu} = \frac{e\eta_i\alpha T P_o(0)}{h\nu}\exp(-\alpha x)\delta x$$

We can integrate this from $x = 0$ (assuming ℓ_p is very thin) to the end of $x = W$, and assuming $W \gg L_h$ to find

Steady state photocurrent pin photodiode

$$I_{ph} \approx \frac{e\eta_i T P_o(0)}{h\nu}\left[1 - \exp(-\alpha W)\right] \qquad \textbf{(5.5.4)}$$

where the approximate sign embeds the many assumptions we made in deriving Eq. (5.5.4). Consider a *pin* photodiode without an AR coating so that $T = 0.68$. Assume $\eta_i = 1$. The SCL width is 20 μm. If the device is to be used at 900 nm, what would be the photocurrent if the incident radiation power is 100 nW? What is the responsivity? Find the photocurrent and the responsivity if a perfect AR coating is used. What is the primary limiting factor? What is the responsivity if $W = 40$ μm?

Solution

From Figure 5.5, at $\lambda = 900$ nm, $\alpha \approx 3 \times 10^4\,\text{m}^{-1}$. Further for $\lambda = 0.90$ μm, the photon energy $h\nu = 1.24/0.90 = 1.38$ eV. Given $P_o(0) = 100$ nW, we have

$$I_{ph} \approx \frac{(1.6 \times 10^{-19})(1)(0.68)(100 \times 10^{-9})}{(1.38 \times 1.6 \times 10^{-19})}\left[1 - \exp(-3 \times 10^4 \times 20 \times 10^{-6})\right] = 22\,\text{nA}$$

and the responsivity $R = 22\,\text{nA}/100\,\text{nW} = 0.22\,\text{A W}^{-1}$, which is on the low-side.

Consider next a perfect AR coating so that $T = 1$, and using Eq. (5.5.4) again, we find $I_{ph} = 32.7$ nA and $R = 0.33\,\text{A W}^{-1}$, a significant improvement.

The factor $\left[1 - \exp(-\alpha W)\right]$ is only 0.451, and can be significantly improved by making the SCL thicker. Setting $W = 40$ μm gives $\left[1 - \exp(-\alpha W)\right] = 0.70$ and $R = 0.51$, which is close to values for commercial devices.

The maximum theoretical photocurrent would be obtained by setting $\exp(-\alpha W) \approx 0$, $T = 1$, $\eta_i = 1$, which gives $I_{ph} = 73$ nA and $R = 0.73\,\text{A W}^{-1}$.

5.6 AVALANCHE PHOTODIODE

A. Principles and Device Structures

Avalanche photodiodes (APDs) are widely used in optical communications due to their high speed and internal gain. They are also used in various applications where sensitivity is important. A simplified schematic diagram of a Si **reach-through APD** is shown in Figure 5.13 (a). The n^+-side is thin and it is the side that is illuminated through a window. There are three p-type layers of different doping levels next to the n^+-layer to suitably modify the field distribution across the diode. The first is a thin p-type layer and the second is a thick, lightly p-type doped (almost intrinsic) layer, called the π-layer, and the third is a heavily doped p^+-layer. The diode is reverse biased to increase the fields in the depletion regions. The net space charge distribution across the diode due to exposed dopant ions is shown in Figure 5.13 (b). Under zero bias, the depletion layer in the p-region (between n^+p) does not normally extend across this layer. But when a sufficient reverse bias is applied, the depletion region in the p-layer widens to *reach through* to the π-layer (and hence the name *reach-through*). The field extends from the exposed positively charged donors in the thin depletion layer in n^+-side, all the way to the exposed negatively charged acceptors in the thin depletion layer in p^+-side.

The electric field is given by the integration of the net space charge density ρ_{net} across the diode subject to an applied voltage V_r across the device. The variation in the field across the diode is shown in Figure 5.13 (c). The field lines start at positive ions and end at negative ions

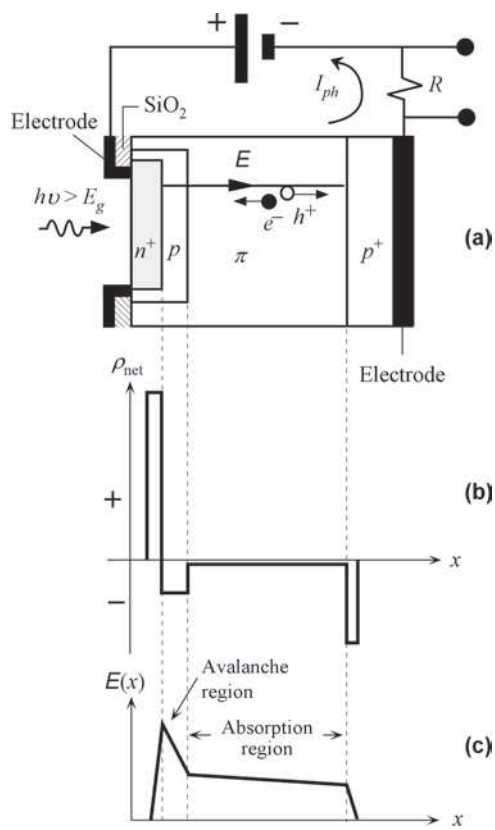


FIGURE 5.13 (a) A schematic illustration of the structure of an avalanche photodiode (APD) biased for avalanche gain. (b) The net space charge density across the photodiode. (c) The field across the diode and the identification of absorption and multiplication regions.

which exist through the p-, π-, and p^+-layers. This means that E is maximum at the n^+p junction, then decreases slowly through the p-layer. Through the π-layer, it decreases only slightly as the net space charge density here is small. The field vanishes at the end of the narrow depletion layer in the p^+-side.

The absorption of photons, and hence photogeneration, takes place mainly in the long π-layer. The nearly uniform field here separates the electron–hole pairs and drifts them at velocities near saturation toward the n^+- and p^+-sides, respectively. When the drifting electrons reach the p-layer, they experience even greater fields and therefore acquire sufficient kinetic energy (greater than E_g) to **impact-ionize** some of the Si covalent bonds and release EHPs. We can visualize the impact ionization process as shown in Figure 5.14 (a),[10] where an electron entering the avalanche region (width w) gains energy from the field as it "drifts" in the opposite direction to the field, and its energy (which is kinetic energy) increases with respect to E_c. Eventually, the energy gained from the field is sufficient to excite an electron across the bandgap E_g as illustrated in Figure 5.14 (b). These impact-ionization-generated carriers are called **secondary carriers**. These secondary EHPs themselves can also be accelerated by the high fields in this region to sufficiently large kinetic energies to further cause impact ionization and release more EHPs, which leads to an **avalanche of impact ionization processes**. Thus, from a single electron entering the p-layer one can generate a large number of EHPs, all of which contribute to the observed photocurrent. The photodiode possesses an **internal gain mechanism** in that single photon absorption leads to a large number of EHPs being generated. The photocurrent in the APD in the presence of avalanche multiplication, therefore, corresponds to an effective quantum efficiency in excess of unity.

The reason for keeping the photogeneration within the π-region and reasonably separate from the avalanche p-region in Figure 5.13 (a) is that avalanche multiplication is a statistical process and hence leads to carrier generation fluctuations, which lead to **excess noise** in the avalanche-multiplied photocurrent. This is minimized if impact ionization is restricted to the carrier with the highest impact ionization efficiency, which in Si is the electron. Thus, the structure in Figure 5.13 (a) allows the photogenerated electrons to drift and reach the avalanche region but not the photogenerated holes.

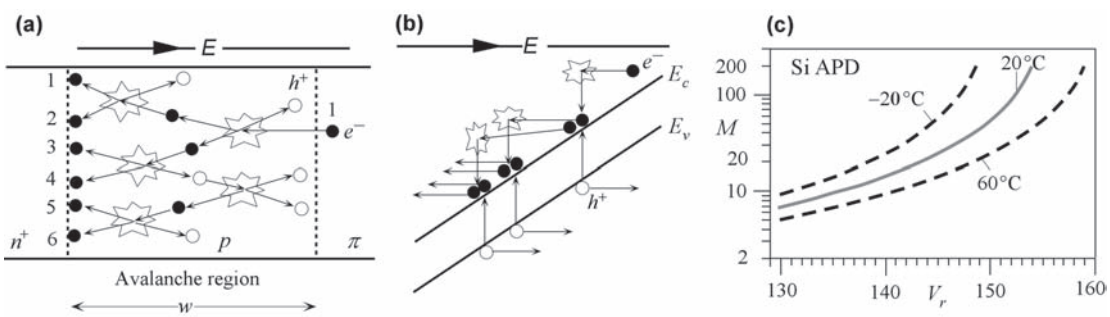


FIGURE 5.14 (a) A pictorial view of impact ionization processes releasing EHPs and the resulting avalanche multiplication. (b) Impact of an energetic conduction electron with crystal vibrations transfers the electron's kinetic energy to a valence electron and thereby excites it to the conduction band. (c) Typical multiplication (gain) M vs. reverse bias characteristics for a typical commercial Si APD, and the effect of temperature. (M measured for a photocurrent generated at 650 nm of illumination.)

[10]The impact of an energetic electron with a Si-Si bond to release an electron–hole pair is only an intuitive picture of the process.

The multiplication of carriers in the avalanche region depends on the probability of impact ionization, which depends strongly on the field in this region and hence on the reverse bias V_r. The overall or effective **avalanche multiplication factor** M, also known as the **gain**, of an APD is defined as[11]

Definition of multiplication M

$$M = \frac{\text{Multiplied photocurrent}}{\text{Primary unmultiplied photocurent}} = \frac{I_{ph}}{I_{pho}} \tag{5.6.1}$$

where I_{ph} is the APD photocurrent that has been multiplied and I_{pho} is the **primary or unmultiplied photocurrent**, the photocurrent that is measured in the absence of multiplication, for example, under a small reverse bias V_r. The multiplication M is a strong function of the reverse bias, as shown in Figure 5.14 (c), and also the temperature. The multiplication M can be approximately described by an empirical relationship of the form[12]

Multiplication and Miller's equation

$$M = \frac{1}{1 - \left(\dfrac{V_r}{V_{\text{br}}}\right)^m} \tag{5.6.2}$$

where V_{br} is a parameter called the **avalanche breakdown voltage** and m is a characteristic index that provides the best fit to the experimental data. Both V_{br} and m are temperature dependent. For Si APDs M values can be as high as 1000 or more, but for many commercial Ge and InGaAs APDs M values are typically around 10–20. The multiplication M defined in Eq. (5.6.1) is also called the **gain** of the APD. Both V_r and V_{br} are positive for a reverse biased photodetector.

The speed of the reach-through APD, shown in Figure 5.13 (a), depends on three factors. First is the time it takes for the photogenerated electron to cross the absorption region (the π-layer) to the multiplication region (the p-layer). Second is the time it takes for the avalanche process to build up in the p-region and generate EHPs. The third is the time it takes for the last hole released in the avalanche process to transit through the π-region. The response time of an APD to an optical pulse is therefore somewhat longer than a corresponding *pin* structure but, in practice, the multiplicative gain often makes up for the reduction in the speed. The overall speed of a photodetector circuit also includes limitations from the electronic preamplifier connected to the photodetector. The APD requires less subsequent electronic amplification, which translates to an overall speed that can be faster than a corresponding detector circuit using a *pin* photodiode and a preamplifier.

One of the drawbacks of the simple reach-through APD structure is that the field around the n^+p junction peripheral edge reaches avalanche breakdown before the n^+p regions under the illuminated area as illustrated in Figure 5.15 (a). Ideally avalanche multiplication should occur uniformly in the illuminated region to encourage the avalanche multiplication of the primary photocurrent rather than the multiplication of the dark current (*i.e.*, the thermally generated random electron hole pairs). In a practical Si APD, an n-type doped region acting as a **guard ring** surrounds the central n^+-region as shown in Figure 5.15 (b) so that the breakdown voltage around the periphery is now higher and avalanche is confined more to the illuminated region

[11] This definition is for an *average* avalanche multiplication for two reasons. First is that the avalanche process is a statistical process with a mean and a standard deviation. Second, the field is not entirely uniform in the avalanche region and impact ionization probability is extremely sensitive to the electric field, which means that multiplication is not uniform over the multiplication region.

[12] This is usually called Miller's equation (S. L. Miller, *Phys. Rev.*, *99*, 1234, 1955).

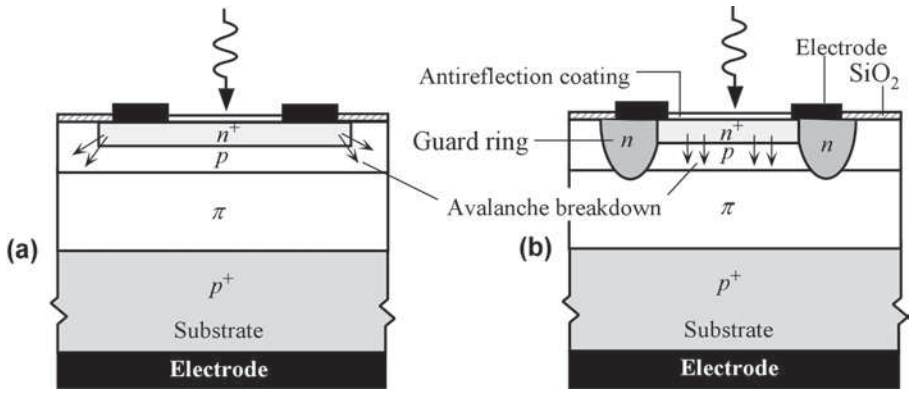


FIGURE 5.15 (a) A Si APD structure without a guard ring. (b) A schematic illustration of the structure of a more practical Si APD.

(n^+p junction). The n^+- and p-layers are very thin (<2 μm) to reduce any absorption in this region; the main absorption occurs in the thick π-region.

Table 5.2 summarizes some typical characteristics of *pn* junction, *pin*, and APD photodiodes based on GaAsP, Si, Ge, InGaAs, InAs, and InSb, covering the range from the ~200 nm in the UV, based on GaP, to ~5 μm in the infrared (InSb). Si-based detectors, because of the well-established microelectronics technology, tend to be less expensive and more prevalent over the 200–1100 nm range. For telecom applications, the detector must be able to respond to 1310 nm and 1550 nm optical signals and also have sufficient speed. These photodiodes are typically *pin* or avalanche photodiodes that have been carefully designed to provide a wide bandwidth detection.

TABLE 5.2 Typical characteristics of some *pn* junction–, *pin*-, and APD-type photodiodes based on GaP, GaAsP, Si, Ge, InGaAs, InAs, and InSb, covering the range from the ~200 nm in the UV (GaP), to ~5 μm in the infrared (InSb)

Photodiode	λ_{range} (nm)	λ_{peak} (nm)	R at λ_{peak} (A W^{-1})	Gain	I_d for 1 mm^2	Features
GaP *pin*	150–550	450	0.1	<1	1 nm	UV detection[a]
GaAsP *pn*	150–750	500–720	0.2–0.4	<1	0.005–0.1 nA	UV to visible, covering the human eye, low I_d
GaAs *pin*	570–870	850	0.4–0.5	<1	0.1–1 nA	High speed, low I_d
Si *pn*	200–1100	600–900	0.5–0.6	<1	0.005–0.1 nA	Inexpensive, general purpose, low I_d
Si *pin*	300–1100	800–1000	0.5–0.6	<1	0.1–1 nA	Faster than *pn*
Si APD	400–1100	800–900	0.4–0.6[b]	$10–10^3$	1–10 nA[c]	High gains, fast
Ge *pin*	700–1800	1500–1580	0.4–0.7	<1	0.1−1 μA	IR detection, fast
Ge APD	700–1700	1500–1580	0.4–0.8[b]	10–20	1−10 μA[c]	IR detection, fast
InGaAs *pin*	800–1700	1500–1600	0.7–1	<1	1–50 nA	Telecom, high speed, low I_d
InGaAs APD	800–1700	1500–1600	0.7–0.95[b]	10–20	0.05−10 μA[c]	Telecom, high speed, and gain
InAs *pn*	2−3.6 μm	3.0−3.5 μm	1–1.5	<1	>100 μA	Photovoltaic mode; normally cooled
InSb *pn*	4−5.5 μm	5 μm	3	<1	Large	Photovoltaic mode; normally cooled

Notes: I_d is the typical dark current under normal operating conditions. The dark current depends on the device area, and the values are typical for 1 mm^2. [a]FGAP71 (Thorlabs); [b]At $M = 1$; [c]At operating multiplication.

EXAMPLE 5.6.1 InGaAs APD responsivity

An InGaAs APD has a quantum efficiency (QE, η_e) of 60% at 1.55 μm in the absence of multiplication ($M = 1$). It is biased to operate with a multiplication of 12. Calculate the photocurrent if the incident optical power is 20 nW. What is the responsivity when the multiplication is 12?

Solution

The responsivity at $M = 1$ in terms of the quantum efficiency is

$$R = \eta_e \frac{e\lambda}{hc} = (0.6)\frac{(1.6 \times 10^{-19}\,\text{C})(1550 \times 10^{-9}\,\text{m})}{(6.626 \times 10^{-34}\,\text{J s})(3 \times 10^{8}\,\text{m s}^{-1})} = 0.75\ \text{A W}^{-1}$$

If I_{pho} is the primary photocurrent (unmultiplied) and P_o is the incident optical power, then by definition $R = I_{pho}/P_o$ so that

$$I_{pho} = RP_o = (0.75\ \text{A W}^{-1})(20 \times 10^{-9}\,\text{W}) = 1.5 \times 10^{-8}\,\text{A} \quad \text{or} \quad 15\ \text{nA}$$

The photocurrent I_{ph} in the APD will be I_{pho} multiplied by M,

$$I_{ph} = MI_{pho} = (12)(1.5 \times 10^{-8}\,\text{A}) = 1.80 \times 10^{-7}\,\text{A} \quad \text{or} \quad 180\ \text{nA}$$

The responsivity at $M = 12$ is

$$R' = I_{ph}/P_o = MR = (12)/(0.75) = 9.0\ \text{A W}^{-1}$$

EXAMPLE 5.6.2 Silicon APD

A Si APD has a QE of 70% at 830 nm in the absence of multiplication, that is, $M = 1$. The APD is biased to operate with a multiplication of 100. If the incident optical power is 10 nW, what is the photocurrent?

Solution

The unmultiplied responsivity is given by

$$R = \eta_e \frac{e\lambda}{hc} = (0.70)\frac{(1.6 \times 10^{-19}\,\text{C})(830 \times 10^{-9}\,\text{m})}{(6.626 \times 10^{-34}\,\text{J s})(3 \times 10^{8}\,\text{m s}^{-1})} = 0.47\ \text{A W}^{-1}$$

The unmultiplied primary photocurrent from the definition of R is

$$I_{pho} = RP_o = (0.47\ \text{A W}^{-1})(10 \times 10^{-9}\,\text{W}) = 4.7\ \text{nA}$$

The multiplied photocurrent is

$$I_{ph} = MI_{pho} = (100)(4.67\ \text{nA}) = 470\ \text{nA} \quad \text{or} \quad 0.47\ \mu\text{A}$$

B. Impact Ionization and Avalanche Multiplication

The impact ionization process illustrated in Figure 5.14 (a) involves impact ionization in the avalanche region that has a width w. Intuitively, if we make the width longer, there should be more impact ionization processes, and hence M should be larger. Although the exact treatment can be quite complicated, we can at least derive some simple relationships between M and the width w and the

reverse bias V_r through semiquantitative argument. The probability that an electron in the avalanche region (w) causes impact ionization per unit distance is called the **electron ionization coefficient** α_e. We can view $1/\alpha_e$ as the average distance an electron travels in w before it causes impact ionization. Put differently, it is the average separation between two consecutive impact ionizations. Similarly, the probability that a hole causes impact ionization per unit distance would be the **hole ionization coefficient** α_h; and $1/\alpha_h$ is the average distance between two hole-initiated impact ionizations. The ratio k of these ionization coefficients is important in APD designs, and is defined as

$$k = \frac{\text{Ionization coefficient for holes}}{\text{Ionization coefficient for electrons}} = \frac{\alpha_h}{\alpha_e} \tag{5.6.3}$$

Ionization coefficient ratio

For Si, $\alpha_e > \alpha_h$ and k is less than unity. The ionization coefficients increase sharply with the field E in an exponential-like manner and can be written as[13]

$$\alpha_e = A\exp(-B/E) \tag{5.6.4}$$

Ionization coefficient and field

where A and B are material-specific constants, different for holes and electrons. α_e and E in Eq. (5.6.4) are at the same point inside the avalanche region. The increase in the field leads to large increases in α_e and α_h and hence M. Consider a small volume of narrow width dx, within w, in which there are N number of electrons. First, we will ignore the impact ionization by holes. The increase dN in N depends on the number electrons we have, N, and the probability of impact ionization in this volume, $\alpha_e dx$, so that $dN = N\alpha_e dx$. Put differently $(1/N)dN = \alpha_e dx$. We can easily integrate this from the left to the right end of the avalanche width w in Figure 5.14 (a), from N_1 to N_2 electrons, to obtain the increase in the number of electrons and hence the multiplication factor $M = N_2/N_1$. The integration is simple and gives an exponential dependence for M,

$$M = \exp(\alpha_e w) \tag{5.6.5}$$

Multiplication by electrons only

Inasmuch as α_e depends on the field by virtue of Eq. (5.6.4), we end up with M having a very strong dependence on the field. In a proper treatment we would need to include the hole-initiated ionization events as well in accounting for multiplication in w. Equation (5.6.5) must be modified, and final result is[14]

$$M = \frac{1 - k}{\exp\left[-(1 - k)\alpha_e w\right] - k} \tag{5.6.6}$$

Multiplication by electrons and holes

Clearly, if $k = 0$ (holes do not ionize), Eq. (5.6.6) reduces to Eq. (5.6.5). In our APD model in Figure 5.14 (a), we allowed only the electrons to enter the avalanche region (not holes). We need to use Eq. (5.6.6) to properly account for the impact ionized holes (secondary holes) avalanching in w as well, as shown in Example 5.6.3. If electrons cannot impact ionize in this region, that is, when $k = \infty$, then $M = 1$, even if the hole ionization coefficient is very large. Equation (5.6.6) is easily modified to account for multiplication when holes are allowed to drift into the avalanche region. We would invert the definition of k and use α_h for α_e.

---

[13] Equation (5.6.4) is called the Chyoweth's law. More generally α_e would be given by $\alpha_e = A\exp\left[-(B/E)^n\right]$ where n is a material-specific index. Since impact ionization coefficient measurements usually have some scatter (due to experimental uncertainties), many researchers simply use Eq. (5.6.4) with $n = 1$.

[14] See, for example, B. E. A. Saleh and M. C. Teich, *Fundamental of Photonics*, 2nd Edition (Wiley, 2007), Ch. 18, pp. 769–773.

Notice that M in Eq. (5.6.6) approaches infinity when the denominator $\exp[-(1-k)\alpha_e w] - k \approx 0$. This occurs at sufficiently high fields when α_e and k satisfy this equation, and corresponds to a runaway multiplication in the avalanche region, *i.e.,* both electrons and holes, and their secondary carriers, are all heavily involved in avalanche multiplication. There is no such runaway, $M \approx \infty$, when $k = 0$, and we have only electrons ionizing, that is, only one type of carrier is involved. Further, in the presence of both electron and hole impact ionization, there is higher excess noise generated by the avalanche process as discussed below.

EXAMPLE 5.6.3 Avalanche multiplication in Si APDs

The electron and hole ionization coefficients α_e and α_h in silicon are approximately given by Eq. (5.6.4) with[15] $A \approx 0.740 \times 10^6\ \text{cm}^{-1}$, $B \approx 1.16 \times 10^6\ \text{V cm}^{-1}$ for electrons (α_e), and $A \approx 0.725 \times 10^6\,\text{cm}^{-1}$ and $B \approx 2.2 \times 10^6\ \text{V cm}^{-1}$ for holes (α_h). Suppose that the width w of the avalanche region is 0.5 μm. Find the multiplication gain M when the applied field in this region reaches $4.00 \times 10^5\ \text{V cm}^{-1}$, $4.30 \times 10^5\ \text{V cm}^{-1}$, and $4.38 \times 10^5\ \text{V cm}^{-1}$. What is your conclusion?

Solution

At the field of $E = 4.00 \times 10^5\ \text{V cm}^{-1}$, from Eq. (5.6.4)

$$\begin{aligned}\alpha_e &= A\exp(-B/E)\\ &= (0.740 \times 10^6\,\text{cm}^{-1})\exp\left[-(1.16 \times 10^6\ \text{V cm}^{-1})/(4.00 \times 10^5\ \text{V cm}^{-1})\right] = 4.07 \times 10^4\,\text{cm}^{-1}\end{aligned}$$

Similarly using Eq. (5.6.4) for holes, $\alpha_h = 2.96 \times 10^3\,\text{cm}^{-1}$. Thus $k = \alpha_h/\alpha_e = 0.073$. Using this k and α_e above in Eq. (5.6.6) with $w = 0.5 \times 10^{-4}\,\text{cm}$,

$$M = \frac{1 - 0.073}{\exp\left[-(1 - 0.073)(4.07 \times 10^4\ \text{cm}^{-1})(0.5 \times 10^{-4}\ \text{cm})\right] - 0.073} = 11.8$$

Note that if we had only electron avalanche without holes ionizing, then the multiplication would be

$$M_e = \exp(\alpha_e w) = \exp\left[(4.07 \times 10^4\ \text{cm}^{-1})(0.5 \times 10^{-4}\ \text{cm})\right] = 7.65$$

TABLE 5.3 Avalanche multiplication in a Si APD and breakdown

E (V cm^{-1})	α_e (cm^{-1})	α_h (cm^{-1})	k	M	M_e	Comment
4.00×10^5	4.07×10^4	2.96×10^3	0.073	11.8	7.65	M and M_e not too different at low E
4.30×10^5	4.98×10^4	4.35×10^3	0.087	57.2	12.1	7.5% increase in E, large difference between M and M_e
4.38×10^5	5.24×10^4	4.77×10^3	0.091	647	13.7	1.9% increase in E

We can now repeat the calculations for $E = 4.30 \times 10^5\ \text{V cm}^{-1}$ and again for $E = 4.38 \times 10^5\ \text{V cm}^{-1}$. The results are summarized in Table 5.3 for both M and M_e. Notice how quickly M builds up with the field and how a very small change at high fields causes an enormous change in M that eventually leads to a breakdown. (M running away to infinity as V_r increases.) Notice also that in the presence of only electron-initiated ionization, M_e simply increases without a sharp runaway to breakdown.

[15] Values from Table 12.12 in S. Adachi, *Properties of Group IV, III-V and II-VI Semiconductors* (Wiley, UK, 2005).

5.7 HETEROJUNCTION PHOTODIODES

A. Separate Absorption and Multiplication APD

Various III–V compound-based APDs have been developed for use at the communications wavelengths 1.3 μm and 1.55 μm. We will consider a few examples. As in the reach-through Si APD, the **absorption** or **photogeneration region** is separated from the **avalanche** or **multiplication region**, which allows the multiplication to be initiated by one type of carrier. Figure 5.16 is a simplified schematic diagram of the structure of an InGaAs-InP APD with a **separate absorption and multiplication (SAM) regions**. InP has a wider bandgap than InGaAs and the *p*- and *n*-type doping of InP is indicated by capital letters *P* and *N*. The main depletion layer is between P^+-InP and *N*-InP layers and it is within the *N*-InP. This is where the field is greatest and therefore it is in this *N*-InP layer where avalanche multiplication takes place. With sufficient reverse bias, the depletion layer in the *N*-InP *reaches* (or *punches*) *through* to the *n*-InGaAs layer. Eventually, both *N*-InP ansd *n*-InGaAs are depleted. The field in the depletion layer in *n*-InGaAs is not as great as that in *N*-InP. The variation of the field across the device is also shown in Figure 5.16 under an applied reverse bias that has caused reach-through. Although the long-wavelength photons are incident onto the InP side, they are not absorbed by InP since the photon energy is less than the bandgap energy of InP ($E_g = 1.35\,\text{eV}$). Photons pass through the InP layer and become absorbed in the *n*-InGaAs layer. The field in the *n*-InGaAs layer drifts the holes to the multiplication region where impact ionization multiplies the carriers.

There are a number of practical features which are not shown in the highly simplified diagram in Figure 5.16. Photogenerated holes drifting from *n*-InGaAs to *N*-InP become trapped at the interface because there is a sharp increase in the bandgap and a sharp change ΔE_v in E_v (valence band edge) between the two semiconductors, and holes cannot easily surmount the potential energy barrier ΔE_v as illustrated in Figure 5.17 (a). This problem is overcome by using thin layers of *n*-type InGaAsP with intermediate bandgaps to provide a graded transition from InGaAs to InP as illustrated in Figure 5.17 (b). Effectively ΔE_v has been broken up into two or more steps. The hole has sufficient energy to overcome the first step and enter the InGaAsP layer. It drifts and accelerates in the InGaAsP layer to gain sufficient energy to surmount the second step. These devices are called **separate absorption, grading and multiplication**

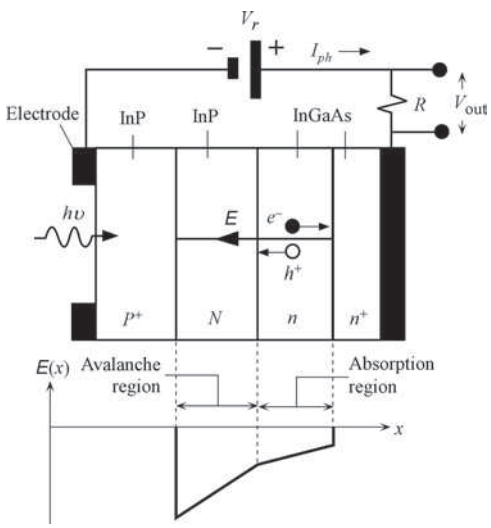


FIGURE 5.16 Simplified schematic diagram of a separate absorption and multiplication APD using a heterostructure based on InGaAs-InP. *P* and *N* refer to *p*- and *n*-type wider bandgap semiconductor.

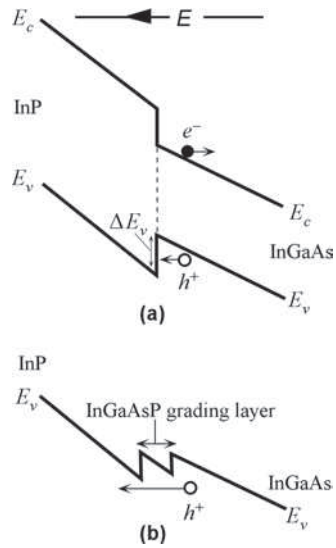


FIGURE 5.17 (a) Energy band diagrams for a SAM detector with a step junction between InP and InGaAs. There is a valence band step ΔE_v from InGaAs to InP that slows hole entry into the InP layer. (b) An interposing grading layer (InGaAsP) with an intermediate bandgap breaks ΔE_v and makes it easier for the hole to pass to the InP layer for a detector with a graded junction between InP and InGaAs. This is the SAGM structure.

(SAGM) APDs. Both the InP layers are grown epitaxially on an InP substrate. The substrate itself is not used directly to make the *P-N* junction to prevent crystal defects (*e.g.*, dislocations) in the substrate appearing in the multiplication region and hence deteriorating the device performance. The schematic diagram of a more practical SAGM APD is illustrated in Figure 5.18.

The current vs. voltage characteristics for an InGaAs reach-through APD are shown in Figure 5.19. The dark current without any illumination is I_d and, like the photocurrent, it is also multiplied by the gain. The photodiode current has both the dark current I_d and the photocurrent I_{ph}. Initially the depletion region in the *N*-InP does not cover this entire region. The photogenerated hole in the *n*-InGaAs layer therefore has to diffuse some distance to the depletion region to find the field to be drifted. Thus, initially, as the reverse bias increases, the photocurrent I_{ph} and the dark current I_d increase because the reverse bias widens the depletion region further in the *N*-InP layer. The holes have to diffuse less and less distance as the voltage is increased, and the current increases until reach-through is attained, and the whole of the *N*-InP is depleted. From then onward, there is a field in the whole of the depleted *N*-layer and the photogenerated holes become drifted toward to P^+-InP. Impact ionization starts soon after reach-through, and the gain

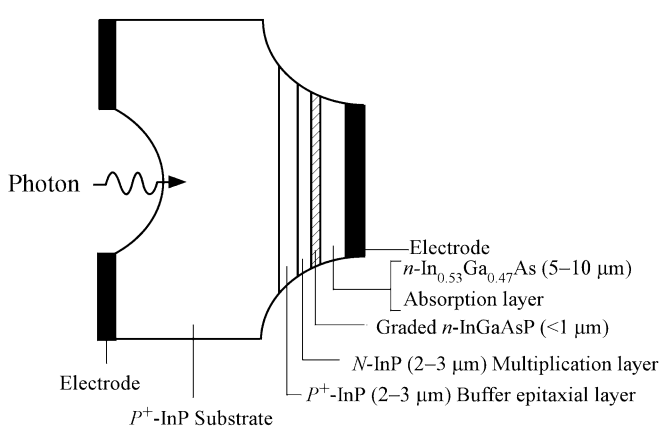


FIGURE 5.18 Simplified schematic diagram of a more practical mesa-etched SAGM layered APD.

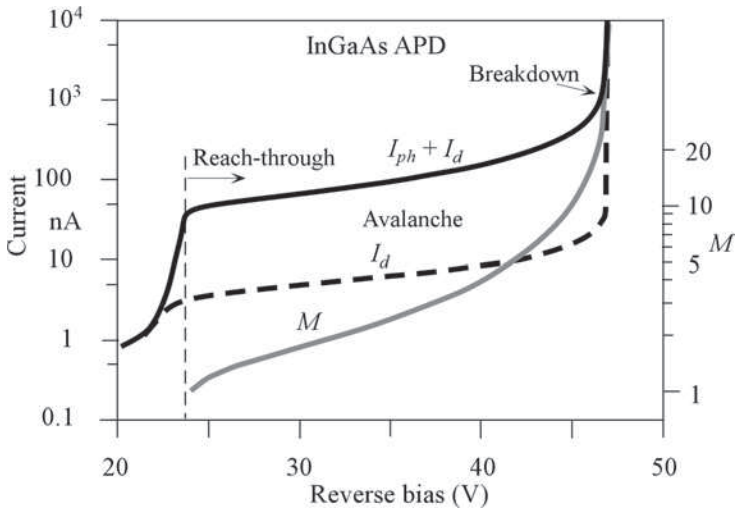


FIGURE 5.19 Typical current and gain (M) vs. reverse bias voltage for a commercial InGaAs reach-through APD. I_d and I_{ph} are the dark current and photocurrent, respectively. The input optical power is ~100 nW. The gain M is 1 when the diode has attained reach-through and then increases with the applied voltage. (*Source:* The data extracted selectively from Voxtel Catalog, Voxtel, Beaverton, OR 97006.)

M increases as the bias voltage increases as shown in Figure 5.19. Typical gains for commercial InGaAs APD are in the 10–20 range. At a sufficiently high reverse bias, the device will eventually break down with a runaway current; this voltage is the **breakdown voltage**.

B. Superlattice APDs

As mentioned earlier, APDs exhibit excess noise in the photocurrent (above the expected shot noise) due to inherent statistical variations in the avalanche multiplication process. This excess avalanche noise is reduced to minimum when only one type of carrier—for example, the electron in the case of Si APDs—is involved in impact ionizations. One method of achieving single carrier multiplication is by fabricating multilayer devices that have alternating layers of different bandgap semiconductors, as in **multiple quantum well** (MQW) devices discussed in Chapters 3 and 4.

The multilayered structure consisting of many alternating layers of different bandgap semiconductors is called a **superlattice**. Figure 5.20 (a) shows a highly simplified energy band diagram of a superlattice APD. There are thin alternating layers of narrow bandgap (E_{g1}) and wider bandgap (E_{g2}) semiconductor. Their widths are not the same. The wider bandgap semiconductors form the **barrier layers** surrounding the narrower bandgap semiconductors. The bandgap difference $E_{g2} - E_{g1}$ is taken up by the discontinuities ΔE_c and ΔE_v. For example, for GaAs/AlGaAs superlattice, ΔE_c to ΔE_v ratio is roughly 3 to 2. For InGaAs/InAlAs

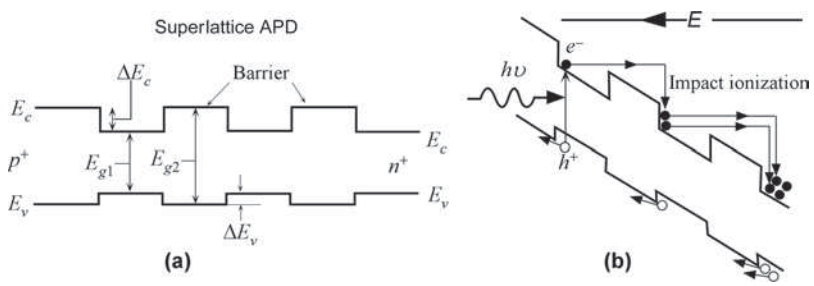


FIGURE 5.20 (a) Energy band diagram of an MQW superlattice APD. (b) Energy band diagram with an applied field and impact ionization.

superlattices, useful in the optical communications wavelengths 1.3–1.55 μm, ΔE_c to ΔE_v ratio is roughly 2 to 1.

When a large field E is applied, the whole band diagram is bent as shown in Figure 5.20 (b) to account for the potential energy of the electron (charge × voltage). An electron in the conduction band (CB) in the E_{g1}-layer accelerates, passes into the E_{g2}-layer, continues to accelerate, and when it arrives and enters the E_{g1}-layer again it has high kinetic energy, and also an additional energy ΔE_c above E_c in this E_{g1}-layer. With its high kinetic energy and the additional ΔE_c, it can easily impact ionize the E_{g1}-semiconductor as shown in Figure 5.20 (b). Of course, the hole accelerates as well but ΔE_v is small. Thus, the electron impact ionization is enhanced much more than the hole impact ionization, a favorable outcome in terms of avalanche multiplication noise. In fact, with proper design, hole impact ionization in this structure can be prevented.

Another important superlattice structure is the **staircase superlattice APD** in which the bandgap is graded within each layer as shown in Figure 5.21 (a) without any bias. The bandgap in each layer changes from a minimum E_{g1} to a maximum E_{g2} which is more than twice E_{g1}. There is a step change ΔE_c in the conduction band edge between two neighboring graded layers that is greater than E_{g1}.

In very simple terms, as illustrated in Figure 5.21 (b), when a bias voltage is applied, that is a field E, the bands bend down. The photogenerated electron initially drifts in the graded layer conduction band. When the electron drifts into the neighboring layer, it now has a kinetic energy ΔE_c above E_c in this layer. It therefore enters the neighboring layer as a highly energetic electron and loses the excess energy ΔE_c by impact ionization. The process repeats itself from layer to layer, leading to an avalanche multiplication of the photogenerated electron. Since the impact ionization is primarily achieved as a result of transition over ΔE_c, the device does not need the high fields typical of avalanche multiplication in bulk semiconductors; it can operate at lower fields. Further, the impact-ionized holes experience only a small ΔE_v which is insufficient to lead to multiplication. Thus, effectively, only electrons are multiplied and the device behaves as if it were a **solid state photomultiplier**.

Such staircase superlattice APDs are difficult to fabricate and typically involve varying the composition of a quaternary III–V semiconducting alloy to obtain the necessary bandgap grading. Superlattice structures that are simply alternating layers of low and high bandgap semiconductor layers are easier to fabricate. Molecular beam epitaxy (MBE) is typically used to fabricate such multilayer structures.

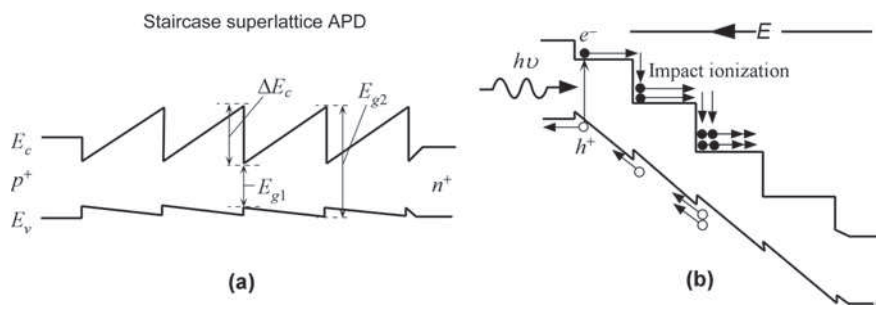


FIGURE 5.21 (a) Energy band diagram of a staircase superlattice APD without an applied bias, and (b) with an applied bias and impact ionization taking place.

5.8 SCHOTTKY JUNCTION PHOTODETECTOR

We consider the junction formed when a metal and an *n*-type semiconductor (*n*-SC) are brought into contact. In practice this process is frequently carried out by the evaporation of a metal onto the surface of a semiconductor crystal in vacuum. The energy band diagrams for the metal and the semiconductor are shown in Figure 5.22 (a). The work function, denoted as Φ, is the energy difference between the vacuum level and the Fermi level. Vacuum level defines the energy where the electron is free from that particular solid, and where the electron has zero KE.

We assume that the work function of the metal is greater than that of the semiconductor, $\Phi_m > \Phi_n$. When the two solids come into contact, the more energetic electrons in the CB of the *n*-SC can readily tunnel into the metal in search of lower empty energy levels (just above E_{Fm}) and accumulate near the surface of the metal as illustrated in Figure 5.22 (a). Electrons tunneling from the *n*-SC leave behind an electron-depleted region of width W in which there are exposed positively charged donors, in other words, net positive space charge. A **built-in potential** V_o therefore develops between the metal and the *n*-SC. There is obviously also a **built-in electric field** E_o from the positive charges to the negative charges on the metal surface as shown in Figure 5.22 (b). Eventually this built-in potential reaches a value that prevents further accumulation of electrons at the metal surface, and equilibrium is reached. The depletion region has been depleted of free carriers (electrons) and hence contains the exposed positive donors. The situation is actually similar to the *pn* junction. This region thus constitutes a **space charge layer** in which there is a nonuniform internal field directed from the semiconductor to the metal surface. The maximum value of this built-in field is denoted as E_o and occurs at the metal–semiconductor junction (this is where there are maximum number of field lines from positive to negative charges).

Once in contact, the Fermi level throughout the whole solid, the metal and *n*-SC in contact, must be uniform in equilibrium. Otherwise a change in the Fermi level, ΔE_F, going from one end to the other end will be available to do external (electrical) work. Thus, E_{Fm} and E_{Fn} line up as shown in Figure 5.22 (b). The SCL, however, has been depleted of electrons so in this region $E_c - E_{Fn}$

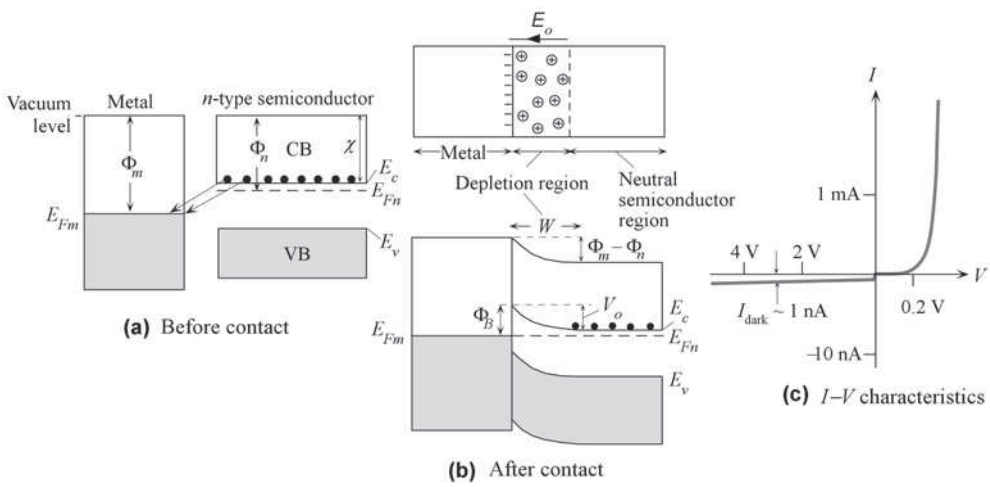


FIGURE 5.22 (a) Metal and an *n*-type semiconductor before contact. The metal work function Φ_m is greater than that of the *n*-type semiconductor. (b) A Schottky junction forms between the metal and the semiconductor. There is a depletion region in the semiconductor next to the metal and a built-in field E_o. (c) Typical *I*–*V* characteristics of a Schottky contact device.

must increase toward the junction so that n decreases. The bands must bend to increase $E_c - E_{Fn}$ toward the junction as illustrated in Figure 5.22 (b). Far away from the junction we, of course, still have an n-type semiconductor. The bending is just enough for the vacuum level to be continuous and changing by $\Phi_m - \Phi_n$ from the semiconductor to the metal as this much energy is needed to take an electron across from the semiconductor to the metal. The built-in potential energy barrier eV_o is simply $\Phi_m - \Phi_n$, given this is the amount of band bending, that is, the change in E_{Fn} to line it up with E_{Fm}. The PE barrier for electrons moving from the metal to the semiconductor is called the **Schottky barrier height**, Φ_B, which is greater than eV_o.

The I–V characteristic of the Schottky junction is shown in Figure 5.22 (c), and is very similar to the pn junction.[16] As apparent from Figure 5.22 (b), the barrier against electron injection from the n-SC to the metal is eV_o. The barrier against electron injection from the metal into the n-SC is Φ_B. In equilibrium, these two injection rates (they depend exponentially on the barrier heights) are small and just balance each other. Under forward bias, with the positive terminal connected to the metal and negative to the n-SC, the applied voltage V drops across the SCL, which reduces the built-in voltage to $V_o - V$. The barrier Φ_B remains unaltered. Since the injection rate depends on the Boltzmann factor $\exp\left[-e(V_o - V)/k_BT\right]$, there is an increase in this rate by a factor of $\exp(eV/k_BT)$, which leads to a very large rate of injection from the n-SC to the metal. The forward current is therefore large and depends exponentially on V.

Under reverse bias V_r, as shown in Figure 5.23 (a), V_o increases to $V_o + V_r$. The electron injection rate from n-SC to the metal vanishes and the current is dominated by the small rate of injection from the metal to n-SC over Φ_B and depends on $\exp(-\Phi_B/k_BT)$. The reverse current in the dark I_d depends on the nature of the metal to semiconductor contact (Φ_B) and the device area. It is smaller for wider bandgap semiconductors; some values are shown in Table 5.4 where I_d ranges from a few femtoamps to microamps per mm^2 of device area.

As a photodiode, the Schottky junction (SJ) is reverse biased so that the field in the depletion region is large. If we illuminate the depletion region of the SJ with photons of energy greater than E_g the photogeneration will take place within the SCL, where there is a strong field. The electrons and holes will roll down their respective energy hills as shown in Figure 5.23 (b), that is, they drift in this depletion region as in Figure 5.23 (c). For photon energies less than E_g the device can still respond, providing that the $h\upsilon$ can excite an electron from E_{Fm} in the metal over the PE barrier Φ_B into the CB, from where the electron will roll down toward the neutral

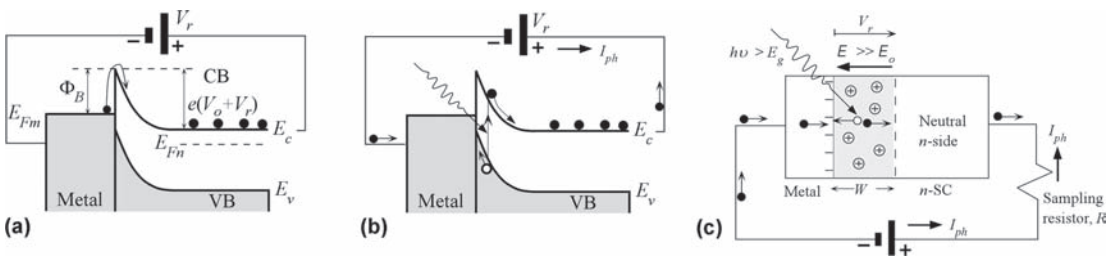


FIGURE 5.23 (a) Reverse biased Schottky junction and the dark current due to the injection of electrons from the metal into the semiconductor over the barrier Φ_B. (b) Photogeneration in the depletion region and the resulting photocurrent. (c) The Schottky junction photodetector.

[16] A straightforward derivation of the Schottky junction current–voltage characteristics can be found in S. O. Kasap, *Principles of Electronic Materials and Devices*, 2nd Edition (McGraw-Hill, 2006), Ch. 5.

TABLE 5.4 Schottky junction–based photodetectors and some of their features. τ_R and τ_F are the rise and fall times of the output of the photodetector for an optical pulse input

Schottky junction	λ range (nm)	R_{peak} (at peak) (A W^{-1})	J_{dark} (per mm^2)	Features with typical values
GaAsP	190–680	0.18 (610 nm)	5 pA	UV to red, $\tau_R = 3.5$ μs (G1126 series[a])
GaP	190–550	0.12 (440 nm)	5 pA	UV to green, $\tau_R = 5$ μs (G1961[a])
AlGaN	220–375	0.13 (350 nm)	1 pA	Measurement of UV; blind to visible light (AG38S[b])
GaAs	320–900	0.2 (830 nm)	~1 nA	Wide bandwidth > 10 GHz, $\tau_R < 30$ ps (UPD-30-VSG-P[c])
InGaAs MSM	850–1650	0.4 (1300 nm)	5 μA	Optical high speed measurements, $\tau_R = 80$ ps, $\tau_F = 160$ ps (G7096[a])
GaAs MSM	450–870	0.3 (850 nm)	0.1 nA	Optical high speed measurements, $\tau_R = 30$ ps, $\tau_F = 30$ ps (G4176[a])

Note: The rise and fall times represent the times required for the output to rise from 10% to 90% of its final steady state value and to fall from 90% to 10% of its value before the optical pulse is turned off. [a]Hamamatsu (Japan); [b]sglux (Germany); [c]Alphalas (Germany).

n-region. In this case $h\nu$ must only be greater than Φ_B. Note also that the SJ photodiode can also be operated in the *photovoltaic mode*, *i.e.,* with no applied bias, since the built-in field E_o can easily separate and drift the photogenerated carriers in the SCL.

One distinct advantage of a SJ photodiode is that the SCL is right next to the metal contact. The short wavelength light that would be absorbed in the neutral region next to the electrode in a *pn* or *pin* photodiode is now absorbed in the SCL in the SJ photodiode. Consequently, the EHPs photogenerated can be separated immediately, drifted, and collected, whereas the minority carriers have to diffuse to the SCL in the *pn* and *pin* photodiodes. In fact, in *pn* and *pin* photodiodes, many of the EHPs diffuse to the surface and disappear by recombination. Clearly, SJ photodiodes are well suited for detecting short-wavelength light that is absorbed very close to the metal–semiconductor interface; and there are many SJ photodiodes for use in the UV region as summarized in Table 5.4. Further, the SJ diodes tend to have a wide spectral responsivity, *e.g.,* from the UV to the red.

Another advantage of SJ photodiodes is that they can be significantly faster than *pn* or *pin* photodiodes. Injected carriers in both forward and the reverse bias are electrons, which are majority carriers and therefore do not suffer from the limitations of minority carrier recombination time. In high-speed applications, such as in optical communications or optical measurements, the SJ photodiodes are used in pairs in a so-called interdigital electrode structure on the surface of the semiconductor as shown in Figure 5.24 (a) and (b). The light is absorbed in the semiconductor between the electrodes. We consider two of the digital electrodes, *A* and *B*, which would have been connected to positive and negative terminals of the battery as in Figure 5.24 (a). The Schottky junctions face each other, which means that whatever the polarity of the applied bias, one would always be reverse biased, *e.g., A*, and the other, *B*, forward biased. Without any bias, the energy band diagram is symmetric as shown in Figure 5.24 (c). The SCLs for *A* and *B* are shown as SCL_1 and SCL_2, respectively. The built-in voltages of *A* and *B* junctions are V_{o1} and V_{o2}, respectively. If an EHP is photogenerated in the semiconductor, they cannot drift and become collected as in the single SJ device. They are stuck between the two potential barriers (eV_{o1} and eV_{o2}) as indicated in Figure 5.24 (c).

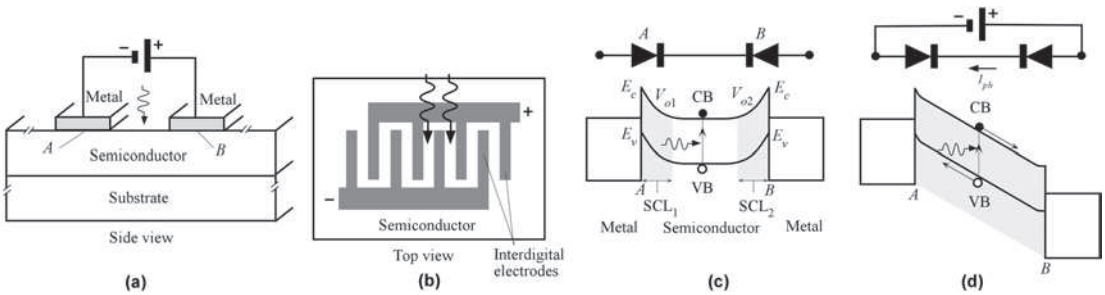
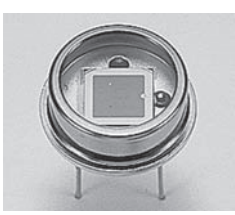


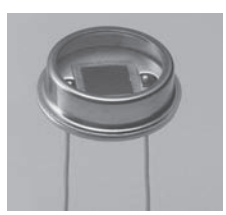
FIGURE 5.24 (a) The metal electrodes are on the surface of the semiconductor crystal (which is grown on a suitable substrate). (b) The electrodes are configured to be interdigital and on the surface of the crystal. (c) Two neighboring Schottky junctions are connected end-to-end, but in opposite directions as shown for *A* and *B*. The energy band diagram without any bias is symmetrical. The gray areas represent the SCL_1 and SCL_2 at *A* and *B*. (d) Under a sufficiently large bias, the SCL_1 from *A* extends and meets that from *B* so that the whole semiconductor between the electrodes is depleted. There is a large field in this region, and the photogenerated EHPs become separated and then drifted, which results in a photocurrent.

In contrast, in a single SJ, the EHP photogenerated inside the SCL can be collected as they become separated and drifted.

When a voltage is applied as in Figure 5.24 (d), V_{o1} is increased and V_{o2} is reduced since *A* is reverse and *B* is forward biased. The dark current is very small since *A* is reverse biased. However, we must bend the right-hand side (*B*) of the band down by eV_r to account for the applied voltage as shown in Figure 5.24 (d), because when the electron from *A* reaches *B* its PE must have dropped by eV_r. The applied bias extends the width of SCL_1 in the semiconductor and shrinks SCL_2, which means that further voltage increases will drop more across SCL_1 than SCL_2. With a sufficiently large voltage, the reverse bias at *A* will extend SCL_1 so much that it will punch through, or *reach-through*, onto the SCL_2, and the whole semiconductor between the electrodes becomes depleted. The corresponding bias voltage is called the **reach-trough voltage** (also known as the **flat-band voltage**). The advantage is that there is now a strong field in this depletion region and photogeneration would create an electron and a hole that can roll down their respective energy hills, *i.e.,* drift toward the electrodes, and give rise to a photocurrent. The photodiode we have considered is based on the basic **metal-semiconductor-metal** (MSM) structure, as in Figure 5.24 (a), that has two Schottky junctions facing each other as in Figure 5.24 (c). These devices are marketed as **MSM photodiodes**; two examples based on GaAs and InGaAs are listed in Table 5.4 and are capable of high-speed operation.

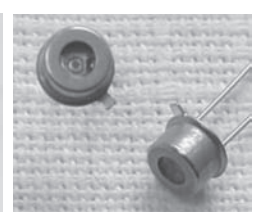


GaAsP Schottky junction photodiode for 190–680 nm detection, from UV to red. *(Courtesy of Hamamatsu.)*



GaP Schottky junction photodiode for 190–550 nm detection. *(Courtesy of Hamamatsu.)*

Schottky junction–type metal-semiconductor-metal (MSM)-type photodetectors. *(Courtesy of Hamamatsu.)*



AlGaN Schottky junction photodiode for UV detection. *(Courtesy of sglux, Germany.)*

5.9 PHOTOTRANSISTORS

The **phototransistor** is a bipolar junction transistor (BJT) that operates as a photodetector with a photocurrent gain. The basic principle is illustrated in Figure 5.25. In an ideal device, only the depletion regions, or the space charge layers, contain an electric field. The base terminal is normally open and there is a voltage applied between the collector and emitter terminals just as in the normal operation of a common emitter BJT. An incident photon is absorbed in the SCL between the base and collector to generate an electron–hole pair. The field E in the SCL separates the electron hole pair and drifts them in opposite directions. This is the **primary photocurrent** and it effectively constitutes a base current even though the base terminal is open circuit (current is flowing into the base from the collector rather than from the external base terminal). When the drifting electron reaches the collector, it becomes collected (and thereby neutralized) by the battery. On the other hand, when the hole enters the neutral base region, it can only be neutralized by injecting a large number of electrons into the base, as explained below. It effectively forces a large number of electrons to be injected from the emitter. Typically the electron recombination time in the base is very long compared with the time it takes for the electron to diffuse across the base. This means that only a small fraction of electrons injected from the emitter can recombine with photogenerated holes that have entered the base. Thus, the emitter has to inject a large number of electrons to neutralize any extra hole in the base. These electrons (except one) diffuse across the base and reach the collector and thereby constitute an amplified photocurrent.

Alternatively, in a more intuitive way, one can argue that the photogeneration of EHPs in the collector SCL decreases the resistance of this region, which decreases the voltage V_{CB} across the base-collector junction. Consequently the base-emitter voltage V_{BE} must increase inasmuch as $V_{BE} + V_{CB} = V_{CC}$ (Figure 5.25). This increase in V_{BE} acts as if it were a forward bias across the base-emitter junction and injects electrons into the base due to the transistor action, that is, $I_E \propto \exp(eV_{BE}/k_BT)$.

Since the photon-generated primary photocurrent I_{pho} is amplified as if it were a base current (I_B), the photocurrent flowing in the external circuit is

Photocurrent in a phototransistor

$$I_{ph} \approx \beta I_{pho} \tag{5.9.1}$$

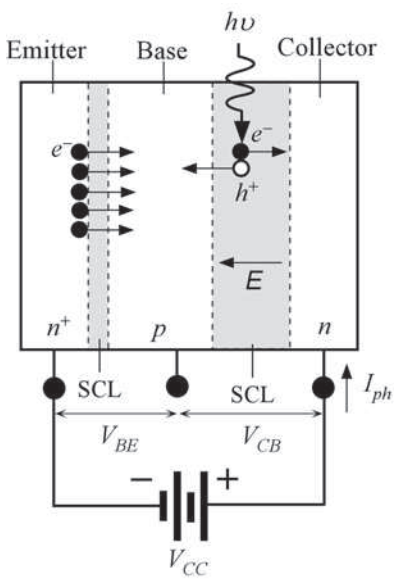


FIGURE 5.25 The principle of operation of the phototransistor. SCL is the space charge layer or the depletion region. The primary photocurrent acts as a base current and gives rise to a large photocurrent in the emitter-collector circuit.

where β is the current gain (or h_{FE}) of the transistor. The phototransistor construction normally allows the incident radiation to be absorbed in the base–collector junction SCL. Most commercial phototransistors only have two terminals, the emitter and collector, and generally do not have an external base terminal.

It is possible to construct a **heterojunction phototransistor** that has different bandgap materials for the emitter, base, and collector. For example, if the emitter in Figure 5.25 is InP ($E_g = 1.35$ eV) and the base is an InGaAsP alloy (*e.g.*, $E_g \approx 0.85$ eV) then photons with energies less than 1.35 eV but more than 0.85 eV will pass through the emitter and become absorbed in the base. This means the device can be illuminated through the emitter.

5.10 PHOTOCONDUCTIVE DETECTORS AND PHOTOCONDUCTIVE GAIN

Photoconductive detectors have the simple structure schematically shown in Figure 5.26. Two electrodes are attached to a semiconductor that has the desired absorption coefficient and quantum efficiency over the wavelengths of interest. A bias voltage V is applied to the electrodes. Incident photons become absorbed in the semiconductor and photogenerate electron–hole pairs (EHPs). The result is an increase in the conductivity of the semiconductor and hence an increase in the external current, which constitutes the photocurrent I_{ph} as illustrated in Figure 5.26.

The actual response of the detector depends on whether the contacts to the semiconductor are ohmic or blocking (*e.g.*, a reverse biased Schottky junction that does not inject carriers), and on the nature of carrier recombination kinetics. We will consider a photoconductor with ohmic contacts, that is, the contacts do not limit the current flow. With ohmic contacts, the photoconductor exhibits **photoconductive gain**, that is, *the external photocurrent is due to more than one electron flow per absorbed photon*, as illustrated in Figure 5.27 and explained below.

An absorbed photon photogenerates an EHP, which drift in opposite directions as shown in Figure 5.27 (a). The electron drifts much faster than the hole and therefore leaves the sample quickly. The sample, however, must be neutral which means another electron must enter the sample from the negative electrode as in (b) (the electrode is ohmic). This new electron also drifts across quickly as in (b) and (c) to leave the sample while the hole is still drifting slowly in the sample. Thus, another electron must enter the sample to maintain neutrality as in (d) and (e),

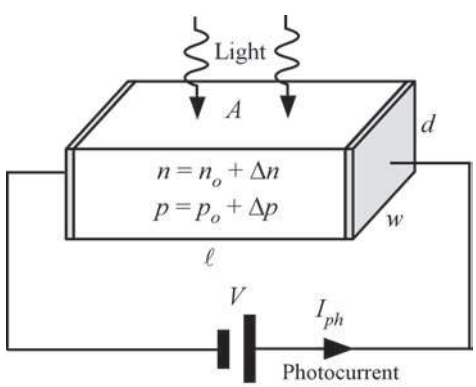
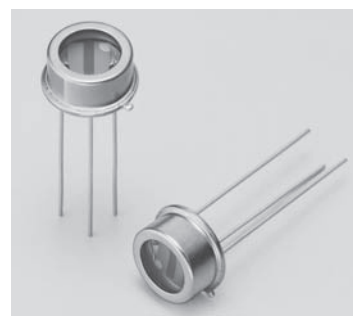


FIGURE 5.26 A semiconductor slab of length ℓ, width w, and depth d is illuminated with light of wavelength λ.



PbS (lead sulfide) photoconductive detectors for the detection of IR radiation up to 2.9 μm. They are typically used in such applications as radiation thermometers, flame monitors, water content and food ingredient analyzers, and spectrophotometers (P9217 series). *(Courtesy of Hamamatsu.)*

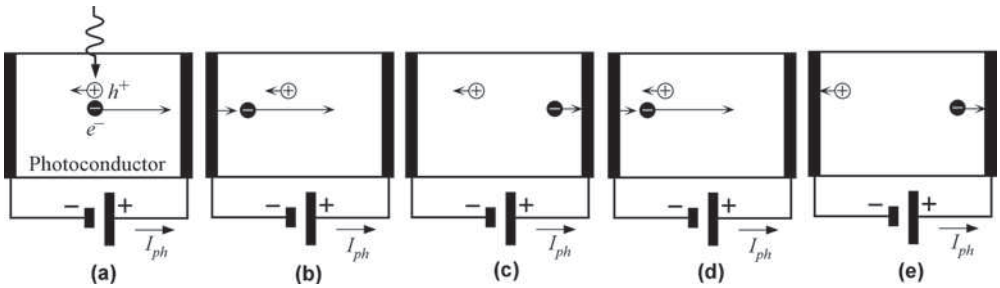


FIGURE 5.27 A photoconductor with ohmic contacts (contacts not limiting carrier entry) can exhibit gain. As the slow hole drifts through the photoconductors, many fast electrons enter and drift through the photoconductor because, at any instant, the photoconductor must be neutral. Electrons drift faster which means that as one leaves, another must enter. (a) An electron and hole are photogenerated. The hole drifts slowly but the electron drifts quickly and leaves the sample. (b) To maintain neutrality, an electron is injected from the negative electrode, which drifts quickly. (c) The electron reaches the positive electrode and leaves the sample while the hole is still slowly drifting. Figures (d) and (e) show the process in (b) and (c) repeated, except that the hole has drifted closer to the negative electrode. Eventually, the hole reaches the negative electrode, shortly after (e), and the photocurrent ceases. The hole can also recombine with one of the injected electrons during its drift.

and so on, until either the hole reaches the negative electrode or recombines with one of these electrons entering the sample. The external photocurrent therefore corresponds to the flow of many electrons per absorbed photon which represents a gain, called **photoconductive gain**. The gain depends on the drift time of the carriers and their recombination lifetime.

Suppose that the photoconductor is suddenly illuminated by a step light as shown in Figure 5.26. If Φ_{ph} is the number of photons arriving per unit area per unit second (the photon flux), then $\Phi_{\text{ph}} = I/h\nu$, where I is the light intensity (radiation energy flowing per unit area per second) and $h\nu$ is the photon energy. If all incident photons are absorbed then the number of EHPs generated per unit volume per second, *i.e.,* the photogeneration rate per unit volume g_{ph}, is given by

$$g_{\text{ph}} = \frac{\eta_i A \Phi_{\text{ph}}}{Ad} = \frac{\eta_i \left(\frac{I}{h\nu}\right)}{d} = \frac{\eta_i I \lambda}{hcd} \tag{5.10.1}$$

where A is the illuminated surface area (ℓw) and η_i is the internal quantum efficiency.

Suppose that at any instant the electron concentration is n (includes photogenerated electrons) and the thermal equilibrium concentration (in the dark) is n_o. Then $\Delta n = n - n_o$ is the **excess electron concentration**. For photogeneration, $\Delta n = \Delta p$ inasmuch as we need to maintain charge neutrality in the sample. At any instant in the sample we have,

$$\begin{bmatrix}\text{The rate of increase in}\\ \text{the excess electron concentration}\end{bmatrix} = \begin{bmatrix}\text{Rate of photogeneration}\\ \text{of excess electrons}\end{bmatrix} - \begin{bmatrix}\text{Rate of recombination}\\ \text{of excess electrons}\end{bmatrix}$$

If τ is the mean recombination time of excess electrons, then

$$\frac{d\Delta n}{dt} = g_{\text{ph}} - \frac{\Delta n}{\tau} \tag{5.10.2}$$

It is clear from Eq. (5.10.2) that Δn increases exponentially from the instant the light is turned on until a steady state is reached when

$$\frac{d\Delta n}{dt} = g_{\text{ph}} - \frac{\Delta n}{\tau} = 0$$

so that

$$\Delta n = \tau g_{\text{ph}} = \frac{\tau \eta_i I \lambda}{hcd} \tag{5.10.3}$$

The conductivity of a semiconductor is given by $\sigma = e\mu_e n + e\mu_h p$, so that the change in the conductivity, called **photoconductivity**, is

Photo-conductivity

$$\Delta\sigma = e\mu_e \Delta n + e\mu_h \Delta p = e\Delta n(\mu_e + \mu_h) \tag{5.10.4}$$

since electrons and holes are generated in pairs, $\Delta n = \Delta p$. Thus, substituting for Δn in the $\Delta\sigma$ expression above we get

Photo-conductivity

$$\Delta\sigma = \frac{e\eta_i I \lambda \tau(\mu_e + \mu_h)}{hcd} \tag{5.10.5}$$

The photocurrent density is simply

$$J_{\text{ph}} = \Delta\sigma \frac{V}{\ell} = \Delta\sigma E \tag{5.10.6}$$

where E is the applied field. The number of electrons flowing in the external circuit can be found from the photocurrent because

$$\text{Rate of electron flow} = \frac{I_{ph}}{e} = \frac{wdJ_{\text{ph}}}{e} = \frac{w\eta_i I \lambda \tau(\mu_e + \mu_h)E}{hc} \tag{5.10.7}$$

However, the rate of electron (*i.e.*, EHP) photogeneration is

$$\text{Rate of electron generation} = (\text{Volume})g_{\text{ph}} = (wd\ell)g_{\text{ph}} = w\ell \frac{\eta_i I \lambda}{hc}$$

The photoconductive gain is then simply

Photo-conductive gain

$$G = \frac{\text{Rate of electron flow in external circuit}}{\text{Rate of electron generation by light absorption}} = \frac{\tau(\mu_e + \mu_h)E}{\ell} \tag{5.10.8}$$

Equation (5.10.8) can be simplified further by noting that the drift velocities of the electrons and holes in the photoconductor are $\mu_e E$ and $\mu_h E$ respectively, so that their corresponding transit times (time to cross the semiconductor) are

$$t_e = \ell/(\mu_e E) \quad \text{and} \quad t_h = \ell/(\mu_h E)$$

Using these transit times in Eq. (5.10.8) we can obtain

Photo-conductive gain

$$G = \frac{\tau}{t_e} + \frac{\tau}{t_h} = \frac{\tau}{t_e}\left(1 + \frac{\mu_h}{\mu_e}\right) \tag{5.10.9}$$

The photoconductive gain can be quite high if τ/t_e is kept large, which means a long recombination time and a short transit time. The transit time can be made shorter by applying a

greater field but this will also lead to an increase in the dark current and more noise. The speed of response of the device is limited by the recombination time of the injected carriers. A long τ means a slow device.

5.11 BASIC PHOTODIODE CIRCUITS

The reverse biased photodiode (PD) mode of operation is shown in Figure 5.28 (a). We will assume steady state operation. There is usually a resistor R_L in the reverse bias circuit, which may be a load or a sampling resistor across which the signal is taken, or simply a current limiting resistor to reverse bias the PD. The definitions for positive I and V, by convention, are for the case in which the PD is forward biased; they are also shown in Figure 5.28 (a). The I–V characteristics of the photodiode are shown in Figure 5.28 (b), where both I and V have negative values representing reverse current and voltage. In the dark, there is a small reverse current, which usually increases with the reverse bias. (Ideally, it should be simply constant and equal to the reverse saturation current.) Upon illumination, a photocurrent I_{ph} is generated in the photodiode; I_{ph} is proportional to the incident light power P_o through the responsivity R. If the photodiode terminals were shorted, the short circuit current I_{sc} under illumination would be $-I_{ph}$ as indicated in Figure 5.28 (b).

The current through R and the voltage across it obey Ohm's law. However, care must be taken to properly account for the current and voltage signs. From Figure 5.28 (a), the current through R_L is

$$I = \left[(-V) - (+V_r)\right]/R_L = -(V + V_r)/R_L$$

This equation is called the **load line** because it represents the load R on the same axes I and V as the photodiode.[17] For $I = 0$, open circuit $V = -V_r$, and is a point on the load line. For $V = 0$, $I = -V_r/R$, and this is also a point on the load line. The load line can thus be quickly drawn to pass through these two points as shown in Figure 5.28 (b). It has a negative slope

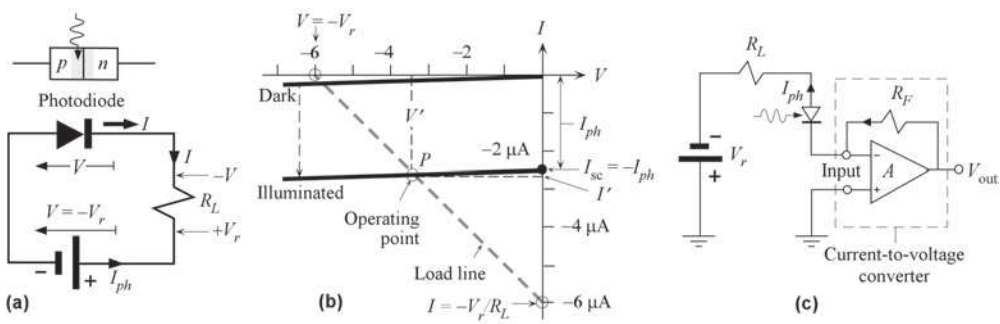


FIGURE 5.28 (a) The photodiode is reverse biased through R_L and illuminated. Definitions of positive I and V are shown as if the photodiode were forward biased. (b) I–V characteristics of the photodiode with positive I and V definitions in (a). The load line represents the behavior of the load R. $V_r = 6$ V and $R = 1$ MΩ have been used. The operating point is P, where the current and voltage are I' and V'. (c) A simple circuit for the measurement of the photocurrent I_{ph} by using a current–voltage converter or a transimpedance amplifier. (The reverse bias V_r is a positive number.)

[17] It is left as an exercise to show that if we move the resistor R_L from the right to the left side of the circuit, we would get the same load line equation.

of $-1/R_L$. It cuts the PD characteristics at P, which represents the operating point of the circuit; the current and voltage will be I' and V' as indicated in Figure 5.28 (b). The particular example in Figure 5.28 (b) is for $V_r = 6\,\text{V}$ and $R = 1\,\text{M}\Omega$.

It is apparent from Figure 5.28 (b) that as long as the PD is reverse biased, the magnitude of the operating current I' under illumination is very close to the generated photocurrent I_{ph} (the short circuit current), *i.e.,* $I' \approx -I_{ph}$. This is not the case in the photovoltaic mode in which there is a load resistor right across the PD without any V_r. Further, we can redirect the photocurrent into a current-to-voltage converter to convert I_{ph} to an output voltage V_{out} as in Figure 5.28 (c), in which the output $V_{\text{out}} = R_F I_{ph}$. The feedback resistor R_F determines the transimpedance gain V_{out}/I_{ph}. Notice that V_{out} increases with increasing I_{ph}, that is, the gain is positive.

There are many applications for which we need to represent the PD by an equivalent circuit. The basic operation of a reverse biased PD is shown in Figure 5.29 (a) where A and B represent the external terminals of the PD. Suppose the EHP generation occurs in the SCL; then the drift of electrons and holes will give rise to a photocurrent I_{ph}. We can represent this effect with a current generator in the equivalent circuit as shown in Figure 5.29 (b). The points A' and B' represent the terminals of the ideal *pn* junction diode that generates the photocurrent I_{ph}. The origin of the capacitance C_t and resistances R_s and R_p in the equivalent circuit are explained below. A and B are the actual external terminals of the PD. The definitions for positive current and voltage are shown as I and V. I flows into A and V is the voltage at A with respect to B. Notice that, due to a voltage drop across R_s, the voltage that appears across the *pn* junction (between A' and B') is V_1.

If we are not limited by the drift and diffusion times of the photogenerated carriers, then I_{ph} will follow the input optical power modulation. Under reverse bias, $V = -V_r$, and the voltage across the ideal *pn* junction V_1 is nearly $-V_r$ because R_s is usually small. There will also be some dark current I_d from the normal diode action (or *pn* junction behavior), *i.e.,* with reverse bias, $I_{\text{diode}} = I_{so}\left[\exp(eV_1/k_BT) - 1\right] = -I_{so}$, reverse saturation current. This reverse current is represented by an ideal diode D in parallel with the current generator as shown in Figure 5.29 (b). I_{diode} is the ideal diode current and along the positive current direction for D. The dark current I_d is quoted as a positive number in data sheets so that $I_d = -I_{\text{diode}} = I_{so}$.

Since the PD is reverse biased, there will be a depletion region capacitance C_{dep} appearing across the PD terminals as in Figure 5.29 (a). We also need to consider any stray capacitances, including terminal to terminal capacitance (so-called packaging capacitances), all of

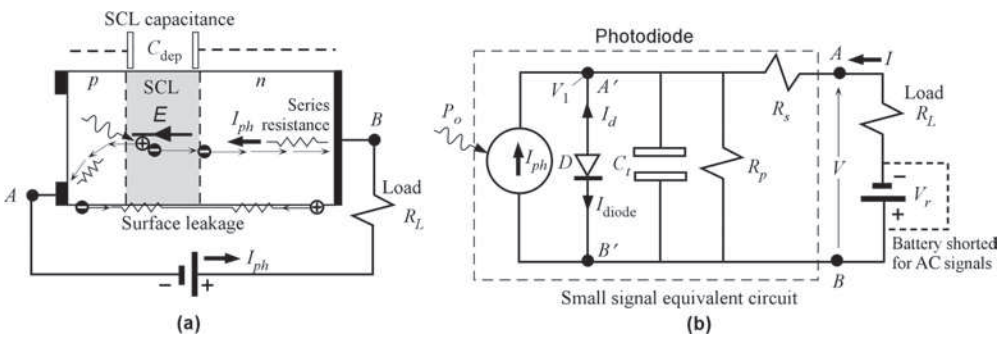
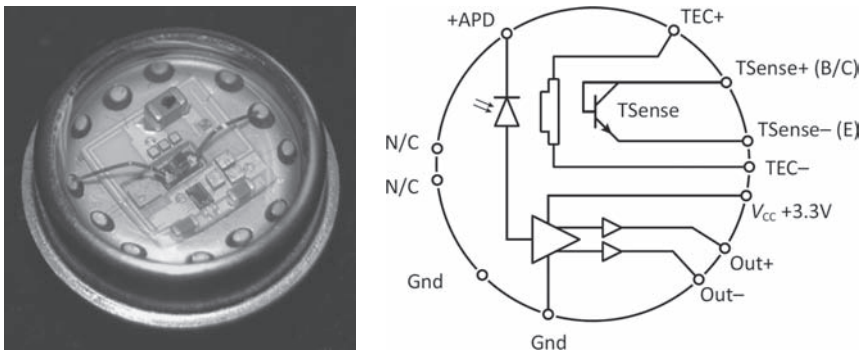


FIGURE 5.29 (a) A real photodiode has series and parallel resistances R_s and R_p and an SCL capacitance C_{dep}. A and C represent anode and cathode terminals. (b) The equivalent circuit of a photodiodes. For AC (or transient) signals, the battery can be shorted since AC signals will simply pass through the battery.

which will add to C_{dep}, and can be lumped into a single total capacitance called the **terminal capacitance** C_t. In Figure 5.29 (b), C_t appears across the photocurrent generator. Because of C_{dep}, C_t decreases with increasing reverse bias, and is quoted at a certain bias voltage in data sheets. It depends on the diode area, and can be anything from a few picofarads (small area PDs) to a few nanofarads.

The photocurrent inside the PD also has to flow through the neutral *p*- and *n*-sides, that is, through bulk semiconductor regions that will have a finite resistance (not zero) as shown in Figure 5.29 (a). We can represent the bulk semiconductor regions by introducing a **series resistance** R_s into the equivalent circuit as shown in shown in Figure 5.29 (b). There will also be surface leakage currents which represent the flow of carriers on the surfaces of the crystal. The net resistance of all these surface paths are represented by using a **shunt resistance** or a **parallel resistance** R_p across the current generator as shown in Figure 5.29 (b). The load R_L is connected to the external terminals *A* and *B* of this equivalent circuit. Although Figure 5.29 (b) has also included the battery (V_r) to indicate reverse bias, this would be shorted in using the circuit for AC signals. The equivalent circuit shown in Figure 5.29 (b) is probably one of the simplest but quite useful in predicting the output from a PD. It can also be used for photovoltaic behavior (without a battery but just a load across *A* and *B*).



A photoreceiver that has an InGaAs APD and peripheral electronics (ICs) to achieve high gain and high sensitivity. There is also a thermoelectric cooler (TEC) and a temperature sensor (TSense). *(Courtesy of Voxtel Inc.)*

Photodiodes are widely used in optoelectronics to measure modulated light signals, which may be digital, in the form of optical pulses, or analog-type, in which the amplitude of the light wave is modulated. As a rough estimate of the PD speed, we can assume R_p is very large, and can be ignored. Further, R_s and the load R_L in Figure 5.29 (b) can be lumped into one resistance, but R_s is typically much smaller than R_L. We then have C_t and R_L in parallel with a time constant $R_L C_t$. The cutoff frequency f_c or the bandwidth of the PD will then be determined by $R_L C_t$, *i.e.*, $f_c = 1/(2\pi R_L C_t)$. Manufacturers usually quote C_t and the bandwidth (the cutoff frequency) for a given load (R_L). Suppose that we make $R_L C_t$ as small as possible, then the bandwidth will be determined by the transit and diffusion times of carriers inside the PD as discussed above.

We can easily build a simple and fast PD circuit as shown in Figure 5.30 (a). The PD is reverse biased through R_B, and C_B shorts both R_B and the battery for transient signals; it is an AC short. There is a small load or a sampling resistor R_L. The transient photocurrent generated by an optical pulse develops a transient voltage across R_L, which can be coupled into a fast buffer amplifier.

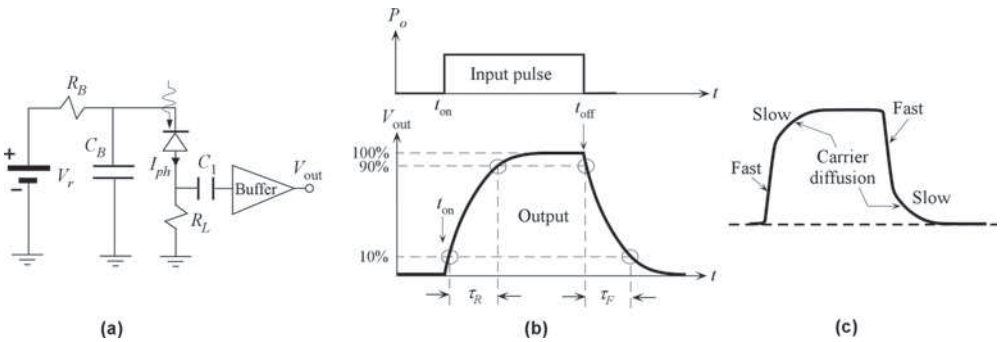


FIGURE 5.30 (a) The photodiode is reverse biased by V_r through R_B. C_B shorts both the R_B and the battery for transient signals. C_1 is a signal coupling capacitor to pass the photocurrent signal into a fast buffer amplifier. (b) Definitions of rise and fall times in the PD output in response to a pulse optical excitation. (c) Carrier diffusion, which is a slow process, results in slow regions in the output pulse shape.

The time constant $R_L C_t$ must be smaller than the optical pulse to be measured. C_t in the circuit of Figure 3.30 (a) should include the capacitance of the buffer amplifier as well, which maybe comparable to that of the PD. Instead of using a sampling resistor R_L, one can, of course, feed the output directly into the input of a fast current to voltage converter amplifier as in Figure 5.28 (c). The critical time constant is then $R_F C_t$.

Manufacturers' data sheets typically provide rise and fall times τ_R and τ_F in the output waveform when the PD is excited from an optical pulse. τ_R is time required for the output to rise from its 10% to 90% of final value. τ_F is the time needed for the output to fall from its 90% to 10% value after it has reached a stated state before the turn-off. Both definitions are shown in Figure 5.30 (b). When the $R_t C_t$ time constant is very small, the rise and fall times are limited by the drift and diffusion of photogenerated carriers in the PD. If the photogeneration occurs in the depletion region the rise and fall times would be quite short and limited by transit times across the depletion width W.

The exact output waveform depends on the excitation wavelength and the PD structure. For long wavelength excitation, for which the absorption depth is long, photogeneration will occur both in the SCL (width, W) and within diffusion lengths L_e and L_h from the SCL as shown in Figure 5.8. The output waveform will have an initial fast and then a slow region in both the rise and fall parts of the waveform as shown in Figure 5.30 (c). Initially, the photogenerated carriers in the SCL will drift very quickly through W and give a fast rise in the output. Those carriers generated outside the SCL have to slowly diffuse to the SCL. Upon arriving at the SCL, they will be drifted quickly. The prolonged arrival of these diffusing carriers is responsible for the slow rise part of the waveform. As soon as the optical pulse is turned off, those carriers in W can be drifted and collected quickly, and give rise to a fast initial fall. However, those carriers that are generated outside W have to diffuse to the SCL before they can be drifted and collected, and this gives rise to the slowly decaying final part in the waveform.

5.12 NOISE IN PHOTODETECTORS

A. The *pn* Junction and *pin* Photodiodes

The lowest signal that a photodetector can detect is determined by the extent of random fluctuations in the current through the detector and the voltage across it as a result of various statistical processes in the device. When a *pn* junction is reverse biased, there is still a dark

current I_d present, which is mainly due to the thermal generation of electron–hole pairs in the SCL and within diffusion lengths to the SCL. If the dark current were absolutely constant with no fluctuations, then any change in the PD current, however small (even a tiny fraction of I_d), due to an optical signal could be easily detected by blocking or removing I_d. The dark current however exhibits **shot noise** or fluctuations about I_d, as shown in Figure 5.31. This shot noise is due to the fact that electrical conduction is by *discrete charges*, which means that there is a statistical distribution in the transit times of the carriers across the photodiode. Carriers are collected as discrete amounts of charge (e) that arrive at random times and not continuously. It is unlike a continuous flow of water through a pipe but rather like rolling ball bearings down a pipe at random times. There will be fluctuations in the outflow rate of ball bearings at the collection end.

The *root mean square* (rms) value of the fluctuations in the dark current represents the shot noise current $i_{n\text{-dark}}$,

$$i_{n\text{-dark}} = \left[2eI_dB\right]^{1/2} \tag{5.12.1}$$

Dark current shot noise

where B is the frequency bandwidth of the photodetector. The photocurrent signal must be greater than this shot noise in the dark current.

The photodetection process involves the interaction of discrete photons with valence electrons. The discrete nature of photons means that there is an unavoidable random fluctuation in the rate of arrival of photons even if we did our best to keep the rate constant. The quantum nature of the photon therefore gives rise to a statistical randomness in the EHP photogeneration process. This type of fluctuation is called **quantum noise** (or **photon noise**) and it is equivalent to shot noise as far as its effects are concerned. Thus, the photocurrent will always exhibit fluctuations about its mean value due to quantum noise. If I_{ph} is the mean photocurrent, the fluctuations about this mean has an rms value that is called **shot noise current** due to **quantum noise**

$$i_{n\text{-quantum}} = \left[2eI_{ph}B\right]^{1/2} \tag{5.12.2}$$

Quantum noise

Generally the dark current shot noise and quantum noise are the main sources of noise in *pn* junction– and *pin*-type photodiodes. The total shot noise i_n generated by the photodetector is not a simple sum of Eqs. (5.12.1) and (5.12.2) because the two processes are due to independent

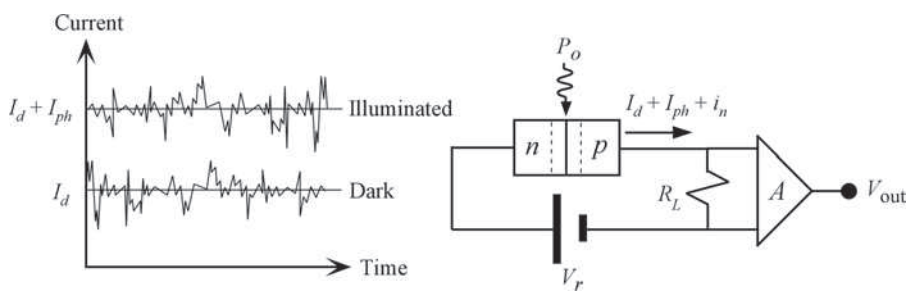


FIGURE 5.31 In *pn* junction and *pin* devices, the main source of noise is shot noise due to the dark current and photocurrent.

random fluctuations. We need to sum the power in each or add the mean squares of the shot noise currents, *i.e.,*

$$i_n^2 = i_{n\text{-dark}}^2 + i_{n\text{-quantum}}^2$$

so that the rms total shot noise current is

Shot noise in pn junction

$$i_n = \left[2e(I_d + I_{ph})B\right]^{1/2} \tag{5.12.3}$$

In Figure 5.31 the photodetector current, $I_d + I_{ph} + i_n$, flows through a load resistor R_L, which acts as a sampling resistor for measuring the current. The voltage across R_L is amplified. In considering the noise of the receiver we also have to include the thermal noise in the resistor R_L and the noise in the input stage of the amplifier. **Thermal noise** is random voltage fluctuations across any conductor due to random motions of the conduction electrons; it is also called **Johnson noise**. In receiver design we are often interested in the **signal to noise ratio**, SNR or S/N, which is defined as the ratio of signal power to noise power,[18]

Signal to noise ratio

$$\text{SNR} = \frac{\text{Signal power}}{\text{Noise power}} \tag{5.12.4}$$

For the photodetector alone, SNR is simply the ratio of I_{ph}^2 to i_n^2. SNR for the receiver must include the noise power generated in the sampling resistor R_L (thermal noise) and in the input elements of the amplifier (*e.g.*, resistors and transistors).

The **noise equivalent power** (NEP) is an important property of a photodetector that is frequently quoted. NEP is defined as the required optical input power to achieve a SNR of 1 within a bandwidth of 1 Hz. Stated differently, it is the optical signal power required to generate a photocurrent signal (I_{ph}) that is equal to the total noise current (i_n) in the photodetector at a given wavelength and within a bandwidth of 1 Hz. Suppose that B is the bandwidth of the detector and P_1 is the incident power that results in SNR $= 1$, then by definition

NEP definition

$$\text{NEP} = \frac{\text{Input power for SNR} = 1}{\sqrt{\text{Bandwidth}}} = \frac{P_1}{B^{1/2}} \tag{5.12.5}$$

If $\mathcal{R}$ is the responsivity and P_o is the monochromatic incident optical power, then the generated photocurrent is

$$I_{ph} = \mathcal{R}P_o \tag{5.12.6}$$

Suppose that the photogenerated current I_{ph} is equal to the noise current i_n in Eq. (5.12.3), when the incident optical power P_o is P_1. Then

$$\mathcal{R}P_1 = \left[2e(I_d + I_{ph})B\right]^{1/2}$$

[18] Some authors define SNR as the ratio of the actual signal magnitude to the rms noise, that is, the square root of the definition in Eq. (5.12.4). The definition used here is common in optical communications. On the other hand, imaging applications tend to define SNR based on the signal magnitude and rms noise.

From this we can find the optical power per square root of bandwidth and hence the NEP from Eq. (5.12.5) as

$$\text{NEP} = \frac{P_1}{B^{1/2}} = \frac{1}{R}\left[2e(I_d + I_{ph})\right]^{1/2} \tag{5.12.7}$$

Noise equivalent power

It is clear that if we put $B = 1\,\text{Hz}$, we obtain numerically NEP $= P_1$, that value of P_o which makes I_{ph} equal to the total noise current i_n. From Eq. (5.12.7), the units for NEP are W $\text{Hz}^{-1/2}$. It is clear from the definition of NEP that we need to lower the NEP for better photodetection.

We can easily extend the NEP definition to photoconductive (PC) detectors as well. In this case, the noise added by the detector is the thermal noise of its resistance and that from the bias or load resistor. The NEP definition stays the same as in Eq. (5.12.5), though Eq. (5.12.7) would not be applicable as we need to consider the thermal noise generated by the photoconductor and its load resistor.

The **detectivity** D is defined as the reciprocal of noise equivalent power (NEP), $D = 1/\text{NEP}$; it is a measure of detector's sensitivity taking into account various noise contributions. However, for a proper comparison of detectors, we should consider all detectors to have the same photosensitive area. If A is the photosensitive area receiving the radiation, then **specific detectivity** D^* is defined by

$$D^* = \frac{A^{1/2}}{\text{NEP}} \tag{5.12.8}$$

Specific detectivity

The units are m $\text{Hz}^{1/2}\,\text{W}^{-1}$. However, cm $\text{Hz}^{1/2}\,\text{W}^{-1}$ is more common, and is called Jones. Typically, the detectivity increases as the temperature of the detector is lowered due to the lowering of the dark current. Table 5.5 lists typical NEP and D^* values for a range of photodetectors. Photoconductive detectors such as PbS, PbSe, and InSb are based on measuring their photoconductivity, that is, the change in the resistance upon illumination. They are used for the detection of radiation at long wavelengths, *e.g.,* longer than $\sim 2\,\mu\text{m}$. Notice also that these photoconductive detectors have lower detectivity than wider bandgap photodiodes used at shorter wavelengths, and have to be cooled.

According to Eq. (5.12.7), the NEP should depend on the dark current as $I_d^{1/2}$. Figure 5.32 shows a plot of NEP vs. I_d on a log-log scale for a collection of commercial Si and InGaAs *pin*, Ge *pn* and GaAsP Schottky junction photodiodes. NEP does indeed increase with I_d, very roughly as $I_d^{1/2}$ for a given class of photodiodes. The InGaAs *pin* photodiode case covers the widest range of dark currents, and involves room temperature, $-10°\text{C}$ and $-20°\text{C}$ values.

TABLE 5.5 Typical noise characteristics of a few selected commercial photodetectors

Photodiode	GaP Schottky	Si *pin*	Ge *pin*	InGaAs *pin*	PbS (PC) −10°C	PbSe (PC) −10°C	InSb (PC) −10°C
λ_{peak} (μm)	0.44	0.96	1.5	1.55	2.4	4.1	5.5
I_d or R_d	10 pA	0.4 nA	3 μA	5 nA	0.1−1 MΩ	0.1−1 MΩ	1−10 kΩ
NEP (W $\text{Hz}^{-1/2}$)	5.4×10^{-15}	1.6×10^{-14}	1×10^{-12}	4×10^{-14}			–
D^* (cm $\text{Hz}^{1/2}\,\text{W}^{-1}$)	1×10^{13}	1×10^{12}	1×10^{11}	5×10^{12}	1×10^{9}	5×10^{9}	1×10^{9}

Notes: PC means a photoconductive detector, whose photoconductivity is used to detect light. For PC detectors, what is important is the dark resistance R_d, which depends on the temperature.

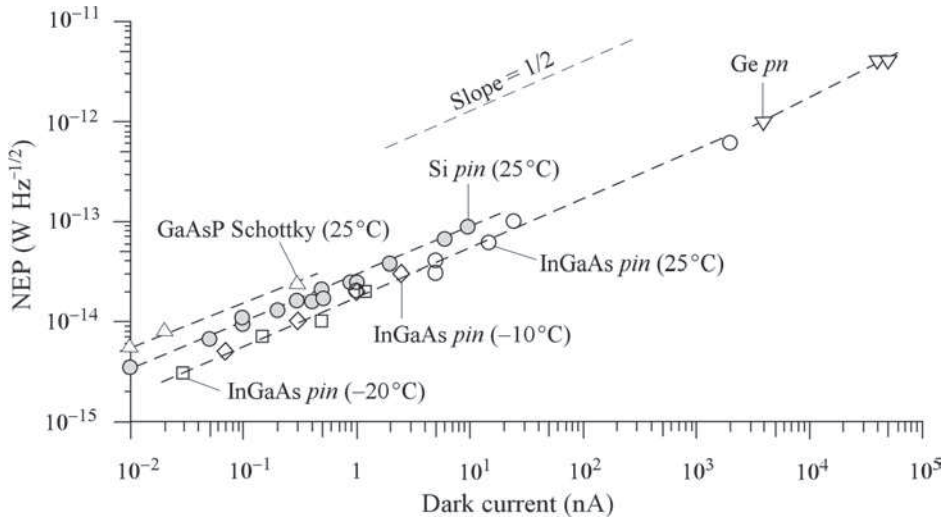


FIGURE 5.32 The dependence of NEP (W Hz$^{-1/2}$) on the photodetector dark current I_d for Si and InGaAs *pin*, Ge *pn* junction, and GaP Schottky photodiodes. Dashed lines indicate observed trends. Filled circle, Si *pin*; open circle, InGaAs *pin* at 25°C, open diamond at −10°C, open square at −20°C; inverted triangle, Ge *pn*; triangle, GaAsP Schottky. (*Source:* Data extracted from the datasheets of 35 commercial photodiodes.)

I_d decreases with decreasing temperature, and so does NEP, roughly along the NEP–$I_d^{1/2}$ line. In fact, at −20°C, InGaAs *pin* photodiodes NEP is quite low, ~3 × 10^{-13} W Hz$^{-1/2}$, which is the reason that it is cooled to achieve higher detectability.

EXAMPLE 5.12.1 NEP of a Si *pin* photodiode

A Si *pin* photodiode has a quoted NEP of 1 × 10^{-13} W Hz$^{-1/2}$. What is the optical signal power it needs for a signal to noise ratio (SNR) of 1 if the bandwidth of operation is 1 GHz?

Solution

By definition, NEP is that optical power per square root of bandwidth which generates a photocurrent equal to the noise current in the detector.

$$\text{NEP} = P_1/B^{1/2}$$

Thus,

$$P_1 = \text{NEP}B^{1/2} = (10^{-13}\,\text{W Hz}^{-1/2})(10^9\,\text{Hz})^{1/2} = 3.16 \times 10^{-9}\,\text{W} \quad \text{or} \quad 3.16\,\text{nW}$$

EXAMPLE 5.12.2 Noise of an ideal photodetector

Consider an ideal photodiode with $\eta_e = 1$ (QE = 100%) and no dark current, $I_d = 0$. Show that the minimum optical power required for a signal to noise ratio (SNR) of 1 is

$$P_1 = \frac{2hc}{\lambda} B \tag{5.12.9}$$

Calculate the minimum optical power for a SNR = 1 for an ideal photodetector operating at 1300 nm with a bandwidth of 1 GHz. What is the corresponding photocurrent?

Solution

We need the incident optical power P_1 that makes the photocurrent I_{ph} equal to the noise current i_n, so that SNR = 1. The photocurrent (signal) is equal to the noise current when

$$I_{\text{ph}} = i_n = \left[2e(I_d + I_{ph})B\right]^{1/2} = (2eI_{ph}B)^{1/2}$$

since $I_d = 0$. Solving the above, $I_{ph} = 2eB$

From Eqs. (5.4.3) and (5.4.4), the photocurrent I_{ph} and the incident optical power P_1 are related by

$$I_{ph} = \frac{\eta_e e P_1 \lambda}{hc} = 2eB$$

Thus,

$$P_1 = \frac{2hc}{\eta_e \lambda} B$$

For an ideal photodetector, $\eta_e = 1$, which leads to Eq. (5.12.9). We note that for a bandwidth of 1 Hz, NEP is numerically equal to P_1 or NEP $= 2hc/\lambda$.

For an ideal photodetector operating at 1.3 μm and at 1 GHz

$$\begin{aligned} P_1 &= 2hcB/\eta_e\lambda = 2(6.63 \times 10^{-34}\,\text{J s})(3 \times 10^8\,\text{m s}^{-1})(10^9\,\text{Hz})/(1)(1.3 \times 10^{-6}\,\text{m}) \\ &= 3.1 \times 10^{-10}\,\text{W} \quad \text{or} \quad 0.31\,\text{nW} \end{aligned}$$

This is the minimum signal for an SNR = 1. The noise current is due to quantum noise. The corresponding photocurrent is

$$I_{ph} = 2eB = 2(1.6 \times 10^{-19}\,\text{C})(10^9\,\text{Hz}) = 3.2 \times 10^{-10}\,\text{A} \quad \text{or} \quad 0.32\,\text{nA}$$

Alternatively, we can calculate I_{ph} from $I_{ph} = \eta_e e P_1 \lambda / hc$ with $\eta_e = 1$.

EXAMPLE 5.12.3 SNR of a receiver

Consider an InGaAs *pin* photodiode used in a receiver circuit as in Figure 5.31 with a load resistor of 10 kΩ. The photodiode has a dark current of 2 nA. The bandwidth of the photodiode and the amplifier together is 1 MHz. Assuming that the amplifier is noiseless, calculate the SNR when the incident optical power generates a mean photocurrent of 5 nA (corresponding to an incident optical power of about 6 nW since $\boldsymbol{R}$ is about 0.8–0.9 nA/nW at the peak wavelength of 1550 nm).

Solution

The noise generated comes from the photodetector as shot noise and from R_L as thermal noise. The mean thermal noise power in the load resistor R_L is $4k_BTB$. If I_{ph} is the photocurrent and i_n is the shot noise in the photodetector then

$$\text{SNR} = \frac{\text{Signal power}}{\text{Noise power}} = \frac{I_{ph}^2 R_L}{i_n^2 R_L + 4k_BTB} = \frac{I_{ph}^2}{\left[2e(I_d + I_{ph})B\right] + 4k_BTB/R_L}$$

The term $4k_BTB/R_L$ in the denominator represents the mean square of the thermal noise current in the resistor. We can evaluate the magnitude of each noise current by substituting, $I_{ph} = 5$ nA, $I_d = 2$ nA, $B = 1$ MHz, $R_L = 10^4\,\Omega$, $T = 300$ K.

$$\text{Shot noise current from the detector} = \left[2e(I_d + I_{ph})B\right]^{1/2} = 0.047 \text{ nA}$$

$$\text{Thermal noise current from } R_L = \left(\frac{4k_BTB}{R_L}\right)^{1/2} = 1.29 \text{ nA}$$

Thus, the noise contribution from R_L is greater than that from the photodiode. The SNR is

$$\text{SNR} = \frac{(5 \times 10^{-9} \text{ A})^2}{(0.047 \times 10^{-9} \text{ A})^2 + (1.29 \times 10^{-9} \text{ A})^2} = 15.0$$

Generally SNR is quoted in decibels. We need $10\log(\text{SNR})$, or $10\log(15.0)$, *i.e.,* 11.8 dB. Clearly, the load resistance has a dramatic effect on the overall noise performance.

B. Avalanche Noise in the APD

In the avalanche photodiode, both photogenerated and thermally generated carriers entering the avalanche zone are multiplied. The shot noise associated with these carriers are also multiplied. If I_{do} and I_{pho} are the unmultiplied ($M = 1$) dark current and photocurrent (primary photocurrent), respectively, in the APD then the shot noise current (as an rms value) in the APD should be

$$i_{n\text{-APD}} = M\left[2e(I_{do} + I_{pho})B\right]^{1/2} = \left[2e(I_{do} + I_{pho})M^2B\right]^{1/2} \quad \textbf{(5.12.10)}$$

The APDs exhibit excess avalanche noise that is above this multiplied shot noise of the photocurrent and dark current. This excess noise is due to the randomness of the impact ionization process in the multiplication region. Some carriers travel far and some short distances within this zone before they cause impact ionization. Furthermore, the impact ionization does not occur uniformly over the multiplication region but is more frequent in the highest field zone. Thus, the multiplication M fluctuates about a mean value. The result of the statistics of impact ionization is an *excess noise* contribution, called **avalanche noise**, to the multiplied shot noise. The noise current in an APD is then given by

APD noise current

$$i_{n\text{-APD}} = \left[2e(I_{do} + I_{pho})M^2FB\right]^{1/2} \quad \textbf{(5.12.11)}$$

where F is called the **excess noise factor** and is a function of M and the impact ionization probabilities (called coefficients). Generally, F is approximated by the relationship $F \approx M^x$ where x is an index that depends on the semiconductor, the APD structure, and the type of carrier that initiates the avalanche (electron or hole). For Si APDs x is 0.3–0.5, whereas for Ge and III–V (such as InGaAs) alloys it is 0.7–1.

EXAMPLE 5.12.4 Noise in an APD

Consider an InGaAs APD with $x \approx 0.7$ which is biased to operate at $M = 10$. The unmultiplied dark current is 10 nA and bandwidth is 700 MHz.

(a) What is the APD noise current per square root of bandwidth?
(b) What is the APD noise current for a bandwidth of 700 MHz?
(c) If the responsivity (at $M = 1$) is 0.8 A W^{-1} what is the minimum optical power for a SNR of 10?

Solution

(a) In the absence of any photocurrent, the noise in the APD comes from the dark current. If the unmultiplied dark current is I_{do} then using $F = M^x$ in Eq. (5.12.11), the noise current (rms) is

$$i_{n\text{-dark}} = \left[2eI_{do}M^{2+x}B\right]^{1/2} \tag{5.12.12}$$

Thus,

$$\frac{i_{n\text{-dark}}}{\sqrt{\text{B}}} = \sqrt{2eI_{do}M^{2+x}} = \sqrt{2(1.6 \times 10^{-19}\,\text{C})(10 \times 10^{-9}\,\text{A})(10)^{2+0.7}}$$
$$= 1.27 \times 10^{-12}\,\text{A Hz}^{-1/2} \quad \text{or} \quad 1.27\,\text{pA Hz}^{-1/2}$$

(b) In a bandwidth B of 700 MHz, the noise current is

$$i_{n\text{-dark}} = (700 \times 10^6\,\text{Hz})^{1/2}(1.27\,\text{pA Hz}^{-1/2}) = 3.35 \times 10^{-8}\,\text{A} \quad \text{or} \quad 33.5\,\text{nA}$$

(c) The SNR with a primary photocurrent I_{pho} in the APD is

$$\text{SNR} = \frac{\text{Signal power}}{\text{Noise power}} = \frac{M^2I_{pho}^2}{\left[2e(I_{do} + I_{pho})M^{2+x}B\right]} \tag{5.12.13}$$ *SNR APD*

Rearranging to obtain I_{pho} we get

$$(M^2)I_{pho}^2 - \left[2eM^{2+x}B(\text{SNR})\right]I_{pho} - \left[2eM^{2+x}B(\text{SNR})I_{do}\right] = 0$$

This is a quadratic equation in I_{pho} with defined coefficients since M, x, B, I_{do}, and SNR are given. Solving this quadratic with a SNR $= 10$ for I_{pho} we find

$$I_{pho} \approx 1.76 \times 10^{-8}\,\text{A} \quad \text{or} \quad 17.6\,\text{nA}$$

While it may seem odd that I_{pho} is less than the dark noise current (33.5 nA) itself, the actual photocurrent I_{ph} is 176 nA, since it is multiplied by M. Further, the total noise current $i_{n\text{-APD}} = \left[2e(I_{do} + I_{pho})M^{2+x}B\right]^{1/2}$ is 55.7 nA, so that one can easily check that SNR $= I_{ph}^2/i_{n\text{-APD}}^2$ is indeed 10.

By the definition of responsivity, $\mathcal{R} = I_{pho}/P_o$, we find

$$P_o = I_{pho}/\mathcal{R} = (1.76 \times 10^{-8}\,\text{A})/(0.8\,\text{A W}^{-1}) = 2.2 \times 10^{-8}\,\text{W} \quad \text{or} \quad 22\,\text{nW}$$

5.13 IMAGE SENSORS

A. Basic Principles

An image sensor, as illustrated in Figure 5.33 (a), is an integrated circuit chip made up of an array of photosensitive elements, that is able to capture an image and provide an output in the form of an electrical signal such as current, charge, or voltage. The output is usually put through a multiplexer and an analog-to-digital converter to obtain a digital form of the image as shown in Figure 5.33 (b). The sensor usually consists of an array of elements, called **pixels** (*pixel* from "picture element"), in N rows and M columns, where each element or pixel has a detector and is able to provide an electrical signal that is proportional to the light intensity received at this pixel. A lens forms an image of the object on the image sensor, and the light intensity $I(X,Y)$ at each point of this image (on the sensor) excites the corresponding pixel at the X,Y location. The image point X,Y excites the pixel at that location, that is, the pixel in row i and column j, and the

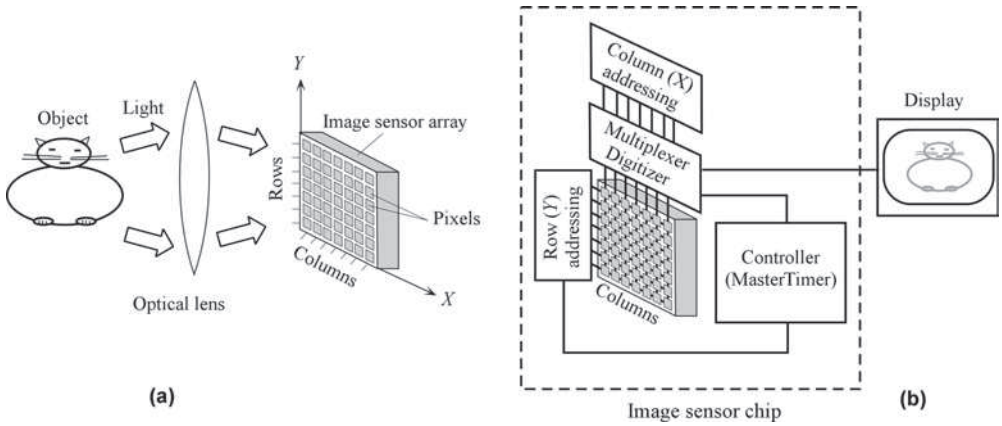


FIGURE 5.33 (a) The basic image-sensing operation using an array of photosensitive pixels. (b) The image sensor chip that incorporates the auxiliary electronics that run the sensor array (CMOS technology).

pixel (i,j) has an electrical signal (*e.g.*, charge) proportional to $I(X,Y)$. Thus, each pixel (i,j) carries a piece of the image information, a pixel of it.

The imaging array shown in Figure 5.33 (b) is addressed from a master timer to read out the signals (charges) at the pixels and hence provide an electronically recorded image. There are two major solid state technologies: **charge-coupled device** (CCD) and **complementary metal oxide semiconductor** (CMOS) sensors, each with its own advantages and drawbacks as discussed in Sections B and C below. CMOS refers to the technology that is used to fabricate the sensor. In the CCD technology, the incident light at a pixel generates charge that is stored in the CCD pixel (called a well) at this location (and discussed below). This charge (signal) is read out by a sequential clocking mechanism, a characteristic of the CCD operation, to shift the charges to an eventual shift register for read-out. The CMOS sensor is more like the image sensor chip shown in Figure 5.33 (b). Each pixel has MOS transistors and a photodiode. The electrical signals at pixels are read out line by line. A row i is addressed by the controller, and all the pixels in this row provide their signals to the columns (data lines) $j = 1$ to M. The parallel data are multiplexed and digitized to provide a digital image. The image sensor quantizes the image because it breaks into $N \times M$ number of pixels. The pixel size imposes a resolution limit because the image is sampled by the pixels, which are of finite size.

Color rendering with image sensors can be done in three ways. The most common is to use red, green, and blue filters to separate out the colors at a point onto three different adjacent pixels,

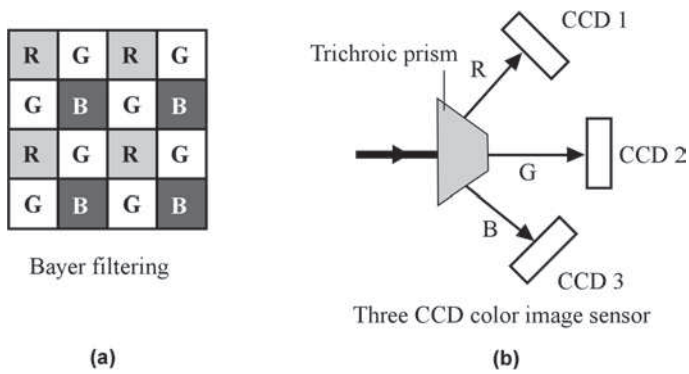


FIGURE 5.34 (a) Color imaging by using two different techniques. Bayer filtering uses red (R), green (G), and blue (B) filters on three pixels for capturing the R, G, and B information. (b) Three CCD sensor uses a trichroic prism to separate the image colors into red, green, and blue, and uses three CCD chips for each.

such as in the **Bayer color image sensor** shown in Figure 5.34 (a). In **three CCD imaging**, as shown in Figure 5.34 (b), a trichroic prism assembly is used to separate the light from the object into its red, green, and blue components, and three different CCDs are used to detect each component. A third technique uses different absorption depths for red, green, and blue light in silicon. The red, green, and blue light have different penetration depths into silicon so that, in principle, we can use photodiodes at different depths to distinguish between the colors. The Bayer filtering technique is the most common among color image sensors.

B. Active Matrix Array and CMOS Image Sensors

Many image sensors, as well as various flat panel TV technologies, make use of active matrix arrays (AMAs). An **active matrix array** is a two-dimensional array of pixels in which each pixel has a thin film transistor (a field effects transistor that has been fabricated by using thin film technologies) that can be externally addressed to read a signal from a sensor located at that pixel. Depending on the application, an AMA can have few pixels or millions of pixels. The TFT AMA technology was pioneered by Peter Brody using CdSe TFTs in early 1970s. A simplified diagram is shown in Figure 5.35 (a) for the array and in (b) for each pixel. Each pixel has a photodiode and a capacitor C_{px}. When a pixel receives light from a particular point on the illuminated object, the photodiode generates a current I_{signal} which charges the pixel capacitor C_{px}. The signal is the charge Q_{signal} stored on C_{px}. The array of pixels therefore has the image stored as charges on pixel capacitances. All we have to do is read the charges out.

Each pixel is identical with its TFT gate connected to a particular address line and the source to a particular data line as shown in Figure 5.35 (b). When row *a* is addressed, all pixels and hence all signals on that row are read out, that is, the data are read out onto the data lines as *parallel data*, which can be converted into serial data. We can scan the whole image row by row (line by line) by starting from the top row and sequentially activating one row after another. Thus, each time we activate and access a row, we extract the signals (Q_{signal} on each C_{px}) on that row as parallel data.

A **CMOS image sensor** is basically an AMA in which each pixel has a photodiode or a photogate and one or more CMOS transistors to read and amplify the electrical signal that has been

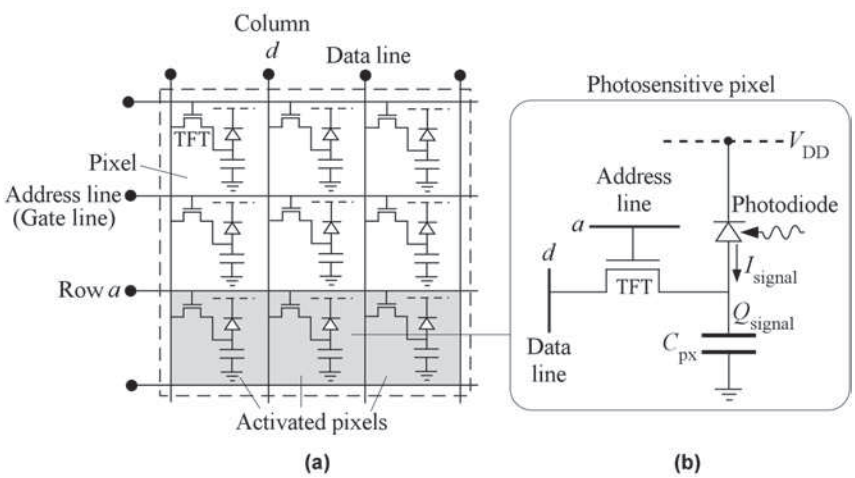


FIGURE 5.35 (a) An active matrix array (AMA). (b) A basic photosensitive pixel structure for detecting the photons arriving at the pixel defined by row *a* and column *d*.

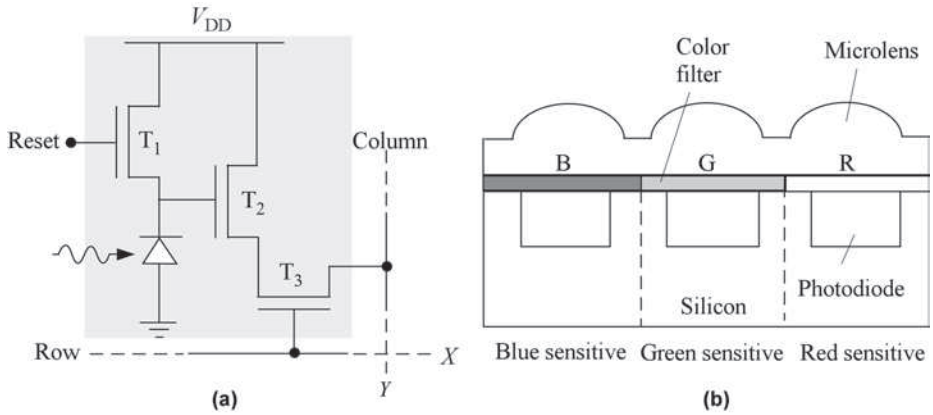
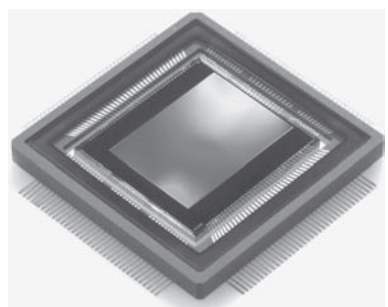


FIGURE 5.36 (a) The pixel architecture in a CMOS image sensor. (b) A cross-section(s) of a CMOS imager with microlenses and color filters (B = blue, G = green, R = red) for color imaging.

generated by the light incident on this pixel. The pixel read-out operation is essentially the same as that in the AMA. The pixel architecture, however, can change from application to application.

CMOS is a well-established silicon fabrication technology, which means that CMOS image sensors can be fabricated relatively inexpensively. If the pixel is active with gain, then the imager is called an **active pixel image sensor**, or **active pixel sensor**. In a **passive pixel sensor** there is only a switching transistor to read the charge out when the transistor switch is addressed. Nearly all modern CMOS imagers are active pixel sensors. One possible active pixel architecture using a photodiode is shown in Figure 5.36 (a). T_1 is the reset transistor, T_2 works as a source follower (*i.e.*, a buffer), and T_3 is a pixel switch transistor. With T_1 off, the photocurrent charges the self-capacitance of the photodiode to a certain voltage. When the row *X* receives a signal, T_3 is turned on, and the signal voltage on the photodiode is transferred through the buffer T_2 to the column *Y*. The pixel is then reset, and T_1 connects the photodiode to V_{DD} to clear the accumulated charge.



Four Megapixel CMOS image sensor. *(Courtesy of Teledyne-DALSA.)*

CMOS imagers have a number of distinct advantages. Standard and well-established CMOS fabrication, and hence lower cost, mean on-chip integration such as a **camera on a chip**. A camera on a chip has a microlens at each pixel—a red, green, or blue filter at each pixel for color imaging—as illustrated in Figure 5.36 (b), analog signal processing after the read-out of the image, and analog-to-digital conversion. Active pixel architecture can include pixel amplification as well to improve the overall efficiency. CMOS low power consumption leads to a longer battery life.

C. Charge-Coupled Devices

Charge-coupled devices (CCDs) are commonly used as image sensors in professional and consumer television cameras and camcorders, and as image sensors in digital still cameras. However, the term CCD in general does not imply an image sensor but a chip that is able to store and transfer signals in the form of charge. The CCD chip is an integrated circuit made from crystalline silicon and has a large number of pixels (each is a detector element); for example, a 2.5 cm × 2.5 cm CCD chip may have 1024 × 1024 or 2048 × 2048 pixels on its surface. The basic pixel structure is a MOS (metal-oxide-semiconductor) or a MIS (metal-insulator-semiconductor) device as shown in Figure 5.37. Notice that the structure is based on a *p*-type Si, an oxide layer, and a metal electrode, which is usually transparent. There is a depletion region inside the *p*-type semiconductor. The EHPs are generated inside the depletion region by illumination either from the top surface or from the backside as shown in Figure 5.37. In back-thinned CCD, light enters the depletion region not from the gate side but from the "substrate" side, which has been thinned to allow the light to pass through.

When a positive voltage $+V$ is applied to the gate V_G in Figure 5.37 ($V_G = +V$), the photogenerated electrons in the depletion region are collected in a layer near the interface as shown in Figure 5.37. (With no gate voltage, photogenerated electrons and holes disappear by recombination.) These electrons are trapped inside a potential energy well, introduced by $+V$ on the gate. Their total charge is proportional to the total light exposure. This charge constitutes the electrical signal. The objective is to read all these charges stored at the illuminated pixels.

In a **three-phase CCD** read-out, there are three line voltages V_1, V_2, V_3 to which the gates are connected in an alternating fashion: G_1 to V_1, G_2 to V_2, and G_3 to V_3, G_4 to V_1 again and so on as shown in Figure 5.38. V_1, V_2, and V_3 are appropriately clocked to shift the charges from pixel to pixel to a register located at the end of the chip. If initially (time $t = t_1$) $V_1 = +V$ and $V_2 = V_3 = 0$, then photogenerated charges will be stored under G_1, G_4, *etc.* as Q_a, Q_b, *etc.* Later ($t = t_2$) we can make $V_1 = +V$, $V_2 = +V$, and $V_3 = 0$. The charge Q_a is shared between the wells under G_1 and G_2. Even later ($t = t_3$), we can bring V_1 down to zero, the charge Q_a must go into the available potential well, which is under G_2. Thus, by toggling gate voltages, Q_a has been shifted from G_1 to G_2, and similarly Q_b from G_4 to G_5 and so on. The charges are therefore clocked progressively along the gates, from pixel to pixel, left to right, until they reach the end of the array where there is a register. The CCD read-out therefore functions like a *shift register* in that clock pulses shift the information along the chain; they are often termed *CCD shift registers*.

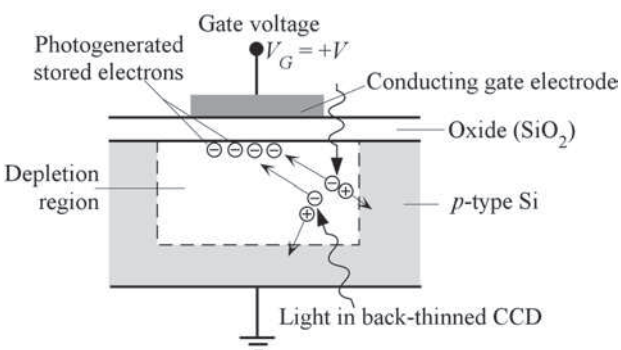
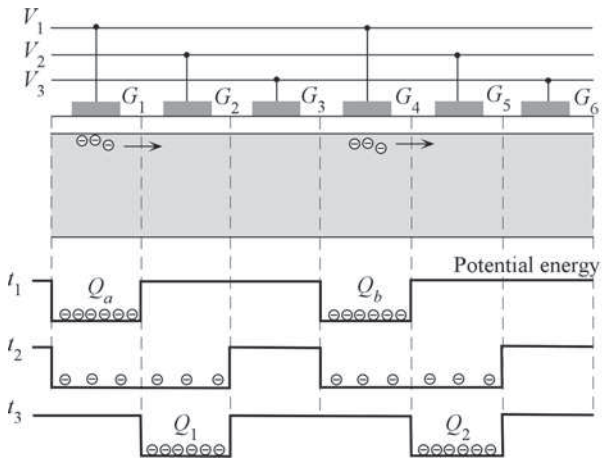


FIGURE 5.37 One element of a CCD imaging sensor, which is a MOS (metal-oxide-semiconductor) device. The absorbed photon actually creates an electron and a hole. The hole simply drifts to the *p*-side.

Various CCD chips. *(Courtesy of Teledyne-DALSA.)*



Time	V_1	V_2	V_3
t_1	$+V$	0	0
t_2	$+V$	$+V$	0
t_3	0	$+V$	0

FIGURE 5.38 Transfer of charge from one well to another by clocking the gate voltages. The table shows the gate voltage sequences in a three-phase CCD.

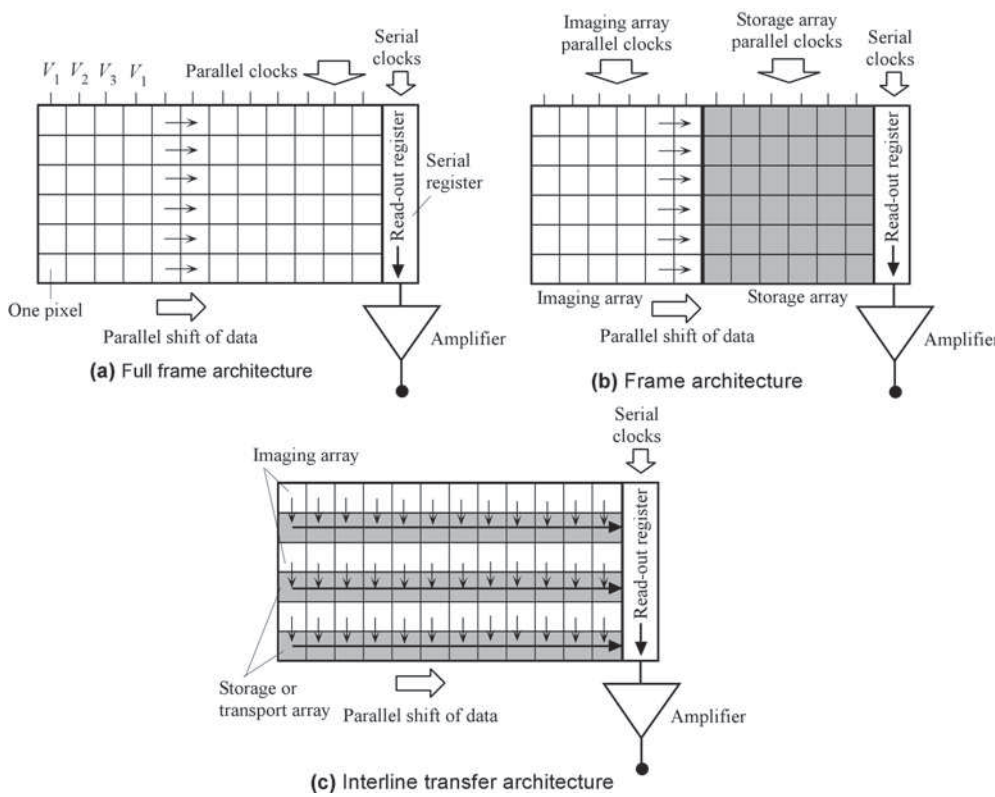


FIGURE 5.39 Three CCD architectures: (a) Full frame architecture. (b) Frame architecture. (c) Interline transfer architecture.

There are three main types of CCD read-out techniques and hence CCD architectures called full frame, frame, and interline transfer architectures. In a **full frame architecture,** the sensor consists of parallel CCD shift registers, a serial CCD shift register as a *read-out register,* and an amplifier as shown in Figure 5.39 (a). The columns of data are fed into the serial register starting soon after the projection of the image onto the array, as the clock signals start transferring the charges pixel to pixel toward the read-out register. A short build-up or exposure time is necessary to build up the charges in the imaging arrays. The process continues until all the columns of data have been transferred. This architecture has the highest resolution and is the densest.

The **frame architecture**, as shown in figure 5.39 (b), is similar to the full frame except that half the array is used as data storage to increase the speed. The storage array is blocked from light. The idea is to shift the data away from the image sensors quickly so that they can capture the next image soon; these have higher frame rates at the expense of resolution.

The **interline transfer architecture** uses every other row of the CCD array for *storage* and *transport* right next to the imaging array as illustrated in figure 5.39 (c). The transport array is blocked from light, and the data from the imaging array are transferred in parallel to the transport array after the image has been built-up in the imaging array. The data on the transport array are shifted, just as in the normal CCD.

Additional Topics

5.14 PHOTOVOLTAIC DEVICES: SOLAR CELLS

A. Basic Principles

A **photovoltaic device** or **a solar cell** converts the incident radiation energy into electrical energy.[19] Incident photons are absorbed to photogenerate charge carriers, which then pass through an external load to do electrical work. Photovoltaic devices may be metal–semiconductor Schottky junctions, *pn* junctions, or *pin* devices. For example, a crystalline Si *pn* junction solar cell may have a thin *n*-type semiconductor layer on a thick *p*-type substrate as shown in Figure 5.40 (a). The electrodes attached to the *n*-side must allow the light to enter the device and at the same time result in a small series resistance. They are deposited on to *n*-side to form an array of **finger electrodes** on the surface as illustrated in Figure 5.40 (a). There is a thin film antireflection (AR) coating (one or two layers) on the surface outside the electrodes. The AR coating significantly reduces the reflectance of the bare semiconductor–air interface.

The *n*-side of the *pn* junction is very narrow to allow most of the photons to be absorbed within the depletion region (W) and within the neutral *p*-side. The photogeneration of electron–hole pairs occurs mainly in these regions. In Si, the electron diffusion length L_e is longer than the hole diffusion length L_h, which is the reason for having the *p*-layer as the main photon absorbing layer. EHP photogeneration then occurs in a bigger volume defined by W and L_e in the *p*-side as indicated in Figure 5.40 (a). Remember that only those electrons in the *p*-side within L_e to the SCL can contribute the photocurrent. EHPs photogenerated in W in Figure 5.40 (a) are

[19] Although the terms *photovoltaic device* and *solar cell* are often used interchangeably, solar cells would convert solar radiation into electrical work, whereas a photovoltaic device could also be used as a photodetector, operating in the photovoltaic mode; their designs would be somewhat different.

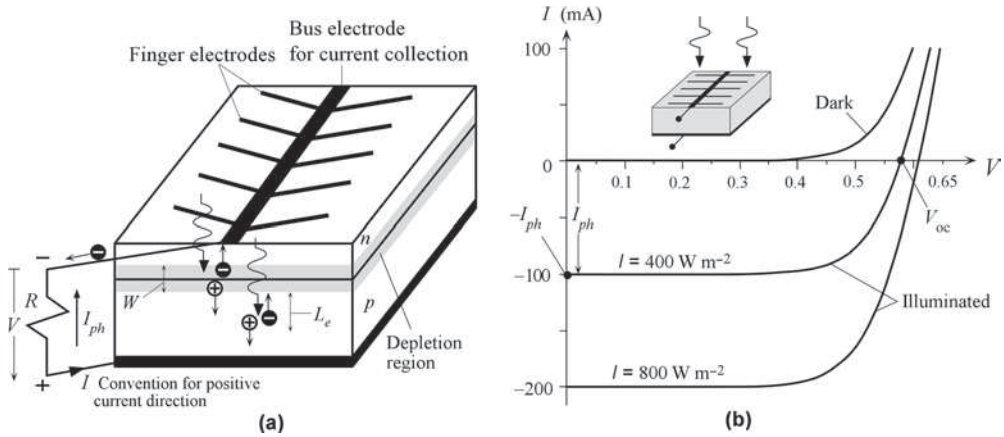


FIGURE 5.40 (a) Typical *pn* junction solar cell. Finger electrodes on the surface of a solar cell reduce the series resistance. (b) *I–V* characteristics in the dark and under illumination at intensities corresponding to 400 W m^{-2} and 800 W m^{-2}.

immediately separated by the built-in field, which drifts them apart. The electrons drift toward the *n*-side and holes toward the *p*-side, which generates a photocurrent. The excess electrons reach the neutral *n*-side, then drift around the external circuit, do work, and reach the *p*-side to recombine with the excess holes on this side. As a result, there will be a continuous external photocurrent during illumination. If the load is a short circuit, then the magnitude of the external current is simply equal to the photocurrent I_{ph} generated by the incident radiation ($I = -I_{ph}$). It is important to realize that without the internal field, it is not possible to drift apart the photogenerated EHPs and hence cause a current flow in the external circuit.

The *I–V* characteristics of a typical Si solar cell are shown in Figure 5.40 (b). The dark *I–V* is the usual forward biased *pn* junction diode equation

pn junction diode current

$$I_{\text{diode}} = I_o\left[\exp(eV/\eta k_B T) - 1\right] \tag{5.14.1}$$

where η is the diode ideality factor (not to be confused with the solar cell efficiency) which is between 1 and 2. Under illumination, the *I–V* dark characteristics are shifted down by an amount that is equal to the photocurrent I_{ph}. The photocurrent I_{ph} is proportional to the photogeneration rate and hence to the incident light intensity *I*, *i.e.,*

Photocurrent and light intensity

$$I_{ph} = KI \tag{5.14.2}$$

where *K* is a device specific constant. As can be seen from Figure 5.40 (b), doubling the light intensity has doubled I_{ph}.

The positive *I* and *V* convention and the short circuit behavior of an illuminated *pn* junction are shown in Figure 5.41 (a) and (b). The short circuit current I_{sc} is $-I_{ph}$, and is an important solar cell quantity. If the solar cell is in open circuit, there is no external current ($I = 0$) but there is a voltage, called the **open circuit voltage**, V_{oc}, across the device as in Figure 5.41 (c). This voltage corresponds to the point where the illuminated *I–V* characteristic cuts the *V*-axis in Figure 5.40 (b). The photogenerated electrons arriving on the *n*-side cannot flow in the external circuit and accumulate in the *n*-side; similarly photogenerated holes accumulate in the *p*-side. We should clearly see a net positive voltage on the *p*-side with respect to the *n*-side. The reality

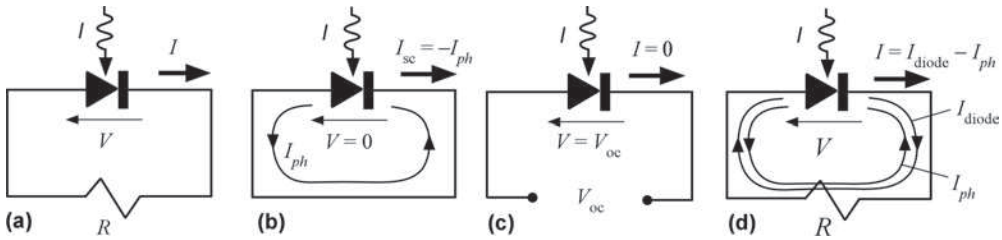


FIGURE 5.41 (a) The solar cell connected to an external load R and the convention for the definitions of positive voltage and positive current. (b) The solar cell in short circuit. The magnitude of the external current I_{sc} in the photocurrent I_{ph}. (c) Under open circuit conditions there is a voltage V_{oc} at the output terminals. (d) The solar cell driving an external load R. There is a voltage V and current I in the circuit.

is a bit more subtle. The extra electrons on the n-side neutralize some of the positive donor ions in the SCL on the n-side. Extra holes neutralize some of negative acceptors in the SCL on the p-side. The internal field becomes decreased, and hence the built-in voltage V_o becomes reduced. It is the reduction in the built-in voltage that appears as V_{oc} at the external terminals.

Suppose that there is an external load R, as shown in Figures 5.40 (a) and 5.41 (d). The external current (no longer simply I_{ph}) will flow through R and generate a voltage V across it. IV is the electric power dissipated in the external load. It should be apparent that the voltage V developed across R now *forward biases* the pn junction, generates a diode current I_{diode} in the normal way, as given by Eq. (5.14.1), and flows in the opposite direction to I_{ph}. The net current, as apparent from Figure 5.41 (d) is

$$I = -I_{ph} + I_{\text{diode}} = -I_{ph} + I_o\left[\exp(eV/\eta k_B T) - 1\right] \qquad \textbf{(5.14.3)}$$

Solar cell I–V

which represents the I–V characteristics of a solar cell. The light intensity increases I_{ph} along Eq. (5.14.2), and hence shifts the normal diode characteristics down as shown in Figure 5.40 (b).

An experimental solar cell aircraft called Helios flying over the coast of Hawaii. *(Courtesy of NASA Dryden Research Center.)*

B. Operating Current and Voltage and Fill Factor

Equation (5.14.3) gives only the I–V characteristics of the solar cell. When the solar cell is connected to a load as in Figure 5.42 (a), the load has the same voltage as the solar cell and carries the same current. But the current I through R is now in the opposite direction to the convention that current flows from high to low potential. Thus, as shown in Figure 5.42 (a)

$$I = -V/R \qquad \textbf{(5.14.4)}$$

The load line

The actual current I' and voltage V' in the circuit must satisfy both the I–V characteristics of the solar cell, Eq. (5.14.3), and that of the load, Eq. (5.14.4). We can find I' and V' by solving these two equations simultaneously or using a **load line construction**. The I–V characteristics

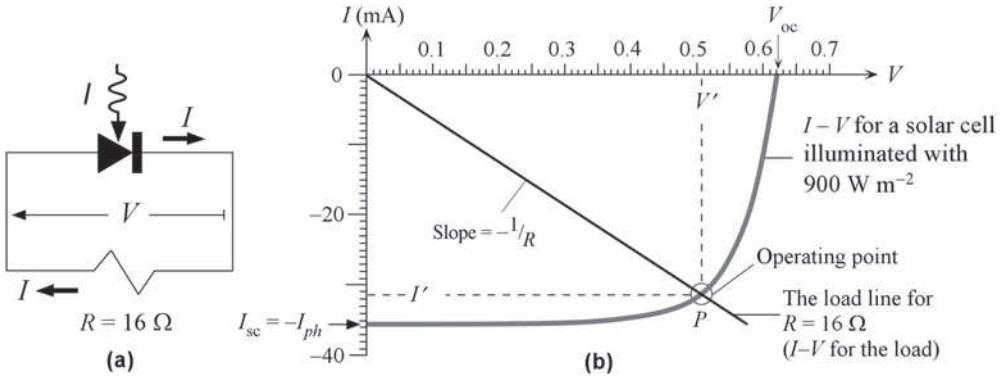


FIGURE 5.42 (a) A solar cell driving a load R and the definitions of positive current I and voltage V. (b) The load line construction for finding the operating point when a load $R_L = 15\,\Omega$ is connected across the solar cell.

of the load in Eq. (5.14.4) is a straight line with a negative slope $-1/R$. This is called the **load line** and is shown in Figure 5.42 (b) along with the I–V characteristics of the solar cell under a given intensity of illumination. The load line cuts the solar cell characteristic at P where the load and the solar cell have the same current and voltage, I' and V'. Point P therefore satisfies both Eqs. (5.14.3) and (5.14.4) and thus represents the **operating point of the circuit**.

The **power delivered** to the load is $P_{\text{out}} = I'V'$, which is the area of the rectangle bound by I- and V-axes and the dashed lines shown in Figure 5.42 (b). Maximum power is delivered to the load when this rectangular area is maximized (by changing R or the intensity of illumination), when $I' = I_m$ and $V' = V_m$. Since the maximum possible current is I_{ph} and the maximum possible voltage is V_{oc}, $I_{ph}V_{\text{oc}}$ represents the desirable goal in power delivery for a given solar cell. It therefore makes sense to compare the maximum power output[20], I_mV_m, with I_{sc} V_{oc}. The **fill factor** (FF), which is a figure of merit for the solar cell, is defined as

Definition of fill factor

$$\text{FF} = \frac{I_mV_m}{I_{\text{sc}}V_{\text{oc}}} \tag{5.14.5}$$

The FF is a measure of the closeness of the solar cell I–V curve to the rectangular shape (the ideal shape). It is clearly advantageous to have FF as close to unity as possible, but the exponential *pn* junction properties prevent this. Typically FF values are in the range 70–85% and depend on the device structure. One can vary the operating point P in Figure 5.42 (b) to examine the effects of different load lines on the rectangular area $I'V'$. A load line that cuts the solar cell I–V characteristics around the "knee" point, as in Figure 5.42 (b), corresponds to nearly the maximum power I_m,V_m point.

C. Equivalent Circuit of a Solar Cell

Practical solar cells can deviate substantially from the ideal *pn* junction solar cell behavior illustrated in Figure 5.40 (b) due to a number of reasons. Consider an illuminated solar cell driving a load resistance R_L and assume that photogeneration takes place in the depletion region. The photogenerated electrons have to traverse a surface semiconductor region to reach the nearest finger electrode as shown in Figure 5.43 (a). All these electron paths in the n-layer surface region of finger electrodes introduce an **effective series resistance** R_s into the photovoltaic circuit. If the finger electrodes are thin, then the resistance of the electrodes themselves will further

[20] One should not be too concerned with the negative sign for I_mV_m inasmuch as the diode is generating electric power during illumination.

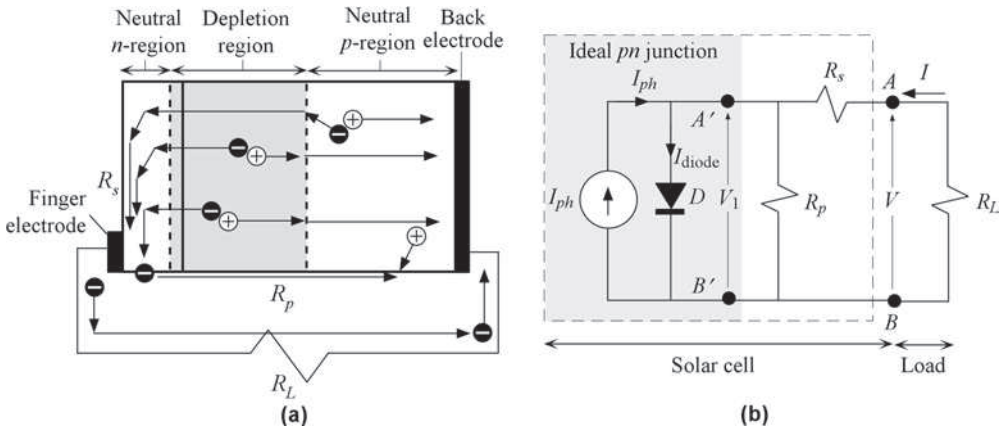


FIGURE 5.43 (a) Series and shunt resistances and various fates of photogenerated EHPs. (b) A simple equivalent circuit diagram for a solar cell. The current I into A, and V across A and B, follow the convention for positive current and voltage. A' and B' represent the ideal diode terminals.

increase R_s. There is also a series resistance due to the neutral p-region but this is generally small compared with the resistance of the electron paths to the finger electrodes.

Figure 5.43 (b) shows the equivalent circuit of a pn junction solar cell. The gray region is the ideal pn junction diode case and has the terminals A' and B'. The photogeneration process is represented by a constant current generator I_{ph}, which generates a current that is proportional to the light intensity. The flow of photogenerated carriers across the junction gives rise to a photovoltaic voltage difference V_1 across the junction, and this voltage leads to the normal diode current $I_{\text{diode}} = I_o\left[\exp(eV_1/\eta kT) - 1\right]$. This diode current I_{diode} is represented by an ideal pn junction diode D in the circuit as shown in Figure 5.43 (b). As apparent, I_{ph} and I_{diode} are in opposite directions (I_{ph} is "up" and I_{diode} is "down") so that, in open circuit, the photovoltaic voltage V_1 is such that I_{ph} and I_{diode} have the same magnitude and cancel each other. By convention, positive current I at the output terminal is normally taken to flow into the terminal, and is given by Eq. (5.14.3) (In reality, of course, the solar cell current is negative, which represents a current that is flowing out into the load.) The voltage V between A and B is the actual voltage measured, and available externally across the solar cell terminals. Notice that $V_1 = V - IR_s$ so that we can substitute for V_1 in terms of V in the diode current equation.

The equivalent circuit in Figure 5.43 (b) outside the ideal gray box contains the series resistance R_s in Figure 5.43 (a), which gives rise to a voltage drop, and therefore prevents the ideal photovoltaic voltage from developing at the output between A and B when a current is drawn. A small fraction of the photogenerated carriers can also flow through the crystal surfaces (edges of the device) or through *grain boundaries in polycrystalline devices* instead of flowing through the external load R_L as shown in Figure 5.43 (a). These effects that prevent photogenerated carriers from flowing in the external circuit can be represented by an effective **shunt** or **parallel resistance** R_p that diverts the photocurrent away from the load R_L. Typically R_p is less important than R_s in overall device behavior, unless the device is highly polycrystalline and the current component flowing through grain boundaries is not negligible. The current I flowing into A, and the voltage V across A and B (of A with respect to B) in Figure 5.43 (b) now represent the solar cell characteristics that the user measures.

The series resistance R_s can significantly deteriorate the solar cell performance as illustrated in Figure 5.44, where $R_s = 0$ is the best solar cell case. It is apparent that the available maximum output power decreases with increasing series resistance, which therefore reduces the

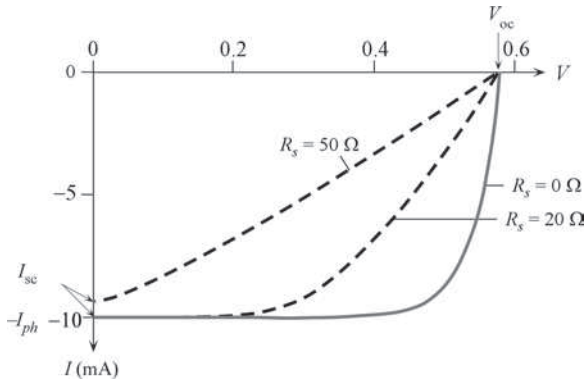


FIGURE 5.44 The effect of series resistance R_s on the *I–V* characteristics. This example is a Si *pn* junction solar cell with $\eta = 1.5$ and $I_o = 3 \times 10^{-6}$ mA. The light intensity is such that it generates $I_{ph} = 10$ mA.

cell efficiency. Notice also that, as expected, V_{oc} is unaffected by R_s, though for sufficiently large R_s, the short circuit current can become reduced. Similarly, low shunt resistance values, due to extensive defects in the material, also reduce the efficiency. The difference is that although R_s does not affect the open circuit voltage V_{oc}, low R_p leads to a reduced V_{oc}. (See Question 5.35.)

D. Solar Cell Structures and Efficiencies

Most solar cells use crystalline silicon because silicon-based semiconductor fabrication is now a mature technology that enables cost-effective devices to be manufactured. One of the most important solar cell metrics is the **efficiency**, which is defined as the maximum output power from the device per unit incident radiation power under well-defined conditions. The Sun's spectrum is different at different heights from Earth due to absorption and scattering effects in the atmosphere, which depend on the wavelength. For terrestrial use, one common standard condition for solar cell comparisons is the so-called global **AM1.5 solar spectrum**[21] that represents the radiation arriving on Earth's surface with an integrated intensity (over all wavelengths) of 1000 W m^{-2}; and the test temperature is 25°C. Table 5.6 lists the efficiencies and various solar cell parameters such as V_{oc}, J_{sc} (short circuit current density), and the fill factor.

TABLE 5.6 Characteristics of a few selected classes of solar cells, and reported efficiencies under global AM1.5 solar spectrum (1000 W m^{-2}) at 25°C

Semiconductor	V_{oc} (V)	$\lvert J_{sc} \rvert$ (mA cm^{-2})	FF (%)	Efficiency (%)	E_g (eV)
Si, single crystal (PERL)	0.707	42.7	82.8	25.0	1.11
Si, polycrystalline	0.664	38.0	80.9	20.4	1.11
Amorphous Si:H (*pin*)	0.886	16.75	67.0	10.1	~1.7
GaAs, single crystal	1.030	29.8	86.0	26.4	1.42
GaAs, thin film	1.107	29.6	84.1	27.6	1.42
InP, single crystal	0.878	29.5	85.4	22.1	1.35
GaInP/GaAs Tandem	2.488	14.22	85.6	30.3	1.95/1.42
GaInP/GaAs/Ge Tandem	2.622	14.37	85.0	32.0	1.95/1.42/0.66

(*Source:* Data have been extracted from M. A. Green, K. Emery, Y. Hishikawa, W. Warta, *Progress in Photovoltaics: Research and Applications, 18*, 346, 2010, and *19*, 84, 2011. The original tables in the latter have extensive confirmed data on a number of important solar cells and modules with references. In addition, the original tables also list the uncertainties and errors involved in the reported efficiency values.)

[21] AM1.5 means Air-Mass1.5. In general, the number m in AMm represents $m = \sec\theta$, where θ is the Zenith angle between the shortest path the sun rays can take and the actual path to the solar cell.

Typical Si-based solar cell efficiencies range from about 18% for polycrystalline to 22–25% in high efficiency single crystal devices that have special structures to absorb as many of the incident photons as possible. The reflection of light from the solar cell surface must be minimized to increase the device efficiency. Indeed, best Si *pn* junction solar cell efficiencies are about 24–25% for expensive single crystal PERL—Passivated Emitter Rear Locally diffused—cells.[22] The PERL and similar cells have a textured surface that is an array of "inverted pyramids" etched into the surface to capture as much of the incoming light as possible as illustrated in Figure 5.45. Normal reflections from a flat crystal surface lead to a loss of light, whereas reflections inside the pyramid allow a second or even a third chance for absorption. Further, after refraction photons would be entering the semiconductor at oblique angles, which means that they will be absorbed in the useful photogeneration volume, that is, within the electron diffusion length of the depletion layer as shown in Figure 5.45. The largest factors reducing the efficiency of a Si solar cell are the unabsorbed photons with $h\upsilon < E_g$ (long wavelengths) and short wavelength photons that are absorbed near the surface. Both of these factors are improved by the use of multifunction heterostructure.

Research over the years has led to the development of high efficiency heterostructure **tandem** solar cells, which are also called **multijunction solar cells.** Multijunction solar cells essentially use two or more cells in tandem, or in cascade, to increase the absorbed photons from the incident light as illustrated in Figure 5.46 (a). The first cell is made from a wider bandgap material and only absorbs photons with $h\upsilon > E_{g1}$. The second cell absorbs photons that pass the first cell and have $h\upsilon > E_{g2}$. The whole structure can be grown by using lattice matched crystalline layers on a suitable substrate, leading to a monolithic tandem cell. The two cells have to be connected, that is, allow the carriers (electrons and holes) to pass. This is done by using a highly doped very thin *pn* junction between the two cells that serves as a tunneling junction; the carriers tunnel through it. All the layers are grown by special techniques on a single substrate. One of the best efficiencies is achieved by using a three junction solar cell, which is illustrated in Figure 5.46 (b). The layers are all grown on a Ge substrate and each cell is an *np* junction. There are two very thin *pn* tunnel junctions that connect the cells in tandem as shown in Figure 5.46 (b), to allow the drifting carriers tunnel through (pass through). The top cell is $GaInP_2$ with $E_g \approx 1.95\,\text{eV}\,(\lambda_g = 0.64\,\mu\text{m})$, the second is GaAs with $E_g \approx 1.42\,\text{eV}\,(\lambda_g = 0.87\,\mu\text{m})$, and the third is Ge with $E_g \approx 0.66\,\text{eV}\,(\lambda_g = 1.9\,\mu\text{m})$. The three cells have a wide spectral range and are able to capture a very high percentage of the solar radiation. The multijunction solar cell in

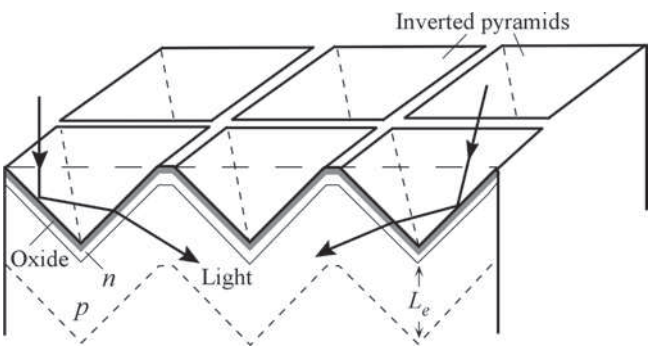


FIGURE 5.45 Inverted pyramid textured surface substantially reduces reflection losses and increases absorption probability in the device.

[22] Much of the pioneering work for high efficiency PERL solar cells was done by Martin Green and coworkers at the University of New South Wales, Australia, (J. Zhao, A. Wang, and M. A. Green, *Appl. Phys. Letts.*, *73*, 1991, 1998.)

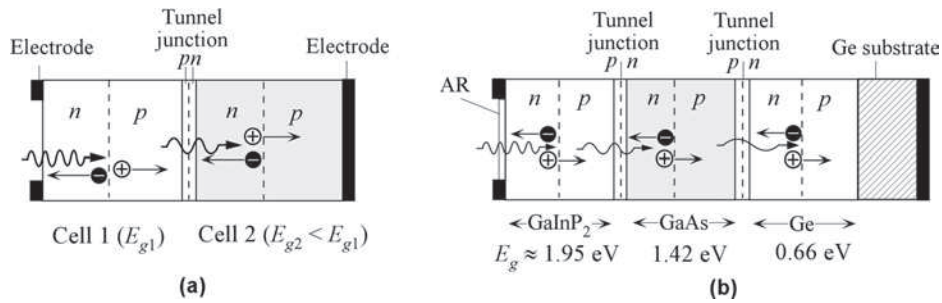


FIGURE 5.46 (a) A heterojunction solar cell that can absorb both high and low energy photons and generate a photocurrent. (b) A tandem solar cell.

Figure 5.46 (b) is commercially available (Spectrolab) with an efficiency of 29.5%. Even higher efficiencies have been reported in research labs using such multijunction heterostructures. If, in addition, light concentrators are also used, the efficiency can be further increased.

Tandem cells are also used in thin film a-Si:H (hydrogenated amorphous Si) *pin* solar cells to obtain efficiencies up to about 12%. These tandem cells are based on a-Si:H and a-SiGe:H and are easily and inexpensively fabricated in large areas. Amorphous Si:H has an E_g of about 1.8 eV. The alloying of a-Si:H with Ge to produce a-SiGe:H decreases E_g. Further, E_g of a-SiGe:H can be graded by controlling the Ge content.

EXAMPLE 5.14.1 Solar cell driving a load

Consider the solar cell driving a 16-Ω resistive load as in Figure 5.42 (b). Suppose that the cell has an area of 1 cm × 1 cm and is illuminated with light of intensity 900 W m^{-2} as in the figure. What are the current and voltage in the circuit? What is the power delivered to the load? What is the efficiency of the solar cell in this circuit? If you assume it is operating close to the maximum deliverable power, what is the FF?

Solution

The *I–V* characteristic of the load is the load line as described by Eq. (5.14.4), $I = -V/R$ with $R = 16\ \Omega$. This line is drawn in Figure 5.42 (b) with a slope $1/(16\ \Omega)$. It cuts the *I–V* characteristics of the solar cell at $I' \approx -31.5$ mA and $V' \approx 0.505$ V, which are the current and voltage in the photovoltaic circuit of Figure 5.42 (b). In fact, from Eq. (5.14.4), V'/I' gives $-16\ \Omega$ as expected. The power delivered to the load is

$$P_{\text{out}} = |I'V'| = (31.5 \times 10^{-3}\ \text{A})(0.505\ \text{V}) = 0.0159\ \text{W} \quad \text{or} \quad 15.9\ \text{mW}$$

This is *not* necessarily the maximum power available from the solar cell. The input sunlight power is

$$P_{\text{in}} = (\text{Light intensity})(\text{Surface area}) = (900\ \text{W m}^{-2})(0.01\ \text{m})^2 = 0.090\ \text{W}$$

The efficiency is

$$\text{Efficiency} = 100 \times (P_{\text{out}}/P_{\text{in}}) = 100(15.9\ \text{mW}/90\ \text{mW}) = 17.7\%$$

This will increase if the load is adjusted to extract the maximum power from the solar cell but the increase will be small as the rectangular area $I'V'$ in Figure 5.42 in it is already close to the maximum. Assuming that $|I'V'|$ is roughly the maximum power available (maximum area for the rectangle $I'V'$), then $I_m \approx I' \approx -31.5$ mA and $V_m \approx V' \approx 0.505$ V. For the solar cell in Figure 5.42 (b), $I_{\text{sc}} = -35.5$ mA and $V_{\text{oc}} = 0.62$ V. Then

$$\text{FF} = I_m V_m / I_{\text{sc}} V_{\text{oc}} \approx (-31.5\ \text{mA})(0.505\ \text{V})/(-35.5\ \text{mA})(0.62\ \text{V}) = 0.72 \quad \text{or} \quad 72\%$$

EXAMPLE 5.14.2 Open circuit voltage and short circuit current

A solar cell under an illumination of 500 W m^{-2} has a short circuit current I_{sc} of −16 mA and an open circuit output voltage V_{oc} of 0.50 V. What are the short circuit current and open circuit voltages when the light intensity is doubled? Assume $\eta = 1$.

Solution

The general *I–V* characteristics under illumination is given by Eq. (5.14.3). The short circuit current corresponds to the photocurrent so that, from Eq. (5.14.2), at double the intensity, the photocurrent is

$$I_{ph2} = \left(\frac{I_2}{I_1}\right) I_{ph1} = (16\text{ mA})(1000/500) = 32\text{ mA}$$

Setting $I = 0$ for open circuit we can obtain the open circuit voltage V_{oc},

$$I = -I_{ph} + I_o\left[\exp(eV_{oc}/\eta k_B T) - 1\right] = 0$$

Assuming that $V_{oc} \gg \eta k_B T/e$, rearranging the above equation we can find V_{oc}

$$V_{oc} = \frac{\eta k_B T}{e} \ln\left(\frac{I_{ph}}{I_o}\right) \tag{5.14.6}$$

Open circuit output voltage

In Eq. (5.14.6), the photocurrent, I_{ph}, depends on the light intensity I via $I_{ph} = KI$. At a given temperature, then the change in V_{oc} is

$$V_{oc2} - V_{oc1} = \frac{\eta k_B T}{e} \ln\left(\frac{I_{ph2}}{I_{ph1}}\right) = \frac{\eta k_B T}{e} \ln\left(\frac{I_2}{I_1}\right)$$

Assuming $\eta = 1$, the new open circuit voltage is

$$V_{oc2} = V_{oc1} + \frac{\eta k_B T}{e} \ln\left(\frac{I_2}{I_1}\right) = 0.50\text{ V} + (1)(0.0259\text{ V})\ln(2) \approx 0.52\text{ V}$$

This is a ~4% increase in V_{oc} compared with the 100% increase in illumination and the short circuit current.

Questions and Problems

5.1 Bandgap and photodetection

(a) Determine the maximum value of the energy gap which a semiconductor, used as a photoconductor, can have if it is to be sensitive to light of wavelength 1550 nm commonly used in optical communication network.

(b) A photodetector whose area is 4×10^{-2} cm^2 is irradiated with light of wavelength 1300 nm, whose intensity is 3 mW cm^{-2}. Assuming that each photon generates one electron–hole pair, calculate the number of pairs generated per second.

(c) From the known energy gap of the semiconductor InGaAsP ($E_g = 1.24$ eV), calculate the primary wavelength of photons emitted from this crystal as a result of electron–hole recombination. Is this wavelength in the visible range?

(d) Will a silicon photodetector be sensitive to the radiation from an InGaAsP laser? Why?

5.2 Absorption coefficient In the following questions assume that all the incident radiation onto the semiconductor surface is transmitted into the semiconductor medium; that is, there is a perfect antirelection coating on the surface of the semiconductor.

(a) If d is the thickness of a photodetector material, I_o is the intensity of the incoming radiation, show that the number of photons absorbed per unit volume of sample is

$$n_{ph} = \frac{I_o\left[1 - \exp(-\alpha d)\right]}{dh\nu}$$

(b) What is the thickness of a Ge and $In_{0.53}Ga_{0.47}As$ crystal layer that is needed for absorbing 90% of the incident radiation at 1.5 μm? (Use Figure 5.5)

(c) Suppose that each absorbed photon liberates one electron (or electron hole pair) in a unity quantum efficiency photodetector and that the photogenerated electrons are immediately collected. Thus, the rate of charge collection is limited by rate of photon generation. What is the external photocurrent density for the photodetectors in (b) if the incident radiation is 100 μW mm^{-2}?

5.3 Absorption coefficient and device design An optoelectronic device engineer has designed a *pn* junction photodiode that has a depletion width of 2 μm. The semiconductor on the light-receiving side is heavily doped (*e.g.*, p^+), the neutral region is very narrow, and can be neglected. If the photogeneration is to occur in the depletion region, what are approximate corresponding wavelengths for the semiconductors in Figure 5.5?

5.4 Absorption coefficient and quantum efficiency Consider the Si *pn* junction photodiode shown in Figure 5.8. Suppose that the incident optical power is P_o. The incident number of photons per unit time is $P_o/h\upsilon$. Of these, a fraction T is transmitted. The number of photons flowing per second along the device is

$$\frac{dN_{\text{ph}}}{dt} = \frac{TP_o}{h\upsilon}\exp(-\alpha x)$$

where x is measured from the illuminated surface of the semiconductor. Show that the photocurrent generated by the photodiode is given by

$$I_{ph} = \frac{e\eta_i TP_o}{h\upsilon}\left\{\exp\left[-\alpha(\ell_p - L_e)\right] - \exp\left[-\alpha(\ell_p - L_e + W + L_h)\right]\right\}$$

where η_i is the internal quantum efficiency and the other symbols are defined in Figure 5.8. Consider two photodiodes, one p^+n and the other *pin*. Their properties are listed in Table 5.7, where the photocurrent has been calculated at 800 nm. Suppose that the incident optical power is 1 μW, and the antireflection coating on the semiconductor provides $T \approx 1$. Find the photocurrent I_{ph} and hence the responsivity of the two photodiodes at 900 nm and 500 nm. What is your conclusion?

TABLE 5.7 Properties of a p^+n and *pin* photodiode

	ℓ_p (μm)	L_e (μm)	W (μm)	L_h (μm)	α (m^{-1}) (at 800 nm)	I_{ph} (μA)
p^+n	0.5	0.1	1	10	1×10^5	0.45
pin	0.5	0.1	30	0.1	1×10^5	0.59

Note: $\eta_i = 1$, $T = 1$ and $P_o = 1$ μW were used.

5.5 Ge Photodiode Consider a commercial Ge *pn* junction photodiode which has the responsivity shown in Figure 5.47. Its photosensitive area is 0.008 mm^2. It is used under a reverse bias of 10 V when the dark current is 0.3 μA and the junction capacitance is 4 pF. The rise time of the photodiode is 0.5 ns.

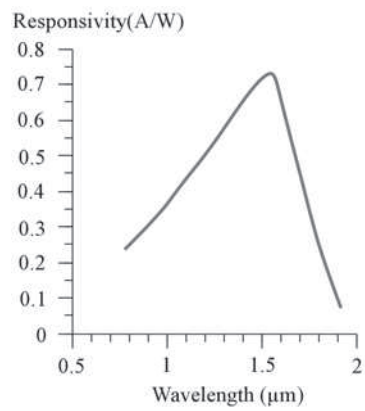


FIGURE 5.47 The responsivity of a commercial Ge *pn* junction photodiode.

(a) Calculate its quantum efficiency at 850 nm, 1300 nm, and 1550 nm.
(b) What is the light intensity of light at 1.55 μm that gives a photocurrent equal to the dark current?
(c) What would be the effect of lowering the temperature on the responsivity curve?
(d) Given that the dark current is in the range of microamperes, what would be the advantage in reducing the temperature?
(e) Suppose that the photodiode is used with a 100-Ω resistance to sample the photocurrent. What limits the speed of response?

5.6 Si *pin* Photodiodes Consider two commercial Si *pin* photodiodes, type A and type B, both classified as fast *pin* photodiodes. They have the responsivity curves shown in Figure 5.48. Differences in the responsivity are due to the *pin* device structure. The photosensitive area is 0.125 mm^2 (0.4 mm in diameter).

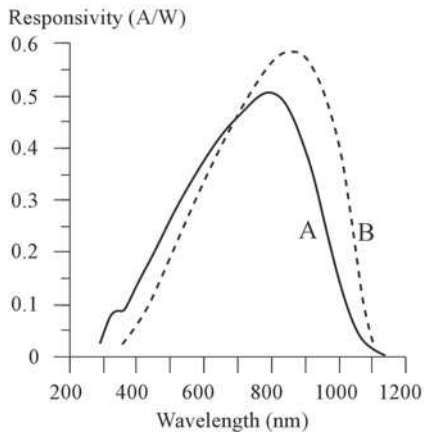


FIGURE 5.48 The responsivity of two commercial Si *pin* photodiodes.

(a) Calculate the photocurrent from each when they are illuminated with blue light of wavelength 450 nm and light intensity 1 μW cm^{-2}. What is the QE of each device?
(b) Calculate the photocurrent from each when they are illuminated with red light of wavelength 700 nm and light intensity 1 μW cm^{-2}. What is the QE of each device?
(c) Calculate the photocurrent from each when they are illuminated with infrared light of wavelength 1000 nm and light intensity 1 μW cm^{-2}. What is the QE of each device?
(d) What is your conclusion?

5.7 InGaAs *pin* Photodiodes Consider a commercial InGaAs *pin* photodiode whose responsivity is shown in Figure 5.49. Its dark current is 5 nA.

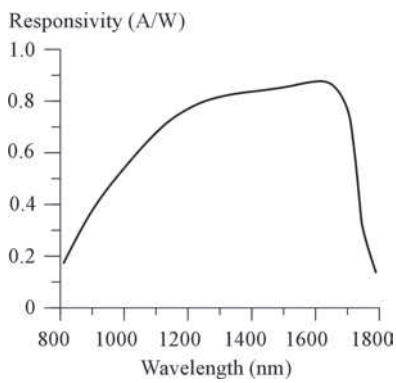


FIGURE 5.49 The responsivity of an InGaAs pin photodiode.

(a) What optical power at a wavelength of 1.55 μm would give a photocurrent that is twice the dark current? What is the QE of the photodetector at 1.55 μm?
(b) What would be the photocurrent if the incident power in (a) was at 1.3 μm? What is the QE at 1.3 μm operation?

5.8 InAs Photodiode Figure 5.50 shows the responsivity of an InAs *pn* junction photodiode for use in the infrared. It is usually cooled under normal operation.

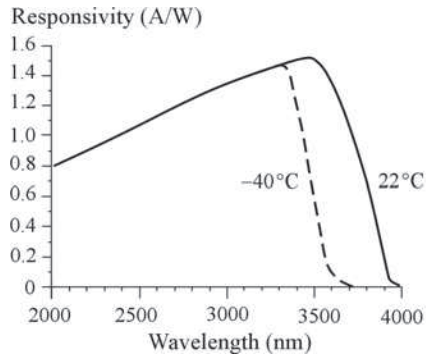


FIGURE 5.50 Responsivity of an InAs photodiode.

(a) What is the wavelength and the quantum efficiency at the peak responsivity at −40°C?
(b) The diode area that is involved in the detection has a diameter of 0.25 mm. What is the intensity of light that corresponds to a photocurrent of 50 nA at peak responsivity?
(c) Why does the cutoff wavelength become shorter at lower temperatures?

5.9 Photocurrent *pin* photodiode Consider a Si *pin* photodiode with a depletion layer width of 30 μm and diameter of 300 μm. The absorption coefficient at 800 nm is 10^3 cm^{-1}.

Find the quantum efficiency of this photodiode, assuming perfect AR coating. If the load resistance is 10 kΩ, obtain the RC time constant. What would happen if there is no AR coating so that $T = 0.70$?

5.10 Shockley–Ramo theorem Consider a *pin* photodiode with an *i*-layer width *W*. A very short pulse of light is absorbed just inside the depletion region on the p^+-side (Figure 5.9). The photogenerated electrons drift through the *i*-layer with a velocity given in Figure 5.10. What is the current generated by this drift? How long does it last? Suppose that *W* is 30 μm, and assume that the quantum efficiency is 0.80. A voltage of 20 V is applied to reverse bias the *pin* detector. The light pulse from a femtosecond laser operating at 515 nm is used for photoexcitation, and the light pulse energy is 37.5 fJ. Assume all of this energy is absorbed very close to the depletion region and p^+-layer boundary, *i.e.* electrons drift across *W* and constitute the photocurrent. Calculate the transient photocurrent. How long does it last?

5.11 QE and Responsivity A photodiode has a quantum efficiency of 70% when photons of energy 1.5×10^{-19} J are incident upon it. What is the operating wavelength of this photodiode? Find out the incident optical power required to obtain a photocurrent of 2.5 μA with the above operating wavelength. What would be the responsivity when the QE is 100%?

5.12 Si *pin* photodiode speed Consider Si *pin* photodiodes with a p^+-layer of thickness 1.0 μm, and an *i*-Si layer of width 7.5 μm. It is reverse biased with a voltage of 25 V.
(a) What is the speed of response due to bulk absorption? What wavelengths would lead to this type of speed of response?
(b) What is the speed of response due to absorption near the surface? What wavelengths would lead to this type of speed of response?

5.13 Transient photocurrents in a *pin* photodiode Consider a reverse biased Si *pin* photodiode as shown in Figure 5.51. It is appropriately reverse biased so that the field in the depletion region (*i*-Si layer), $E = V_r/W$, is the saturation field. Thus, photogenerated electrons and holes in this layer drift at saturation velocities v_{de} and v_{dh}. Assume that the field *E* is uniform and that the thickness of the p^+ is negligible. A very short light pulse (infinitesimally short) photogenerates EHPs in the depletion layer as shown in Figure 5.51, which results in an exponentially decaying EHP concentrations across *W*. Figure 5.51 shows the photogenerated electron concentration at time $t = 0$ and also at a later time *t*, when the electrons have drifted a distance $\Delta x = v_{de}t$. Those that reach the back electrode *B* become collected. The electron distribution shifts at a constant velocity until the

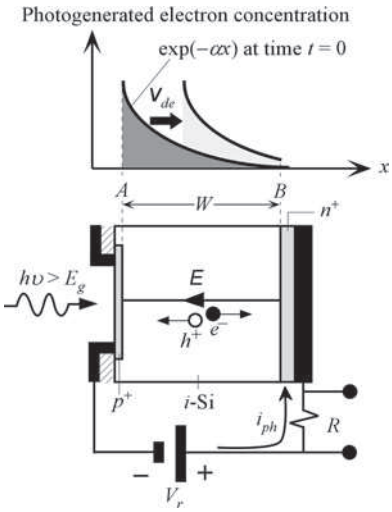


FIGURE 5.51 An infinitesimally short light pulse is absorbed throughout the depletion layer of a reverse-biased ideal *pin* photodetector, and creates an EHP concentration that decays exponentially. The p^+- and n^+-layers are assumed to be very thin.

initial electrons at A reach B, which represents the longest transit time $\tau_e = W/v_{de}$. Similar arguments apply to holes, but they drift in the opposite direction and their transit time $\tau_h = W/v_{dh}$ where v_{dh} is their saturation velocity. The photocurrent density at any instant is

$$j_{ph} = j_e(t) + j_h(t) = eN_e v_{de} + eN_h v_{dh}$$

where N_e and N_h are the overall electron and hole concentration in the sample at time t. Assume for convenience that the cross-sectional area $A = 1$ below (derivations are not affected as we are interested in the photocurrent current densities).

(a) Sketch the hole distribution at time $\tau_h > t > 0$ where τ_h = hole drift time $= W/v_{dh}$.

(b) The electron concentration distribution $n(x)$ at time t corresponds to that at $t = 0$ shifted by $v_{de}t$. Thus the total electrons in W is proportional to integrating this distribution $n(x)$ from A at $x = v_{de}t$ to B at $x = W$. Given $n(x) = n_o \exp(-\alpha x)$ at $t = 0$, where n_o is the electron concentration at $x = 0$ at $t = 0$ we have

$$\text{Total number electrons at time } t = \int_{v_{de}t}^{W} n_o \exp\left[-\alpha(x - v_{de}t)\right] dx$$

and

$$N_e(t) = \frac{\text{Total number electrons at time } t}{\text{Volume}}$$

Then

$$N_e(t) = \frac{1}{W}\int_{v_{de}t}^{W} n_o \exp\left[-\alpha(x - v_{de}t)\right] dx$$

$$= \frac{n_o}{W\alpha}\left\{1 - \exp\left[-\alpha W\left(1 - \frac{t}{\tau_e}\right)\right]\right\}$$

$N_e(0)$ is the initial overall electron concentration at time $t = 0$, that is,

$$N_e(0) = \frac{1}{W}\int_0^W n_o \exp(-\alpha x) dx = \frac{n_o}{W\alpha}\left[1 - \exp(-\alpha W)\right]$$

We note that n_o depends on the intensity I of the light pulse so that $n_o \propto I$. Show that for holes

$$N_h(t) = \frac{n_o \exp(-\alpha W)}{W\alpha}\left\{\exp\left[\alpha W\left(1 - \frac{t}{\tau_h}\right)\right] - 1\right\}$$

(c) Given $W = 40\ \mu m$, $\alpha = 5 \times 10^4\ m^{-1}$, $v_{de} = 10^5\ m\ s^{-1}$, $v_{dh} = 0.8 \times 10^5\ m\ s^{-1}$, $n_o = 10^{13}\ cm^{-3}$, calculate the electron and hole transit time, sketch the photocurrent densities $j_e(t)$ and $j_h(t)$ and hence $j_{ph}(t)$ as a function of time, and calculate the initial photocurrent. What is your conclusion?

5.14 Fiber attenuation and InGaAs *pin* Photodiode Consider the commercial InGaAs *pin* photodiode whose responsivity is shown in Figure 5.49. This is used in a receiver circuit that needs a minimum of 6 nA photocurrent for a discernible output signal (acceptable signal to noise ratio for the customer). Suppose that the InGaAs *pin* PD is used at 1.3-μm operation with a single mode fiber whose attenuation is 0.4 dB km^{-1}. If the laser diode emitter can launch at most 3 mW of power into the fiber, what is the maximum distance for the communication without a repeater?

5.15 Multiplication in APDs Consider the results of multiplication vs. reverse bias experiments on a commercial InGaAs APD summarized in Table 5.8. Plot $(1 - 1/M)$ vs. V_r and hence find m and V_{br} in Eq. (5.6.2). (This can be done, for example, by using Excel and then fitting a power law trend line; or plotting on a log-log axis in which case the slope would give m.)

TABLE 5.8 Multiplication (M) vs. reverse voltage (V_r) data for an InGaAs commercial APD

V_r	30.6	33.5	36.6	39.5	41.9	43.4	44.3	45.46	45.92	46.21
M	1.78	2.16	2.74	3.74	5.34	7.21	9.0	14.2	19.5	24.8

5.16 Effect of temperature on APDs Experiments on two commercial Si APDs A and B have examined how the temperature affects the reverse bias V_r needed to keep the multiplication constant. The results are summarized in Table 5.9 in terms of V_r values at different temperatures that would keep $M = 100$ and 200 for A and 100 for B. Plot V_r vs. T (in K) for A and B. Do you agree that empirically $V_r \propto T^n$? What is n?

TABLE 5.9 The reverse bias voltage V_r for given multiplication M at different temperatures for two APDs A and B

T	−20°C	0°C	20°C	40°C	60°C	Comment
V_r at $M = 100$	125	162.5	200	236	273	APD A
V_r at $M = 100$	135	173.5	211	249	286	APD A
V_r at $M = 100$	147	150	153	155	157	APD B

5.17 Multiplication in APDs The avalanche multiplication gain M is given by

$$M = \frac{1 - k}{\exp\left[-(1 - k)\alpha_e w\right] - k}$$

Consider two cases (a) $k = 0$ and (b) $k = 0.1$. Suppose that w is 2 μm. Increasing the field will increase α_e. Plot M vs. α_e semilogarithmically for the two cases over the range $\alpha_e = 0$ to 2. What is your conclusion? (*Note*: As α_e increases with the field, k does not strictly remain the same. However, the two cases clearly distinguish two distinctly different types of behavior.)

5.18 Si APD The electron and hole ionization coefficients α_e and α_h in silicon are approximately given in Example 5.6.3. Suppose we would like the avalanche to be achieved at a certain applied field in the avalanche region (corresponding to a particular desirable voltage range). Suppose that we would like the multiplication M to be 100 when $E = 5 \times 10^5\ V\ cm^{-1}$. Estimate the width w of the avalanche region.

5.19 InP APD design For InP the impact ionization coefficients are roughly given by $\alpha_h \approx (9.2 \times 10^6) \exp(-3.44 \times 10^6/E)$ and $\alpha_h \approx (4.3 \times 10^6)\exp(-2.72 \times 10^6/E)$, where α_e and α_h are in cm^{-1}, and E is in $V\ cm^{-1}$ (from K. Takuchi *et al.*, *J. Appl. Phys.*, *59*, 476, 1986). Consider the heterojunction InGaAs-InP APD in Figure 5.16. The avalanche is initiated by holes in N-InP. We can define $k = \alpha_e/\alpha_h$ and then the multiplication $M = (1 - k)/\left\{\exp\left[-(1 - k)\alpha_h w\right] - k\right\}$. Suppose that the width of the N-layer is 1 μm and the applied field is $4.6 \times 10^5\ V\ cm^{-1}$. What is the multiplication M? If you increase the field by 2%, what is M?

5.20 Schottky junction photodiode Consider a *pn* junction and a Schottky junction photodiode. Calculate the photocurrent under an illumination of 2 μW at a wavelength of 400 nm for each, and find the responsivity given (a) the *pn* junction has $\ell_p = 0.5\ \mu m$, $L_e = 0.1\ \mu m$, $W = 2\ \mu m$, $L_h = 10\ \mu m$; and (b) the Schottky junction has $W = 3\ \mu m$. Assume that the internal QE is 100%. (*Note*: You can use the equation in Question 5.3.)

5.21 Photoconductive detector An *n*-type Si photoconductor has a length $L = 100\ \mu m$ and a cross-sectional area $A = 10^{-4}\ mm^2$. The applied bias voltage to the photoconductor is 10 V.

(a) What are the transit times, t_e and t_h, of an electron and a hole across L? What is the photoconductive gain?

(b) It should be apparent that as electrons are much faster than holes, a photogenerated electron leaves the photoconductor very quickly. This leaves behind a drifting hole and therefore a positive charge in the semiconductor. Secondary (*i.e.*, additional electrons) then flow into the photoconductor to maintain neutrality in the sample and the current continues to flow. These events will continue until the hole has disappeared by recombination, which takes on average a time τ. Thus, more charges flow through the contact per unit time than charges actually photogenerated per unit time. What will happen if the contacts are not ohmic, *i.e.,* they are not injecting?

(c) What can you say about the product $\Delta\sigma$ and the speed of response, which is proportional to $1/\tau$.

5.22 Photoconductive detectors: PbSe for 4 μm detection The basic principle in photoconductive detectors is the change in the resistance of the semiconductor upon exposure to light. Figure 5.52 shows how PbS and PbSe photodetectors are often used. The photoconductor (PC) has a dark resistance R_d and is biased by V_B through a load R_L. A chopper (either mechanical or electronic) chops the light at a frequency f_c. The resistance of the PC changes periodically at the chopper frequency. The periodic change in R_d causes a periodic change in the current. This periodic change is the photocurrent $i_{ph}(t)$ signal and is an AC-type signal at the frequency f_c. The photocurrent generates a periodic voltage signal $v_{ph}(t)$ across R_d, which can be coupled through a coupling capacitor C_c into a lock-in amplifier (LIA). This amplifier is synchronized with the chopper and only amplifies the signal if it is in phase with the chopper. Its output is a DC signal that represents the magnitude of v_{ph} that is in phase with the chopped light. Consider a PbSe photoconductor for detecting 4 μm radiation. The circuit in Figure 5.52 has been implemented with $V_B = 15\,V$, $R_d = 1\,M\Omega$, and $R_L = R_d$. A test experiment has shown that incident optical power of 50 nW on the detector's photosensitive area ($3 \times 3\ mm^2$) decreases R_d by 6.0 Ω. (Put differently, $\delta R_d = -6\ \Omega$ for incident radiation of 50 nW.) Find the signal voltage V_{signal}, *i.e.,* the change δV in the voltage across R_d, and the photosensitivity, defined as V_{signal}/P_o. (b) What is the photosensitivity if $R_L = 2R_d$ and $R_L = R_d/2$? (c) For the case $R_L = R_d$, find the NEP and hence the specific detectivity D^*. Assume that the only noise source is the thermal noise from R_d and R_L. The quoted D^* for this device is $2.5 \times 10^9\ cm\ Hz^{1/2}\ W^{-1}$.

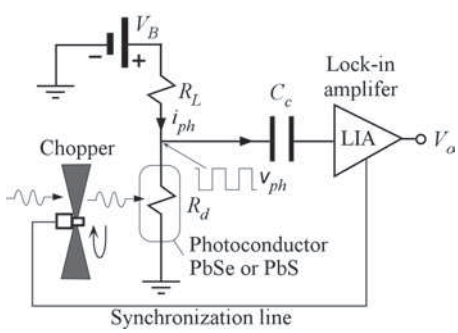


FIGURE 5.52 A photodetection system based on using a photoconductive detector, a chopper to chop the light, and a lock-in amplifier (LIA).

5.23 SNR of an InGaAs *pin* A particular photodetection application requires an InGaAs photodetector and needs a bandwidth of 2 GHz. The dark current of the InGaAs *pin* detector is 6 nA at 30°C. The minimum signal that is to be measured is 5 nW at 1550 nm where the responsivity is $0.95\ A\ W^{-1}$.

(a) Calculate the SNR in dB at 30°C.

(b) When the detector is cooled to −30°C, the dark current becomes 0.1 nA, and the responsivity is about the same. What is the new SNR?

(c) Suppose that we operate the detector over a bandwidth of 20 MHz. What is the new SNR at −30°C?

5.24 Detector and receiver noise Consider an InGaAs *pin* photodiode used in a receiver circuit as in Figure 5.31 with a load resistor of 30 kΩ. The total capacitance of the detector and the input of the amplifier together is 20 pF. The photodiode has a dark current of 1 nA. The incident radiation is 6 nW at 1550 nm where the responsivity is 0.9 A W^{-1}. Assuming that the amplifier is noiseless, calculate the SNR at 27°C.

5.25 The NEP and Ge and InGaAs photodiodes

(a) Show that the noise equivalent power of a photodiode is given by

$$\text{NEP} = \frac{P_1}{B^{1/2}} = \frac{hc}{\eta e \lambda}\left[2e(I_d + I_{ph})\right]^{1/2}$$

How would you improve the NEP of a photodiode? What is *NEP* for an ideal PD at operating at $\lambda = 1.55\,\mu$m?

(b) Given the dark current I_d of a PD, show that for SNR = 1, the photocurrent is

$$I_{ph} = eB\left[1 + \left(1 + \frac{2I_d}{eB}\right)^{1/2}\right]$$

What is the corresponding optical power P_1?

(c) Consider a fast Ge *pn* junction PD which has a photosensitive area of diameter 0.3 mm. It is reverse biased for photodetection and has a dark current of 0.5 μA. Its peak responsivity is 0.7 A V^{-1} at 1.55 μm (see Figure 5.47). The bandwidth of the photodetector and the amplifier circuit together is 100 MHz. Calculate its NEP at the peak wavelength and find the minimum optical power and hence minimum light intensity that gives a SNR of 1. How would you improve the minimum detectable optical power?

(d) Table 5.10 shows the characteristics of typical Ge *pn* junction and InGaAs *pin* photodiodes in terms of responsivity and the current. Fill in the remainder of the columns in the table assuming that there is an ideal, noiseless preamplifier to detect the photocurrent from the photodiode. Assume a working bandwidth, *B*, of 1 MHz. What is your conclusion?

TABLE 5.10 Ge *pn* junction and InGaAs *pin* PDs. Photosensitive area has a diameter of 1 mm

Photodiode	R at 1.55 μm (A V^{-1})	I_d (nA)	I_{ph} (nA) for SNR = 1 at B = 1 MHz	Optical power (nW) for SNR = 1 at B = 1 MHz	NEP (W $Hz^{-1/2}$)	Comment
Ge at 25°C	0.8	400				
Ge at –20°C	0.8	5				Thermoelectric cooling
InGaAs *pin*	0.95	3				

5.26 Specific detectivity of Si photodiodes Consider the NEP measurements shown in Table 5.11 for a collection of Si photodiodes. (a) For each calculate the specific detectivity. Then find the average and the mean standard deviation. (b) Plot NEP vs. $A^{1/2}$ on log-log plot and find the slope. What does the slope represent?

TABLE 5.11 Photosensitive area and NEP for a collection of Si photodiodes

A (mm^2)	7	13	20	25	66	100	324	784
NEP ($\times 10^{-15}$ W $Hz^{-1/2}$)	9.5	13	17	16	25	38	66	86

5.27 Specific detectivity of InGaAs photodiodes Consider the results of NEP measurements shown in Table 5.12 for a few InGaAs pin photodiodes with different photosensitive areas. (a) Plot the dark current vs. A on log-log plot and find the slope. What does the slope represent? (b) Plot NEP vs. A on log-log plot and find the slope. What does the slope mean?

TABLE 5.12 Photosensitive area and NEP for a collection of InGaAs photodiodes at −10°C

A (mm^2)	0.785	3.14	7.07	19.6
Dark current (nA)	0.07	0.3	1	2.5
NEP ($\times 10^{-15}$ W Hz$^{-1/2}$)	5	10	20	30

5.28 Excess avalanche noise in InGaAs communications detector Consider a particular commercial 1550 nm InGaAs communications detector whose properties are as follows: Diameter of sensitive area is 75 μm, the unmultiplied dark current (I_{do}) is 0.7 nA. Its responsivity at 1550 nm (peak responsivity) is 0.75 A W^{-1}. The excess noise factor F has been measured to be 3.5 at $M = 10$. Assume the operating wavelength is 1550 nm. The APD normally biased to operate at a multiplication (gain) of $M = 10$ and has a bandwidth of 2 GHz when connected to a suitable external load (preamplifier). We will ignore the load and the preamplifier connection and simply focus on the APD's performance on its own within a bandwidth of 2 GHz. (a) Find the NEP under $M = 1$ (with $B = 1$ Hz). (b) What is the NEP when the APD is operating at $M = 10$? (c) What is the incident detectable optical power and hence intensity for operation at an SNR of 10 at $M = 1$. Suppose that this incident power is kept the same but we introduce gain and set $M = 10$. What is the new SNR? What is your conclusion?

5.29 Avalanche photodiodes and excess avalanche noise APDs exhibit excess avalanche noise which contributes to the shot noise of the diode current. The total noise current in the APD is given by

$$i_{n\text{–APD}} = \left[2e(I_{do} + I_{pho})M^2FB\right]^{1/2} \quad \textbf{(P1.1)}$$

where F is the excess noise factor which depends in a complicated way not only on M but also on the ionization probabilities of the carriers in the device. It is normally taken simply to be M^x where x in an index that depends on the semiconductor material and device structure.

(a) Table 5.13 provides measurements of F vs. M on a Ge APD using photogeneration at 1.55 μm. Find x in $F = M^x$. How good is the fit?

TABLE 5.13 Data for excess avalanche noise as F vs. M for a Ge APD

M	1	3	5	7	9
F	1.1	2.8	4.4	5.5	7.5

(*Source:* Data from D. Scansen and S. O. Kasap, *Cnd. J. Phys. 70*, 1070, 1992.)

(b) The above Ge APD has an unmultiplied dark current of 0.5 μA and an unmultiplied responsivity of 0.8 A V^{-1} at its peak response at 1.55 μm and is biased to operate at $M = 6$ in a receiver circuit with a bandwidth of 500 MHz. What is the minimum photocurrent that will give a SNR = 1? If the photosensitive area is 0.3 mm in diameter what is the corresponding minimum optical power and light intensity?

(c) What should be the photocurrent and incident optical power for SNR = 10? What would be the effect of cooling the APD so that the dark current is reduced by a factor of 100, but M is the same?

5.30 Photodetector materials, devices, and their limitations

(a) What limits the operation of a photodetector when the absorption depth ($\delta = 1/\alpha$) at short wavelengths becomes so narrow that EHPs are generated almost at the crystal surface?

(b) Quantum efficiency defined in terms of incident photons applies to the whole device and includes the effects arising from reflections from the semiconductor surface. What is the percentage of photons lost due to reflections at a Si crystal surface if the refractive index of Si is 3.5. How can you improve the transmitted number of photos into the semiconductor crystal?

(c) In some applications such as measuring the light intensity in the visible range from a source that also emits extensively in the infrared (such as an incandescent light source), it is necessary to use an infrared heat filter. Why?

(d) Consider a heterojunction APD such as that shown in Figure 5.16. For InP, $E_g = 1.35$ eV and for InGaAs $E_g = 0.75$ eV. Obviously 1.5 μm photons will not be absorbed in InP. What is the effect of mismatch in the refractive index n between the two semiconductors? Is it important if $\Delta n \approx 0.2$, and $n \approx 3.5$?

(e) What determines the speed of operation of the phototransistor in Figure 5.25? Consider how long it takes for the photoinjected hole into the p-type base to become neutralized by recombination.

5.31 CMOS and CCD image sensors Consider the basic principle of operation of the CMOS and CCD image sensors in Section 5.13. (a) What constitutes the signal in the CMOS and CCD sensors? (b) Using the responsivity or sensitivity definition used for photodetectors, how would you define the responsivity of CMOS and CCD sensors? (c) The **fill factor** defines what fraction of the pixel area is able to capture the light and convert it to a signal. For example, what fraction of the area is allocated to the photodiode in the pixel of a CMOS? Why is FF important? (d) Considering the operation of the CMOS and CCD sensors, which one is faster for a given number of megapixels? (e) Considering the principle of operation, which sensor would have lower noise?

5.32 CCD image sensors

(a) A particular full frame square CCD image sensor (Kodak KAF-1001) has 1024 (H) × 1024 (V) pixels, where H and V represent horizontal and vertical directions, respectively. The pixel size is a square with a side 24 μm. What is the optical imaging area of the sensor? If the data transfer rate is 2 MHz, what is the total time to transfer one frame of the image?

(b) In a CCD, charge is transferred from one well to the next, and a quantity of importance is the **charge transfer efficiency** η_t. $1 - \eta_t$ represents the fraction of electrons lost in the transfer process. After n transfers from well to well, the collected charge Q_n with respect to the original signal Q_o is

$$Q_n/Q_o = \eta^n$$

If the CCD in (a) has a $\eta = 0.9997$, calculate the charge collection efficiency of the CCD.

5.33 Solar cell driving a load Consider the *I–V* characteristics of the solar cell shown in Figure 5.40 (b). Suppose that we connect a resistive load of 5 Ω across it. Find the current through and the voltage across the load at 800 W m^{-2} and 400 W m^{-2}.

5.34 Solar cell open circuit voltage and short circuit A solar cell under an illumination of 300 W m^{-2} has a short circuit current I_{sc} of 60 mA and an open circuit output voltage V_{oc}, of 0.60 V. What are the short circuit current and open circuit voltages when the light intensity is 600 W m^{-2}?

5.35 Shunt resistance Consider the equivalent circuit of a solar cell as shown in Figure 5.43 (b).

(a) Show that

$$I = -I_{ph} + I_{\text{diode}} + \frac{V}{R_p} = -I_{ph} + I_o \exp\left(\frac{eV}{\eta k_B T}\right) - I_o + \frac{V}{R_P}$$

(b) Plot I vs. V for a polycrystalline Si solar cell that has $\eta = 2$ and $I_o = 3 \times 10^{-4}$ mA, for an illumination such that $I_{ph} = 5$ mA. Use $R_p = \infty, 1000\,\Omega$ and then $R_p = 100\,\Omega$; and plot the curves on the graph for comparison. What is your conclusion?

5.36 Series resistance Consider a solar cell that has $I_o = 30 \times 10^{-6}$ mA, $\eta = 1.5$, and negligible shunt conductance ($R_p = \infty$). Suppose that the illumination is such that $I_{ph} = 9$ mA. Plot the individual *I–V* characteristics for $R_s = 0\,\Omega$, 40 Ω, and 90 Ω.

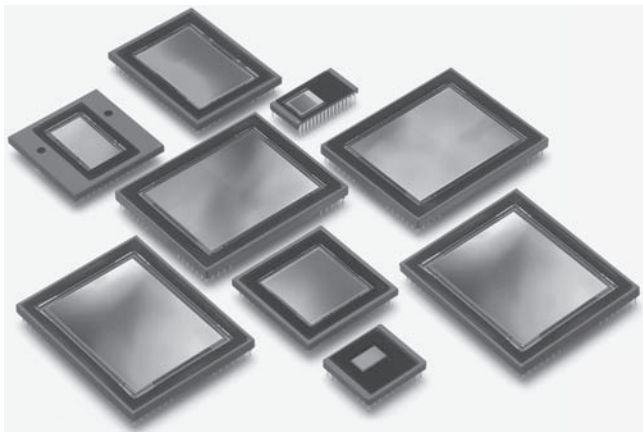
5.37 Solar cell near Eskimo Point

(a) The intensity of light arriving at a point on earth, where the solar latitude is α, can be approximated by the Meinel and Meinel equation

$$I \approx 1.353(0.7)^{(\text{cosec}\alpha)^{0.678}} \quad (\text{kW m}^{-2})$$

where $\text{cosec}\,\alpha = 1/\sin\alpha$. The solar latitude α is the angle between the sun rays and the horizon. Around September 23rd and March 22nd, the sun rays arrive parallel to the plane of the equator. Plot intensity I vs. α. What is your conclusion? What is the maximum power available for a photovoltaic device panel of area 1 m^2 if its efficiency of conversion is 20%?

(b) Manufacturer's characterization tests on a particular Si *p-n* junction solar cell at 27°C specifies an open circuit output voltage of 0.45 V and a short circuit current of 400 mA when illuminated directly with light of intensity 1 kW m^{-2}. The fill factor for the solar cell is 0.73. This solar cell is to be used in a portable equipment application near Eskimo Point (Nunavut, Canada) at a latitude (ϕ) of 63°. The latitude ϕ and α are related by $\phi + \alpha = \pi/2$. Calculate the open circuit output voltage and the maximum available power when the solar cell is used at noon on September 23rd, when the temperature is around −10°C. What is the maximum current this solar cell can supply to an electronic equipment? What is your conclusion? (Assume that the ideality factor is 1.)



Various image sensors. *(Courtesy of Teledyne-DALSA.)*

As there are two different refractions, I conceived also that there are two different emanations of the waves of light…

—Christiaan Huygens

Sailors visiting Iceland during the seventeenth century brought back to Europe calcite crystals (Iceland spar) which had the unusual property of showing double images when objects were viewed through it. The Danish scientist Rasmus Bartholin (Erasmus Bartholinus) described this property in 1669 as the effect of double refraction, and later Christiaan Huygens (1629–1695), a Dutch physicist, explained this double refraction in terms of ordinary and extraordinary waves. Christiaan Huygens made many contributions to optics and wrote prolifically on the subject. *(Morphart Creations Inc./Shutterstock.com)*