Chapter 4

Optical Receivers

The role of an optical receiver is to convert the optical signal back into electrical form and recover the data transmitted through the lightwave system. Its main component is a photodetector that converts light into electricity through the photoelectric effect. The requirements for a photodetector are similar to those of an optical source. It should have high sensitivity, fast response, low noise, low cost, and high reliability. Its size should be compatible with the fiber-core size. These requirements are best met by photodetectors made of semiconductor materials. This chapter focuses on photodetectors and optical receivers [1]–[9]. We introduce in Section 4.1 the basic concepts behind the photodetection process and discuss in Section 4.2 several kinds of photodetectors commonly used for optical receivers. The components of an optical receiver are described in Section 4.3 with emphasis on the role played by each component. Section 4.4 deals with various noise sources that limit the signal-to-noise ratio in optical receivers. Sections 4.5 and 4.6 are devoted to receiver sensitivity and its degradation under nonideal conditions. The performance of optical receivers in actual transmission experiments is discussed in Section 4.7.

4.1 Basic Concepts

The fundamental mechanism behind the photodetection process is optical absorption. This section introduces basic concepts such as responsivity, quantum efficiency, and bandwidth that are common to all photodetectors and are needed later in this chapter.

4.1.1 Detector Responsivity

Consider the semiconductor slab shown schematically in Fig. 4.1. If the energy $h\nu$ of incident photons exceeds the bandgap energy, an electron–hole pair is generated each time a photon is absorbed by the semiconductor. Under the influence of an electric field set up by an applied voltage, electrons and holes are swept across the semiconductor, resulting in a flow of electric current. The photocurrent $I_p$ is directly proportional to
the incident optical power $P_{\text{in}}$, i.e.,

$$I_p = R P_{\text{in}}, \quad (4.1.1)$$

where $R$ is the responsivity of the photodetector (in units of A/W).

The responsivity $R$ can be expressed in terms of a fundamental quantity $\eta$, called the quantum efficiency and defined as

$$\eta = \frac{\text{electron generation rate}}{\text{photon incidence rate}} = \frac{I_p/q}{P_{\text{in}}/h\nu} = \frac{h\nu}{q} R, \quad (4.1.2)$$

where Eq. (4.1.1) was used. The responsivity $R$ is thus given by

$$R = \frac{\eta q}{h\nu} \approx \frac{\eta \lambda}{1.24}, \quad (4.1.3)$$

where $\lambda \equiv c/\nu$ is expressed in micrometers. The responsivity of a photodetector increases with the wavelength $\lambda$ simply because more photons are present for the same optical power. Such a linear dependence on $\lambda$ is not expected to continue forever because eventually the photon energy becomes too small to generate electrons. In semiconductors, this happens for $h\nu < E_g$, where $E_g$ is the bandgap. The quantum efficiency $\eta$ then drops to zero.

The dependence of $\eta$ on $\lambda$ enters through the absorption coefficient $\alpha$. If the facets of the semiconductor slab in Fig. 4.1 are assumed to have an antireflection coating, the power transmitted through the slab of width $W$ is $P_{\text{tr}} = \exp(-\alpha W) P_{\text{in}}$. The absorbed power can be written as

$$P_{\text{abs}} = P_{\text{in}} - P_{\text{tr}} = [1 - \exp(-\alpha W)] P_{\text{in}}. \quad (4.1.4)$$

Since each absorbed photon creates an electron–hole pair, the quantum efficiency $\eta$ is given by

$$\eta = P_{\text{abs}}/P_{\text{in}} = 1 - \exp(-\alpha W). \quad (4.1.5)$$
As expected, $\eta$ becomes zero when $\alpha = 0$. On the other hand, $\eta$ approaches 1 if $\alpha W \gg 1$.

Figure 4.2 shows the wavelength dependence of $\alpha$ for several semiconductor materials commonly used to make photodetectors for lightwave systems. The wavelength $\lambda_c$ at which $\alpha$ becomes zero is called the cutoff wavelength, as that material can be used for a photodetector only for $\lambda < \lambda_c$. As seen in Fig. 4.2, indirect-bandgap semiconductors such as Si and Ge can be used to make photodetectors even though the absorption edge is not as sharp as for direct-bandgap materials. Large values of $\alpha (\sim 10^4 \text{ cm}^{-1})$ can be realized for most semiconductors, and $\eta$ can approach 100% for $W \sim 10 \mu\text{m}$. This feature illustrates the efficiency of semiconductors for the purpose of photodetection.

### 4.1.2 Rise Time and Bandwidth

The bandwidth of a photodetector is determined by the speed with which it responds to variations in the incident optical power. It is useful to introduce the concept of *rise time* $T_r$, defined as the time over which the current builds up from 10 to 90% of its final value when the incident optical power is changed abruptly. Clearly, $T_r$ will depend on
the time taken by electrons and holes to travel to the electrical contacts. It also depends on the response time of the electrical circuit used to process the photocurrent.

The rise time $T_r$ of a linear electrical circuit is defined as the time during which the response increases from 10 to 90% of its final output value when the input is changed abruptly (a step function). When the input voltage across an $RC$ circuit changes instantaneously from 0 to $V_0$, the output voltage changes as

$$V_{out}(t) = V_0[1 - \exp(-t/RC)],$$

where $R$ is the resistance and $C$ is the capacitance of the $RC$ circuit. The rise time is found to be given by

$$T_r = (\ln 9)RC \approx 2.2\tau_{RC},$$

where $\tau_{RC} = RC$ is the time constant of the $RC$ circuit.

The rise time of a photodetector can be written by extending Eq. (4.1.7) as

$$T_r = (\ln 9)(\tau_{tr} + \tau_{RC}),$$

where $\tau_{tr}$ is the transit time and $\tau_{RC}$ is the time constant of the equivalent $RC$ circuit. The transit time is added to $\tau_{RC}$ because it takes some time before the carriers are collected after their generation through absorption of photons. The maximum collection time is just equal to the time an electron takes to traverse the absorption region. Clearly, $\tau_{tr}$ can be reduced by decreasing $W$. However, as seen from Eq. (4.1.5), the quantum efficiency $\eta$ begins to decrease significantly for $\alpha W < 3$. Thus, there is a trade-off between the bandwidth and the responsivity (speed versus sensitivity) of a photodetector. Often, the $RC$ time constant $\tau_{RC}$ limits the bandwidth because of electrical parasitics. The numerical values of $\tau_{tr}$ and $\tau_{RC}$ depend on the detector design and can vary over a wide range.

The bandwidth of a photodetector is defined in a manner analogous to that of a $RC$ circuit and is given by

$$\Delta f = [2\pi(\tau_{tr} + \tau_{RC})]^{-1}.$$  

(4.1.9)

As an example, when $\tau_{tr} = \tau_{RC} = 100$ ps, the bandwidth of the photodetector is below 1 GHz. Clearly, both $\tau_{tr}$ and $\tau_{RC}$ should be reduced below 10 ps for photodetectors needed for lightwave systems operating at bit rates of 10 Gb/s or more.

Together with the bandwidth and the responsivity, the dark current $I_d$ of a photodetector is the third important parameter. Here, $I_d$ is the current generated in a photodetector in the absence of any optical signal and originates from stray light or from thermally generated electron–hole pairs. For a good photodetector, the dark current should be negligible ($I_d < 10$ nA).

### 4.2 Common Photodetectors

The semiconductor slab of Fig. 4.1 is useful for illustrating the basic concepts but such a simple device is rarely used in practice. This section focuses on reverse-biased $p$–$n$ junctions that are commonly used for making optical receivers. Metal–semiconductor–metal (MSM) photodetectors are also discussed briefly.
4.2. COMMON PHOTODETECTORS

4.2.1 p–n Photodiodes

A reverse-biased p–n junction consists of a region, known as the depletion region, that is essentially devoid of free charge carriers and where a large built-in electric field opposes flow of electrons from the n-side to the p-side (and of holes from p to n). When such a p–n junction is illuminated with light on one side, say the p-side (see Fig. 4.3), electron–hole pairs are created through absorption. Because of the large built-in electric field, electrons and holes generated inside the depletion region accelerate in opposite directions and drift to the n- and p-sides, respectively. The resulting flow of current is proportional to the incident optical power. Thus a reverse-biased p–n junction acts as a photodetector and is referred to as the p–n photodiode.

Figure 4.3(a) shows the structure of a p–n photodiode. As shown in Fig. 4.3(b), optical power decreases exponentially as the incident light is absorbed inside the depletion region. The electron–hole pairs generated inside the depletion region experience a large electric field and drift rapidly toward the p- or n-side, depending on the electric charge [Fig. 4.3(c)]. The resulting current flow constitutes the photodiode response to the incident optical power in accordance with Eq. (4.1.1). The responsivity of a photodiode is quite high (\(R \sim 1 \text{ A/W}\)) because of a high quantum efficiency.

The bandwidth of a p–n photodiode is often limited by the transit time \(\tau_{tr}\) in Eq. (4.1.9). If \(W\) is the width of the depletion region and \(v_d\) is the drift velocity, the transit time is given by

\[
\tau_{tr} = W / v_d. \tag{4.2.1}
\]

Typically, \(W \sim 10 \mu\text{m}, v_d \sim 10^5 \text{ m/s}\), and \(\tau_{tr} \sim 100 \text{ ps}\). Both \(W\) and \(v_d\) can be optimized to minimize \(\tau_{tr}\). The depletion-layer width depends on the acceptor and donor concentrations and can be controlled through them. The velocity \(v_d\) depends on the applied voltage but attains a maximum value (called the saturation velocity) \(\sim 10^5 \text{ m/s}\) that depends on the material used for the photodiode. The \(RC\) time constant \(\tau_{RC}\) can be
written as
\[ \tau_{RC} = (R_L + R_s)C_p, \] (4.2.2)
where \( R_L \) is the external load resistance, \( R_s \) is the internal series resistance, and \( C_p \) is the parasitic capacitance. Typically, \( \tau_{RC} \sim 100 \text{ ps} \), although lower values are possible with a proper design. Indeed, modern \( p-n \) photodiodes are capable of operating at bit rates of up to 40 Gb/s.

The limiting factor for the bandwidth of \( p-n \) photodiodes is the presence of a diffusive component in the photocurrent. The physical origin of the diffusive component is related to the absorption of incident light outside the depletion region. Electrons generated in the \( p \)-region have to diffuse to the depletion-region boundary before they can drift to the \( n \)-side; similarly, holes generated in the \( n \)-region must diffuse to the depletion-region boundary. Diffusion is an inherently slow process; carriers take a nanosecond or longer to diffuse over a distance of about 1 \( \mu \text{m} \). Figure 4.4 shows how the presence of a diffusive component can distort the temporal response of a photodiode. The diffusion contribution can be reduced by decreasing the widths of the \( p \)- and \( n \)-regions and increasing the depletion-region width so that most of the incident optical power is absorbed inside it. This is the approach adopted for \( p-i-n \) photodiodes, discussed next.

### 4.2.2 \( p-i-n \) Photodiodes

A simple way to increase the depletion-region width is to insert a layer of undoped (or lightly doped) semiconductor material between the \( p-n \) junction. Since the middle
Figure 4.5: (a) A p–i–n photodiode together with the electric-field distribution under reverse bias; (b) design of an InGaAs p–i–n photodiode.

layer consists of nearly intrinsic material, such a structure is referred to as the p–i–n photodiode. Figure 4.5(a) shows the device structure together with the electric-field distribution inside it under reverse-bias operation. Because of its intrinsic nature, the middle i-layer offers a high resistance, and most of the voltage drop occurs across it. As a result, a large electric field exists in the i-layer. In essence, the depletion region extends throughout the i-region, and its width $W$ can be controlled by changing the middle-layer thickness. The main difference from the p–n photodiode is that the drift component of the photocurrent dominates over the diffusion component simply because most of the incident power is absorbed inside the i-region of a p–i–n photodiode.

Since the depletion width $W$ can be tailored in p–i–n photodiodes, a natural question is how large $W$ should be. As discussed in Section 4.1, the optimum value of $W$ depends on a compromise between speed and sensitivity. The responsivity can be increased by increasing $W$ so that the quantum efficiency $\eta$ approaches 100% [see Eq. (4.1.5)]. However, the response time also increases, as it takes longer for carriers to drift across the depletion region. For indirect-bandgap semiconductors such as Si and Ge, typically $W$ must be in the range 20–50 $\mu$m to ensure a reasonable quantum efficiency. The bandwidth of such photodiodes is then limited by a relatively long transit time ($\tau_{tr} > 200$ ps). By contrast, $W$ can be as small as 3–5 $\mu$m for photodiodes that use direct-bandgap semiconductors, such as InGaAs. The transit time for such photodiodes is $\tau_{tr} \sim 10$ ps. Such values of $\tau_{tr}$ correspond to a detector bandwidth $\Delta f \sim 10$ GHz if we use Eq. (4.1.9) with $\tau_{tr} \gg \tau_{RC}$.

The performance of p–i–n photodiodes can be improved considerably by using a double-heterostructure design. Similar to the case of semiconductor lasers, the middle i-type layer is sandwiched between the p-type and n-type layers of a different semiconductor whose bandgap is chosen such that light is absorbed only in the middle i-layer. A p–i–n photodiode commonly used for lightwave applications uses InGaAs for the middle layer and InP for the surrounding p-type and n-type layers [10]. Figure 4.5(b)
Table 4.1 Characteristics of common p–i–n photodiodes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Unit</th>
<th>Si</th>
<th>Ge</th>
<th>InGaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>$\lambda$</td>
<td>$\mu$m</td>
<td>0.4–1.1</td>
<td>0.8–1.8</td>
<td>1.0–1.7</td>
</tr>
<tr>
<td>Responsivity</td>
<td>$R$</td>
<td>A/W</td>
<td>0.4–0.6</td>
<td>0.5–0.7</td>
<td>0.6–0.9</td>
</tr>
<tr>
<td>Quantum efficiency</td>
<td>$\eta$</td>
<td>%</td>
<td>75–90</td>
<td>50–55</td>
<td>60–70</td>
</tr>
<tr>
<td>Dark current</td>
<td>$I_d$</td>
<td>nA</td>
<td>1–10</td>
<td>50–500</td>
<td>1–20</td>
</tr>
<tr>
<td>Rise time</td>
<td>$T_r$</td>
<td>ns</td>
<td>0.5–1</td>
<td>0.1–0.5</td>
<td>0.02–0.5</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>$\Delta f$</td>
<td>GHz</td>
<td>0.3–0.6</td>
<td>0.5–3</td>
<td>1–10</td>
</tr>
<tr>
<td>Bias voltage</td>
<td>$V_b$</td>
<td>V</td>
<td>50–100</td>
<td>6–10</td>
<td>5–6</td>
</tr>
</tbody>
</table>

shows such an InGaAs p–i–n photodiode. Since the bandgap of InP is 1.35 eV, InP is transparent for light whose wavelength exceeds 0.92 $\mu$m. By contrast, the bandgap of lattice-matched In$_{1-x}$Ga$_x$As material with $x = 0.47$ is about 0.75 eV (see Section 3.1.4), a value that corresponds to a cutoff wavelength of 1.65 $\mu$m. The middle InGaAs layer thus absorbs strongly in the wavelength region 1.3–1.6 $\mu$m. The diffusive component of the detector current is eliminated completely in such a heterostructure photodiode simply because photons are absorbed only inside the depletion region. The front facet is often coated using suitable dielectric layers to minimize reflections. The quantum efficiency $\eta$ can be made almost 100% by using an InGaAs layer 4–5 $\mu$m thick. InGaAs photodiodes are quite useful for lightwave systems and are often used in practice. Table 4.1 lists the operating characteristics of three common p–i–n photodiodes.

Considerable effort was directed during the 1990s toward developing high-speed p–i–n photodiodes capable of operating at bit rates exceeding 10 Gb/s [10]–[20]. Bandwidths of up to 70 GHz were realized as early as 1986 by using a thin absorption layer (< 1 $\mu$m) and by reducing the parasitic capacitance $C_p$ with a small size, but only at the expense of a lower quantum efficiency and responsivity [10]. By 1995, p–i–n photodiodes exhibited a bandwidth of 110 GHz for devices designed to reduce $\tau_{RC}$ to near 1 ps [15].

Several techniques have been developed to improve the efficiency of high-speed photodiodes. In one approach, a Fabry–Perot (FP) cavity is formed around the p–i–n structure to enhance the quantum efficiency [11]–[14], resulting in a laserlike structure. As discussed in Section 3.3.2, a FP cavity has a set of longitudinal modes at which the internal optical field is resonantly enhanced through constructive interference. As a result, when the incident wavelength is close to a longitudinal mode, such a photodiode exhibits high sensitivity. The wavelength selectivity can even be used to advantage in wavelength-division multiplexing (WDM) applications. A nearly 100% quantum efficiency was realized in a photodiode in which one mirror of the FP cavity was formed by using the Bragg reflectivity of a stack of AlGaAs/AlAs layers [12]. This approach was extended to InGaAs photodiodes by inserting a 90-nm-thick InGaAs absorbing layer into a microcavity composed of a GaAs/AlAs Bragg mirror and a dielectric mirror. The device exhibited 94% quantum efficiency at the cavity resonance with a bandwidth of 14 nm [13]. By using an air-bridged metal waveguide together with an undercut mesa
4.2. COMMON PHOTODETECTORS

Figure 4.6: (a) Schematic cross section of a mushroom-mesa waveguide photodiode and (b) its measured frequency response. (After Ref. [17]; ©1994 IEEE; reprinted with permission.)

structure, a bandwidth of 120 GHz has been realized [14]. The use of such a structure within a FP cavity should provide a p–i–n photodiode with a high bandwidth and high efficiency.

Another approach to realize efficient high-speed photodiodes makes use of an optical waveguide into which the optical signal is edge coupled [16]–[20]. Such a structure resembles an unpumped semiconductor laser except that various epitaxial layers are optimized differently. In contrast with a semiconductor laser, the waveguide can be made wide to support multiple transverse modes in order to improve the coupling efficiency [16]. Since absorption takes place along the length of the optical waveguide (≈ 10 µm), the quantum efficiency can be nearly 100% even for an ultrathin absorption layer. The bandwidth of such waveguide photodiodes is limited by $\tau_{RC}$ in Eq. (4.1.9), which can be decreased by controlling the waveguide cross-section-area. Indeed, a 50-GHz bandwidth was realized in 1992 for a waveguide photodiode [16].

The bandwidth of waveguide photodiodes can be increased to 110 GHz by adopting a mushroom-mesa waveguide structure [17]. Such a device is shown schematically in Fig. 4.6. In this structure, the width of the i-type absorbing layer was reduced to 1.5 µm while the p- and n-type cladding layers were made 6 µm wide. In this way, both the parasitic capacitance and the internal series resistance were minimized, reducing $\tau_{RC}$ to about 1 ps. The frequency response of such a device at the 1.55-µm wavelength is also shown in Fig. 4.6. It was measured by using a spectrum analyzer (circles) as well as taking the Fourier transform of the short-pulse response (solid curve). Clearly, waveguide p–i–n photodiodes can provide both a high responsivity and a large bandwidth. Waveguide photodiodes have been used for 40-Gb/s optical receivers [19] and have the potential for operating at bit rates as high as 100 Gb/s [20].

The performance of waveguide photodiodes can be improved further by adopting an electrode structure designed to support traveling electrical waves with matching impedance to avoid reflections. Such photodiodes are called traveling-wave photodetectors. In a GaAs-based implementation of this idea, a bandwidth of 172 GHz with 45% quantum efficiency was realized in a traveling-wave photodetector designed with a 1-µm-wide waveguide [21]. By 2000, such an InP/InGaAs photodetector exhibited a bandwidth of 310 GHz in the 1.55-µm spectral region [22].
4.2.3 Avalanche Photodiodes

All detectors require a certain minimum current to operate reliably. The current requirement translates into a minimum power requirement through $P_{in} = I_p/R$. Detectors with a large responsivity $R$ are preferred since they require less optical power. The responsivity of $p$–$i$–$n$ photodiodes is limited by Eq. (4.1.3) and takes its maximum value $R = q/h\nu$ for $\eta = 1$. Avalanche photodiode (APDs) can have much larger values of $R$, as they are designed to provide an internal current gain in a way similar to photomultiplier tubes. They are used when the amount of optical power that can be spared for the receiver is limited.

The physical phenomenon behind the internal current gain is known as the impact ionization [23]. Under certain conditions, an accelerating electron can acquire sufficient energy to generate a new electron–hole pair. In the band picture (see Fig. 3.2) the energetic electron gives a part of its kinetic energy to another electron in the valence band that ends up in the conduction band, leaving behind a hole. The net result of impact ionization is that a single primary electron, generated through absorption of a photon, creates many secondary electrons and holes, all of which contribute to the photodiode current. Of course, the primary hole can also generate secondary electron–hole pairs that contribute to the current. The generation rate is governed by two parameters, $\alpha_e$ and $\alpha_h$, the impact-ionization coefficients of electrons and holes, respectively. Their numerical values depend on the semiconductor material and on the electric field.

![Figure 4.7: Impact-ionization coefficients of several semiconductors as a function of the electric field for electrons (solid line) and holes (dashed line). (After Ref. [24]; ©1977 Elsevier; reprinted with permission.)](image-url)
that accelerates electrons and holes. Figure 4.7 shows $\alpha_e$ and $\alpha_h$ for several semiconductors [24]. Values $\sim 1 \times 10^4 \text{ cm}^{-1}$ are obtained for electric fields in the range 2–4×$10^5 \text{ V/cm}$. Such large fields can be realized by applying a high voltage ($\sim 100 \text{ V}$) to the APD.

APDs differ in their design from that of $p$–$i$–$n$ photodiodes mainly in one respect: an additional layer is added in which secondary electron–hole pairs are generated through impact ionization. Figure 4.8(a) shows the APD structure together with the variation of electric field in various layers. Under reverse bias, a high electric field exists in the $p$-type layer sandwiched between $i$-type and $n^+$-type layers. This layer is referred to as the multiplication layer, since secondary electron–hole pairs are generated here through impact ionization. The $i$-layer still acts as the depletion region in which most of the incident photons are absorbed and primary electron–hole pairs are generated. Electrons generated in the $i$-region cross the gain region and generate secondary electron–hole pairs responsible for the current gain.

The current gain for APDs can be calculated by using the two rate equations governing current flow within the multiplication layer [23]:

$$\frac{d i_e}{dx} = \alpha_e i_e + \alpha_h i_h,$$  \hspace{1cm} (4.2.3)

$$\frac{d i_h}{dx} = \alpha_e i_e + \alpha_h i_h,$$  \hspace{1cm} (4.2.4)

where $i_e$ is the electron current and $i_h$ is the hole current. The minus sign in Eq. (4.2.4) is due to the opposite direction of the hole current. The total current,

$$I = i_e(x) + i_h(x),$$  \hspace{1cm} (4.2.5)
remains constant at every point inside the multiplication region. If we replace $i_h$ in Eq. (4.2.3) by $I - i_e$, we obtain

$$\frac{di_e}{dx} = (\alpha_e - \alpha_h)i_e + \alpha_hI.$$  (4.2.6)

In general, $\alpha_e$ and $\alpha_h$ are $x$ dependent if the electric field across the gain region is nonuniform. The analysis is considerably simplified if we assume a uniform electric field and treat $\alpha_e$ and $\alpha_h$ as constants. We also assume that $\alpha_e > \alpha_h$. The avalanche process is initiated by electrons that enter the gain region of thickness $d$ at $x = 0$. By using the condition $i_h(d) = 0$ (only electrons cross the boundary to enter the $n$-region), the boundary condition for Eq. (4.2.6) is $i_e(d) = I$. By integrating this equation, the multiplication factor defined as

$$M = \frac{i_e(d)}{i_e(0)}$$

is given by

$$M = \frac{1 - k_A}{\exp[-(1 - k_A)\alpha_e d] - k_A},$$  (4.2.7)

where $k_A = \alpha_h / \alpha_e$. The APD gain is quite sensitive to the ratio of the impact-ionization coefficients. When $\alpha_h = 0$ so that only electrons participate in the avalanche process, $M = \exp(\alpha_e d)$, and the APD gain increases exponentially with $d$. On the other hand, when $\alpha_h = \alpha_e$, so that $k_A = 1$ in Eq. (4.2.7), $M = (1 - \alpha_e d)^{-1}$. The APD gain then becomes infinite for $\alpha_e d = 1$, a condition known as the avalanche breakdown. Although higher APD gain can be realized with a smaller gain region when $\alpha_e$ and $\alpha_h$ are comparable, the performance is better in practice for APDs in which either $\alpha_e \gg \alpha_h$ or $\alpha_h \gg \alpha_e$ so that the avalanche process is dominated by only one type of charge carrier. The reason behind this requirement is discussed in Section 4.4, where issues related to the receiver noise are considered.

Because of the current gain, the responsivity of an APD is enhanced by the multiplication factor $M$ and is given by

$$R_{\text{APD}} = MR = M(\eta q / h\nu),$$  (4.2.8)

where Eq. (4.1.3) was used. It should be mentioned that the avalanche process in APDs is intrinsically noisy and results in a gain factor that fluctuates around an average value. The quantity $M$ in Eq. (4.2.8) refers to the average APD gain. The noise characteristics of APDs are considered in Section 4.4.

The intrinsic bandwidth of an APD depends on the multiplication factor $M$. This is easily understood by noting that the transit time $\tau_{tr}$ for an APD is no longer given by Eq. (4.2.1) but increases considerably simply because generation and collection of secondary electron–hole pairs take additional time. The APD gain decreases at high frequencies because of such an increase in the transit time and limits the bandwidth. The decrease in $M(\omega)$ can be written as [24]

$$M(\omega) = M_0[1 + (\omega \tau_e M_0)^2]^{-1/2},$$  (4.2.9)

where $M_0 = M(0)$ is the low-frequency gain and $\tau_e$ is the effective transit time that depends on the ionization coefficient ratio $k_A = \alpha_h / \alpha_e$. For the case $\alpha_h < \alpha_e$, $\tau_e = c_A k_A \tau_0$, where $c_A$ is a constant ($c_A \sim 1$). Assuming that $\tau_{RC} \ll \tau_e$, the APD bandwidth is given approximately by $\Delta f = (2\pi \tau_e M_0)^{-1}$. This relation shows the trade-off between
Table 4.2 Characteristics of common APDs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Unit</th>
<th>Si</th>
<th>Ge</th>
<th>InGaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>$\lambda$</td>
<td>$\mu$m</td>
<td>0.4–1.1</td>
<td>0.8–1.8</td>
<td>1.0–1.7</td>
</tr>
<tr>
<td>Responsivity</td>
<td>$R_{APD}$</td>
<td>A/W</td>
<td>80–130</td>
<td>3–30</td>
<td>5–20</td>
</tr>
<tr>
<td>APD gain</td>
<td>$M$</td>
<td>—</td>
<td>100–500</td>
<td>50–200</td>
<td>10–40</td>
</tr>
<tr>
<td>k-factor</td>
<td>$k_A$</td>
<td>—</td>
<td>0.02–0.05</td>
<td>0.7–1.0</td>
<td>0.5–0.7</td>
</tr>
<tr>
<td>Dark current</td>
<td>$I_d$</td>
<td>nA</td>
<td>0.1–1</td>
<td>50–500</td>
<td>1–5</td>
</tr>
<tr>
<td>Rise time</td>
<td>$T_r$</td>
<td>ns</td>
<td>0.1–2</td>
<td>0.5–0.8</td>
<td>0.1–0.5</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>$\Delta f$</td>
<td>GHz</td>
<td>0.2–1</td>
<td>0.4–0.7</td>
<td>1–10</td>
</tr>
<tr>
<td>Bias voltage</td>
<td>$V_b$</td>
<td>V</td>
<td>200–250</td>
<td>20–40</td>
<td>20–30</td>
</tr>
</tbody>
</table>

the APD gain $M_0$ and the bandwidth $\Delta f$ (speed versus sensitivity). It also shows the advantage of using a semiconductor material for which $k_A \ll 1$.

Table 4.2 compares the operating characteristics of Si, Ge, and InGaAs APDs. As $k_A \ll 1$ for Si, silicon APDs can be designed to provide high performance and are useful for lightwave systems operating near 0.8 $\mu$m at bit rates $\sim$100 Mb/s. A particularly useful design, shown in Fig. 4.8(b), is known as reach-through APD because the depletion layer reaches to the contact layer through the absorption and multiplication regions. It can provide high gain ($M \approx 100$) with low noise and a relatively large bandwidth. For lightwave systems operating in the wavelength range 1.3–1.6 $\mu$m, Ge or InGaAs APDs must be used. The improvement in sensitivity for such APDs is limited to a factor below 10 because of a relatively low APD gain ($M \sim 10$) that must be used to reduce the noise (see Section 4.4.3).

The performance of InGaAs APDs can be improved through suitable design modifications to the basic APD structure shown in Fig. 4.8. The main reason for a relatively poor performance of InGaAs APDs is related to the comparable numerical values of the impact-ionization coefficients $\alpha_e$ and $\alpha_h$ (see Fig. 4.7). As a result, the bandwidth is considerably reduced, and the noise is also relatively high (see Section 4.4). Furthermore, because of a relatively narrow bandgap, InGaAs undergoes tunneling breakdown at electric fields of about $1 \times 10^5$ V/cm, a value that is below the threshold for avalanche multiplication. This problem can be solved in heterostructure APDs by using an InP layer for the gain region because quite high electric fields ($> 5 \times 10^5$ V/cm) can exist in InP without tunneling breakdown. Since the absorption region (i-type InGaAs layer) and the multiplication region (n-type InP layer) are separate in such a device, this structure is known as SAM, where SAM stands for separate absorption and multiplication regions. As $\alpha_h > \alpha_e$ for InP (see Fig. 4.7), the APD is designed such that holes initiate the avalanche process in an n-type InP layer, and $k_A$ is defined as $k_A = \alpha_e/\alpha_h$. Figure 4.9(a) shows a mesa-type SAM APD structure.

One problem with the SAM APD is related to the large bandgap difference between InP ($E_g = 1.35$ eV) and InGaAs ($E_g = 0.75$ eV). Because of a valence-band step of about 0.4 eV, holes generated in the InGaAs layer are trapped at the heterojunction interface and are considerably slowed before they reach the multiplication region (InP layer). Such an APD has an extremely slow response and a relatively small bandwidth.
The problem can be solved by using another layer between the absorption and multiplication regions whose bandgap is intermediate to those of InP and InGaAs layers. The quaternary material InGaAsP, the same material used for semiconductor lasers, can be tailored to have a bandgap anywhere in the range 0.75–1.35 eV and is ideal for this purpose. It is even possible to grade the composition of InGaAsP over a region of 10–100 nm thickness. Such APDs are called SAGM APDs, where SAGM indicates separate absorption, grading, and multiplication regions [25]. Figure 4.9(b) shows the design of an InGaAs APD with the SAGM structure. The use of an InGaAsP grading layer improves the bandwidth considerably. As early as 1987, a SAGM APD exhibited a gain–bandwidth product $M \Delta f = 70$ GHz for $M > 12$ [26]. This value was increased to 100 GHz in 1991 by using a charge region between the grading and multiplication regions [27]. In such SAGCM APDs, the InP multiplication layer is undoped, while the InP charge layer is heavily $n$-doped. Holes accelerate in the charge layer because of a strong electric field, but the generation of secondary electron–hole pairs takes place in the undoped InP layer. SAGCM APDs improved considerably during the 1990s [28]–[32]. A gain–bandwidth product of 140 GHz was realized in 2000 using a 0.1-$\mu$m-thick multiplication layer that required $<20$ V across it [32]. Such APDs are quite suitable for making a compact 10-Gb/s APD receiver.

A different approach to the design of high-performance APDs makes use of a superlattice structure [33]–[38]. The major limitation of InGaAs APDs results from comparable values of $\alpha_e$ and $\alpha_h$. A superlattice design offers the possibility of reducing the ratio $k_A = \alpha_h / \alpha_e$ from its standard value of nearly unity. In one scheme, the absorption and multiplication regions alternate and consist of thin layers (~10 nm) of semiconductor materials with different bandgaps. This approach was first demonstrated for GaAs/AlGaAs multiquantum-well (MQW) APDs and resulted in a considerable enhancement of the impact-ionization coefficient for electrons [33]. Its use is less successful for the InGaAs/InP material system. Nonetheless, considerable progress has been made through the so-called staircase APDs, in which the InGaAsP layer is compositionally graded to form a sawtooth kind of structure in the energy-band diagram that looks like a staircase under reverse bias. Another scheme for making high-speed
4.2. COMMON PHOTODETECTORS

Figure 4.10: (a) Device structure and (b) measured 3-dB bandwidth as a function of $M$ for a superlattice APD. (After Ref. [38]; ©2000 IEEE; reprinted with permission.)

APDs uses alternate layers of InP and InGaAs for the grading region [33]. However, the ratio of the widths of the InP to InGaAs layers varies from zero near the absorbing region to almost infinity near the multiplication region. Since the effective bandgap of a quantum well depends on the quantum-well width (InGaAs layer thickness), a graded “pseudo-quaternary” compound is formed as a result of variation in the layer thickness.

The most successful design for InGaAs APDs uses a superlattice structure for the multiplication region of a SAM APD. A superlattice consists of a periodic structure such that each period is made using two ultrathin ($\sim10$-nm) layers with different bandgaps. In the case of 1.55-µm APDs, alternate layers of InAlGaAs and InAlAs are used, the latter acting as a barrier layer. An InP field-buffer layer often separates the InGaAs absorption region from the superlattice multiplication region. The thickness of this buffer layer is quite critical for the APD performance. For a 52-nm-thick field-buffer layer, the gain–bandwidth product was limited to $M\Delta f = 120$ GHz [34] but increased to 150 GHz when the thickness was reduced to 33.4 nm [37]. These early devices used a mesa structure. During the late 1990s, a planar structure was developed for improving the device reliability [38]. Figure 4.10 shows such a device schematically together with its 3-dB bandwidth measured as a function of the APD gain. The gain–bandwidth product of 110 GHz is large enough for making APDs operating at 10 Gb/s. Indeed, such an APD receiver was used for a 10-Gb/s lightwave system with excellent performance.

The gain–bandwidth limitation of InGaAs APDs results primarily from using the InP material system for the generation of secondary electron–hole pairs. A hybrid approach in which a Si multiplication layer is incorporated next to an InGaAs absorption layer may be useful provided the heterointerface problems can be overcome. In a 1997 experiment, a gain-bandwidth product of more than 300 GHz was realized by using such a hybrid approach [39]. The APD exhibited a 3-dB bandwidth of over 9 GHz for values of $M$ as high as 35 while maintaining a 60% quantum efficiency.

Most APDs use an absorbing layer thick enough (about 1 µm) that the quantum efficiency exceeds 50%. The thickness of the absorbing layer affects the transit time $\tau_t$ and the bias voltage $V_b$. In fact, both of them can be reduced significantly by using a thin absorbing layer ($\sim0.1$ µm), resulting in improved APDs provided that a high
quantum efficiency can be maintained. Two approaches have been used to meet these somewhat conflicting design requirements. In one design, a FP cavity is formed to enhance the absorption within a thin layer through multiple round trips. An external quantum efficiency of $\sim 70\%$ and a gain–bandwidth product of 270 GHz were realized in such a 1.55-\(\mu\)m APD using a 60-nm-thick absorbing layer with a 200-nm-thick multiplication layer [40]. In another approach, an optical waveguide is used into which the incident light is edge coupled [41]. Both of these approaches reduce the bias voltage to near 10 V, maintain high efficiency, and reduce the transit time to $\sim 1$ ps. Such APDs are suitable for making 10-Gb/s optical receivers.

4.2.4 MSM Photodetectors

In metal–semiconductor–metal (MSM) photodetectors, a semiconductor absorbing layer is sandwiched between two metals, forming a Schottky barrier at each metal–semiconductor interface that prevents flow of electrons from the metal to the semiconductor. Similar to a \(p-i-n\) photodiode, electron–hole pairs generated through photoabsorption flow toward the metal contacts, resulting in a photocurrent that is a measure of the incident optical power, as indicated in Eq. (4.1.1). For practical reasons, the two metal contacts are made on the same (top) side of the epitaxially grown absorbing layer by using an \textit{interdigitated} electrode structure with a finger spacing of about 1 \(\mu\)m [42]. This scheme results in a planar structure with an inherently low parasitic capacitance that allows high-speed operation (up to 300 GHz) of MSM photodetectors. If the light is incident from the electrode side, the responsivity of a MSM photodetector is reduced because of its blockage by the opaque electrodes. This problem can be solved by back illumination if the substrate is transparent to the incident light.

GaAs-based MSM photodetectors were developed throughout the 1980s and exhibit excellent operating characteristics [42]. The development of InGaAs-based MSM photodetectors, suitable for lightwave systems operating in the range 1.3–1.6 \(\mu\)m, started in the late 1980s, with most progress made during the 1990s [43]–[52]. The major problem with InGaAs is its relatively low \textit{Schottky-barrier height} (about 0.2 eV). This problem was solved by introducing a thin layer of InP or InAlAs between the InGaAs layer and the metal contact. Such a layer, called the \textit{barrier-enhancement layer}, improves the performance of InGaAs MSM photodetectors drastically. The use of a 20-nm-thick InAlAs barrier-enhancement layer resulted in 1992 in 1.3-\(\mu\)m MSM photodetectors exhibiting 92\% quantum efficiency (through back illumination) with a low dark current [44]. A packaged device had a bandwidth of 4 GHz despite a large 150 \(\mu\)m diameter. If top illumination is desirable for processing or packaging reasons, the responsivity can be enhanced by using semitransparent metal contacts. In one experiment, the responsivity at 1.55 \(\mu\)m increased from 0.4 to 0.7 A/W when the thickness of gold contact was reduced from 100 to 10 nm [45]. In another approach, the structure is separated from the host substrate and bonded to a silicon substrate with the interdigitated contact on bottom. Such an “inverted” MSM photodetector then exhibits high responsivity when illuminated from the top [46].

The temporal response of MSM photodetectors is generally different under back and top illuminations [47]. In particular, the bandwidth $\Delta f$ is larger by about a factor of 2 for top illumination, although the responsivity is reduced because of metal shad-
4.3. Receiver Design

The design of an optical receiver depends on the modulation format used by the transmitter. Since most lightwave systems employ the binary intensity modulation, we focus in this chapter on digital optical receivers. Figure 4.11 shows a block diagram of such a receiver. Its components can be arranged into three groups—the front end, the linear channel, and the decision circuit.

4.3.1 Front End

The front end of a receiver consists of a photodiode followed by a preamplifier. The optical signal is coupled onto the photodiode by using a coupling scheme similar to that used for optical transmitters (see Section 3.4.1); butt coupling is often used in practice. The photodiode converts the optical bit stream into an electrical time-varying signal. The role of the preamplifier is to amplify the electrical signal for further processing.

The design of the front end requires a trade-off between speed and sensitivity. Since the input voltage to the preamplifier can be increased by using a large load resistor $R_L$, a high-impedance front end is often used [see Fig. 4.12(a)]. Furthermore, as discussed in Section 4.4, a large $R_L$ reduces the thermal noise and improves the receiver sensitivity. The main drawback of high-impedance front end is its low bandwidth given by $\Delta f = \left(\frac{2\pi R_L C_T}{R_s R_L + C_T}\right)^{-1}$, where $R_s \ll R_L$ is assumed in Eq. (4.2.2) and $C_T = C_p + C_A$ is the total capacitance, which includes the contributions from the photodiode ($C_p$) and the transistor used for amplification ($C_A$). The receiver bandwidth is limited by its slowest

![Figure 4.11: Diagram of a digital optical receiver showing various components. Vertical dashed lines group receiver components into three sections.](image-url)
component. A high-impedance front end cannot be used if $\Delta f$ is considerably less than the bit rate. An equalizer is sometimes used to increase the bandwidth. The equalizer acts as a filter that attenuates low-frequency components of the signal more than the high-frequency components, thereby effectively increasing the front-end bandwidth. If the receiver sensitivity is not of concern, one can simply decrease $R_L$ to increase the bandwidth, resulting in a low-impedance front end.

Transimpedance front ends provide a configuration that has high sensitivity together with a large bandwidth. Its dynamic range is also improved compared with high-impedance front ends. As seen in Fig. 4.12(b), the load resistor is connected as a feedback resistor around an inverting amplifier. Even though $R_L$ is large, the negative feedback reduces the effective input impedance by a factor of $G$, where $G$ is the amplifier gain. The bandwidth is thus enhanced by a factor of $G$ compared with high-impedance front ends. Transimpedance front ends are often used in optical receivers because of their improved characteristics. A major design issue is related to the stability of the feedback loop. More details can be found in Refs. [5]–[9].

### 4.3.2 Linear Channel

The linear channel in optical receivers consists of a high-gain amplifier (the main amplifier) and a low-pass filter. An equalizer is sometimes included just before the amplifier to correct for the limited bandwidth of the front end. The amplifier gain is controlled automatically to limit the average output voltage to a fixed level irrespective of the incident average optical power at the receiver. The low-pass filter shapes the voltage pulse. Its purpose is to reduce the noise without introducing much intersymbol
4.3. RECEIVER DESIGN

interference (ISI). As discussed in Section 4.4, the receiver noise is proportional to the receiver bandwidth and can be reduced by using a low-pass filter whose bandwidth $\Delta f$ is smaller than the bit rate. Since other components of the receiver are designed to have a bandwidth larger than the filter bandwidth, the receiver bandwidth is determined by the low-pass filter used in the linear channel. For $\Delta f < B$, the electrical pulse spreads beyond the allocated bit slot. Such a spreading can interfere with the detection of neighboring bits, a phenomenon referred to as ISI.

It is possible to design a low-pass filter in such a way that ISI is minimized [1]. Since the combination of preamplifier, main amplifier, and the filter acts as a linear system (hence the name linear channel), the output voltage can be written as

$$V_{\text{out}}(t) = \int_{-\infty}^{\infty} z_T(t - t')I_p(t')\,dt',$$  

(4.3.1)

where $I_p(t)$ is the photocurrent generated in response to the incident optical power ($I_p = R_P \text{in}$). In the frequency domain,

$$\tilde{V}_{\text{out}}(\omega) = Z_T(\omega)\tilde{I}_p(\omega),$$  

(4.3.2)

where $Z_T$ is the total impedance at the frequency $\omega$ and a tilde represents the Fourier transform. Here, $Z_T(\omega)$ is determined by the transfer functions associated with various receiver components and can be written as [3]

$$Z_T(\omega) = G_p(\omega)G_A(\omega)H_F(\omega)/Y_m(\omega),$$  

(4.3.3)

where $Y_m(\omega)$ is the input admittance and $G_p(\omega)$, $G_A(\omega)$, and $H_F(\omega)$ are transfer functions of the preamplifier, the main amplifier, and the filter. It is useful to isolate the frequency dependence of $\tilde{V}_{\text{out}}(\omega)$ and $\tilde{I}_p(\omega)$ through normalized spectral functions $H_{\text{out}}(\omega)$ and $H_p(\omega)$, which are related to the Fourier transform of the output and input pulse shapes, respectively, and write Eq. (4.3.2) as

$$H_{\text{out}}(\omega) = H_T(\omega)H_p(\omega),$$  

(4.3.4)

where $H_T(\omega)$ is the total transfer function of the linear channel and is related to the total impedance as $H_T(\omega) = Z_T(\omega)/Z_T(0)$. If the amplifiers have a much larger bandwidth than the low-pass filter, $H_T(\omega)$ can be approximated by $H_F(\omega)$.

The ISI is minimized when $H_{\text{out}}(\omega)$ corresponds to the transfer function of a raised-cosine filter and is given by [3]

$$H_{\text{out}}(f) = \begin{cases} \frac{1}{2} \left( 1 + \cos(\pi f/B) \right), & f < B, \\ 0, & f \geq B. \end{cases}$$  

(4.3.5)

where $f = \omega/2\pi$ and $B$ is the bit rate. The impulse response, obtained by taking the Fourier transform of $H_{\text{out}}(f)$, is given by

$$h_{\text{out}}(t) = \frac{\sin(2\pi Bt)}{2\pi Bt} \frac{1}{1 - (2Bt)^2}.$$  

(4.3.6)

The functional form of $h_{\text{out}}(t)$ corresponds to the shape of the voltage pulse $V_{\text{out}}(t)$ received by the decision circuit. At the decision instant $t = 0$, $h_{\text{out}}(t) = 1$, and the
Figure 4.13: Ideal and degraded eye patterns for the NRZ format.

signal is maximum. At the same time, $h_{out}(t) = 0$ for $t = m/B$, where $m$ is an integer. Since $t = m/B$ corresponds to the decision instant of the neighboring bits, the voltage pulse of Eq. (4.3.6) does not interfere with the neighboring bits.

The linear-channel transfer function $H_T(\omega)$ that will result in output pulse shapes of the form (4.3.6) is obtained from Eq. (4.3.4) and is given by

$$H_T(f) = H_{out}(f)/H_p(f). \quad (4.3.7)$$

For an ideal bit stream in the nonreturn-to-zero (NRZ) format (rectangular input pulses of duration $T_B = 1/B$), $H_p(f) = B \sin(\pi f/B)/\pi f$, and $H_T(f)$ becomes

$$H_T(f) = (\pi f/2B) \cot(\pi f/2B). \quad (4.3.8)$$

Equation (4.3.8) determines the frequency response of the linear channel that would produce output pulse shape given by Eq. (4.3.6) under ideal conditions. In practice, the input pulse shape is far from being rectangular. The output pulse shape also deviates from Eq. (4.3.6), and some ISI invariably occurs.

### 4.3.3 Decision Circuit

The data-recovery section of optical receivers consists of a decision circuit and a clock-recovery circuit. The purpose of the latter is to isolate a spectral component at $f = B$ from the received signal. This component provides information about the bit slot ($T_B = 1/B$) to the decision circuit and helps to synchronize the decision process. In the case of RZ (return-to-zero) format, a spectral component at $f = B$ is present in the received signal; a narrow-bandpass filter such as a surface-acoustic-wave filter can isolate this component easily. Clock recovery is more difficult in the case of NRZ format because the signal received lacks a spectral component at $f = B$. A commonly used technique generates such a component by squaring and rectifying the spectral component at $f = B/2$ that can be obtained by passing the received signal through a high-pass filter.

The decision circuit compares the output from the linear channel to a threshold level, at sampling times determined by the clock-recovery circuit, and decides whether the signal corresponds to bit 1 or bit 0. The best sampling time corresponds to the situation in which the signal level difference between 1 and 0 bits is maximum. It
can be determined from the eye diagram formed by superposing 2–3-bit-long electrical sequences in the bit stream on top of each other. The resulting pattern is called an eye diagram because of its appearance. Figure 4.13 shows an ideal eye diagram together with a degraded one in which the noise and the timing jitter lead to a partial closing of the eye. The best sampling time corresponds to maximum opening of the eye.

Because of noise inherent in any receiver, there is always a finite probability that a bit would be identified incorrectly by the decision circuit. Digital receivers are designed to operate in such a way that the error probability is quite small (typically \( < 10^{-5} \)). Issues related to receiver noise and decision errors are discussed in Sections 4.4 and 4.5. The eye diagram provides a visual way of monitoring the receiver performance: Closing of the eye is an indication that the receiver is not performing properly.

### 4.3.4 Integrated Receivers

All receiver components shown in Fig. 4.11, with the exception of the photodiode, are standard electrical components and can be easily integrated on the same chip by using the integrated-circuit (IC) technology developed for microelectronic devices. Integration is particularly necessary for receivers operating at high bit rates. By 1988, both Si and GaAs IC technologies have been used to make integrated receivers up to a bandwidth of more than 2 GHz [53]. Since then, the bandwidth has been extended to 10 GHz.

Considerable effort has been directed at developing monolithic optical receivers that integrate all components, including the photodetector, on the same chip by using the optoelectronic integrated-circuit (OEIC) technology [54]–[74]. Such a complete integration is relatively easy for GaAs receivers, and the technology behind GaAs-based OEICs is quite advanced. The use of MSM photodiodes has proved especially useful as they are structurally compatible with the well-developed field-effect-transistor (FET) technology. This technique was used as early as 1986 to demonstrate a four-channel OEIC receiver chip [56].

For lightwave systems operating in the wavelength range 1.3–1.6 \( \mu \text{m} \), InP-based OEIC receivers are needed. Since the IC technology for GaAs is much more mature than for InP, a hybrid approach is sometimes used for InGaAs receivers. In this approach, called flip-chip OEIC technology [57], the electronic components are integrated on a GaAs chip, whereas the photodiode is made on top of an InP chip. The two chips are then connected by flipping the InP chip on the GaAs chip, as shown in Fig. 4.14. The advantage of the flip-chip technique is that the photodiode and the electrical components of the receiver can be independently optimized while keeping the parasitics (e.g., effective input capacitance) to a bare minimum.

The InP-based IC technology has advanced considerably during the 1990s, making it possible to develop InGaAs OEIC receivers [58]–[74]. Several kinds of transistors have been used for this purpose. In one approach, a \( p-i-n \) photodiode is integrated with the FETs or high-electron-mobility transistors (HEMTs) side by side on an InP substrate [59]–[63]. By 1993, HEMT-based receivers were capable of operating at 10 Gb/s with high sensitivity [62]. The bandwidth of such receivers has been increased to \( >40 \) GHz, making it possible to use them at bit rates above 40 Gb/s [63]. A waveguide
Figure 4.14: Flip-chip OEIC technology for integrated receivers. The InGaAs photodiode is fabricated on an InP substrate and then bonded to the GaAs chip through common electrical contacts. (After Ref. [57]; ©1988 IEE; reprinted with permission.)

$p$–$i$–$n$ photodiode has also been integrated with HEMTs to develop a two-channel OEIC receiver.

In another approach [64]–[69], the heterojunction-bipolar transistor (HBT) technology is used to fabricate the $p$–$i$–$n$ photodiode within the HBT structure itself through a common-collector configuration. Such transistors are often called heterojunction phototransistors. OEIC receivers operating at 5 Gb/s (bandwidth $\Delta f = 3$ GHz) were made by 1993 [64]. By 1995, OEIC receivers making use of the HBT technology exhibited a bandwidth of up to 16 GHz, together with a high gain [66]. Such receivers can be used at bit rates of more than 20 Gb/s. Indeed, a high-sensitivity OEIC receiver module was used in 1995 at a bit rate of 20 Gb/s in a 1.55-µm lightwave system [67]. Even a decision circuit can be integrated within the OEIC receiver by using the HBT technology [68].

A third approach to InP-based OEIC receivers integrates a MSM or a waveguide photodetector with an HEMT amplifier [70]–[73]. By 1995, a bandwidth of 15 GHz was realized for such an OEIC by using modulation-doped FETs [71]. By 2000, such receivers exhibited bandwidths of more than 45 GHz with the use of waveguide photodiodes [73]. Figure 4.15 shows the frequency response together with the epitaxial-layer structure of such an OEIC receiver. This receiver had a bandwidth of 46.5 GHz and exhibited a responsivity of 0.62 A/W in the 1.55-µm wavelength region. It had a clear eye opening at bit rates of up to 50 Gb/s.

Similar to the case of optical transmitters (Section 3.4), packaging of optical receivers is also an important issue [75]–[79]. The fiber–detector coupling issue is quite critical since only a small amount of optical power is typically available at the photodetector. The optical-feedback issue is also important since unintentional reflections fed back into the transmission fiber can affect system performance and should be minimized. In practice, the fiber tip is cut at an angle to reduce the optical feedback. Several different techniques have been used to produce packaged optical receivers capable of operating at bit rates as high as 10 Gb/s. In one approach, an InGaAs APD was bonded to the Si-based IC by using the flip-chip technique [75]. Efficient fiber–APD coupling was realized by using a slant-ended fiber and a microlens monolithically fabricated on
4.4. RECEIVER NOISE

Figure 4.15: (a) Epitaxial-layer structure and (b) frequency response of an OEIC receiver module made using a waveguide photodetector (WGPD). (After Ref. [73]; ©2000 IEEE; reprinted with permission.)

the photodiode. The fiber ferrule was directly laser welded to the package wall with a double-ring structure for mechanical stability. The resulting receiver module withstood shock and vibration tests and had a bandwidth of 10 GHz.

Another hybrid approach makes use of a planar-lightwave-circuit platform containing silica waveguides on a silicon substrate. In one experiment, an InP-based OEIC receiver with two channels was flip-chip bonded to the platform [76]. The resulting module could detect two 10-Gb/s channels with negligible crosstalk. GaAs ICs have also been used to fabricate a compact receiver module capable of operating at a bit rate of 10 Gb/s [77]. By 2000, fully packaged 40-Gb/s receivers were available commercially [79]. For local-loop applications, a low-cost package is needed. Such receivers operate at lower bit rates but they should be able to perform well over a wide temperature range extending from −40 to 85°C.

4.4 Receiver Noise

Optical receivers convert incident optical power $P_{in}$ into electric current through a photodiode. The relation $I_p = RP_{in}$ in Eq. (4.1.1) assumes that such a conversion is noise free. However, this is not the case even for a perfect receiver. Two fundamental noise mechanisms, shot noise and thermal noise [80]–[82], lead to fluctuations in the current even when the incident optical signal has a constant power. The relation $I_p = RP_{in}$ still holds if we interpret $I_p$ as the average current. However, electrical noise induced by current fluctuations affects the receiver performance. The objective of this section is to review the noise mechanisms and then discuss the signal-to-noise ratio (SNR) in optical receivers. The $p-i-n$ and APD receivers are considered in separate subsections, as the SNR is also affected by the avalanche gain mechanism in APDs.
4.4.1 Noise Mechanisms

The shot noise and thermal noise are the two fundamental noise mechanisms responsible for current fluctuations in all optical receivers even when the incident optical power \( P_{\text{in}} \) is constant. Of course, additional noise is generated if \( P_{\text{in}} \) itself is fluctuating because of noise produced by optical amplifiers. This section considers only the noise generated at the receiver; optical noise is discussed in Section 4.6.2.

**Shot Noise**

Shot noise is a manifestation of the fact that an electric current consists of a stream of electrons that are generated at random times. It was first studied by Schottky [83] in 1918 and has been thoroughly investigated since then [80]–[82]. The photodiode current generated in response to a constant optical signal can be written as

\[
I(t) = I_p + i_s(t),
\]

where \( I_p = RP_{\text{in}} \) is the average current and \( i_s(t) \) is a current fluctuation related to shot noise. Mathematically, \( i_s(t) \) is a stationary random process with Poisson statistics (approximated often by Gaussian statistics). The autocorrelation function of \( i_s(t) \) is related to the spectral density \( S_s(f) \) by the Wiener–Khinchin theorem [82]

\[
\langle i_s(t)i_s(t+\tau) \rangle = \int_{-\infty}^{\infty} S_s(f) \exp(2\pi if\tau) df,
\]

where angle brackets denote an ensemble average over fluctuations. The spectral density of shot noise is constant and is given by \( S_s(f) = qI_p \) (an example of white noise). Note that \( S_s(f) \) is the two-sided spectral density, as negative frequencies are included in Eq. (4.4.2). If only positive frequencies are considered by changing the lower limit of integration to zero, the one-sided spectral density becomes \( 2qI_p \).

The noise variance is obtained by setting \( \tau = 0 \) in Eq. (4.4.2), i.e.,

\[
\sigma_s^2 = \langle i_s^2(t) \rangle = \int_{-\infty}^{\infty} S_s(f) df = 2qI_p \Delta f,
\]

where \( \Delta f \) is the effective noise bandwidth of the receiver. The actual value of \( \Delta f \) depends on receiver design. It corresponds to the intrinsic photodetector bandwidth if fluctuations in the photocurrent are measured. In practice, a decision circuit may use voltage or some other quantity (e.g., signal integrated over the bit slot). One then has to consider the transfer functions of other receiver components such as the preamplifier and the low-pass filter. It is common to consider current fluctuations and include the total transfer function \( H_T(f) \) by modifying Eq. (4.4.3) as

\[
\sigma_s^2 = 2qI_p \int_0^{\infty} |H_T(f)|^2 df = 2qI_p \Delta f,
\]

where \( \Delta f = \int_0^{\infty} |H_T(f)|^2 df \) and \( H_T(f) \) is given by Eq. (4.3.7). Since the dark current \( I_d \) also generates shot noise, its contribution is included in Eq. (4.4.4) by replacing \( I_p \) by \( I_p + I_d \). The total shot noise is then given by

\[
\sigma_s^2 = 2q(I_p + I_d) \Delta f.
\]
4.4. RECEIVER NOISE

The quantity $\sigma_s$ is the root-mean-square (RMS) value of the noise current induced by shot noise.

**Thermal Noise**

At a finite temperature, electrons move randomly in any conductor. Random thermal motion of electrons in a resistor manifests as a fluctuating current even in the absence of an applied voltage. The load resistor in the front end of an optical receiver (see Fig. 4.12) adds such fluctuations to the current generated by the photodiode. This additional noise component is referred to as thermal noise. It is also called *Johnson noise* [84] or *Nyquist noise* [85] after the two scientists who first studied it experimentally and theoretically. Thermal noise can be included by modifying Eq. (4.4.1) as

\[ I(t) = I_p + i_s(t) + i_T(t), \]  

where $i_T(t)$ is a current fluctuation induced by thermal noise. Mathematically, $i_T(t)$ is modeled as a stationary Gaussian random process with a spectral density that is frequency independent up to $f \sim 1$ THz (nearly white noise) and is given by

\[ S_T(f) = \frac{2k_B T}{R_L}, \]  

where $k_B$ is the Boltzmann constant, $T$ is the absolute temperature, and $R_L$ is the load resistor. As mentioned before, $S_T(f)$ is the two-sided spectral density.

The autocorrelation function of $i_T(t)$ is given by Eq. (4.4.2) if we replace the subscript $s$ by $T$. The noise variance is obtained by setting $\tau = 0$ and becomes

\[ \sigma_T^2 = \langle i_T^2(t) \rangle = \int_{-\infty}^{\infty} S_T(f) df = (4k_B T / R_L) \Delta f, \]  

where $\Delta f$ is the effective noise bandwidth. The same bandwidth appears in the case of both shot and thermal noises. Note that $\sigma_T^2$ does not depend on the average current $I_p$, whereas $\sigma_s^2$ does.

Equation (4.4.8) includes thermal noise generated in the load resistor. An actual receiver contains many other electrical components, some of which add additional noise. For example, noise is invariably added by electrical amplifiers. The amount of noise added depends on the front-end design (see Fig. 4.12) and the type of amplifiers used. In particular, the thermal noise is different for field-effect and bipolar transistors. Considerable work has been done to estimate the amplifier noise for different front-end designs [5]. A simple approach accounts for the amplifier noise by introducing a quantity $F_n$, referred to as the amplifier noise figure, and modifying Eq. (4.4.8) as

\[ \sigma_T^2 = (4k_B T / R_L) F_n \Delta f. \]  

Physically, $F_n$ represents the factor by which thermal noise is enhanced by various resistors used in pre- and main amplifiers.

The total current noise can be obtained by adding the contributions of shot noise and thermal noise. Since $i_s(t)$ and $i_T(t)$ in Eq. (4.4.6) are independent random processes
with approximately Gaussian statistics, the total variance of current fluctuations, $\Delta I = I - I_p = i_s + i_T$, can be obtained simply by adding individual variances. The result is

$$\sigma^2 = \langle (\Delta I)^2 \rangle = \sigma_s^2 + \sigma_T^2 = 2q(I_p + I_d)\Delta f + (4k_B T/R_L)F_n\Delta f.$$  \hspace{1cm} (4.4.10)

Equation (4.4.10) can be used to calculate the SNR of the photocurrent.

### 4.4.2 $p$–$i$–$n$ Receivers

The performance of an optical receiver depends on the SNR. The SNR of a receiver with a $p$–$i$–$n$ photodiode is considered here; APD receivers are discussed in the following subsection. The SNR of any electrical signal is defined as

$$\text{SNR} = \frac{\text{average signal power}}{\text{noise power}} = \frac{I_p^2}{\sigma^2},$$  \hspace{1cm} (4.4.11)

where we used the fact that electrical power varies as the square of the current. By using Eq. (4.4.10) in Eq. (4.4.11) together with $I_p = R P_{\text{in}}$, the SNR is related to the incident optical power as

$$\text{SNR} = \frac{R^2 P_{\text{in}}^2}{2q(R P_{\text{in}} + I_d)\Delta f + 4(k_B T/R_L)F_n\Delta f},$$  \hspace{1cm} (4.4.12)

where $R = \eta q/h\nu$ is the responsivity of the $p$–$i$–$n$ photodiode.

#### Thermal-Noise Limit

In most cases of practical interest, thermal noise dominates receiver performance ($\sigma_T^2 \gg \sigma_s^2$). Neglecting the shot-noise term in Eq. (4.4.12), the SNR becomes

$$\text{SNR} = \frac{R L R^2 P_{\text{in}}^2}{4k_B T F_n\Delta f},$$  \hspace{1cm} (4.4.13)

Thus, the SNR varies as $P_{\text{in}}^2$ in the thermal-noise limit. It can also be improved by increasing the load resistance. As discussed in Section 4.3.1, this is the reason why most receivers use a high-impedance or transimpedance front end. The effect of thermal noise is often quantified through a quantity called the noise-equivalent power (NEP). The NEP is defined as the minimum optical power per unit bandwidth required to produce SNR = 1 and is given by

$$\text{NEP} = \frac{P_{\text{in}}}{\sqrt{\Delta f}} = \left( \frac{4k_B T F_n}{R_L R^2} \right)^{1/2} = \frac{h\nu}{\eta q} \left( \frac{4k_B T F_n}{R_L} \right)^{1/2}.$$  \hspace{1cm} (4.4.14)

Another quantity, called detectivity and defined as $(\text{NEP})^{-1}$, is also used for this purpose. The advantage of specifying NEP or the detectivity for a $p$–$i$–$n$ receiver is that it can be used to estimate the optical power needed to obtain a specific value of SNR if the bandwidth $\Delta f$ is known. Typical values of NEP are in the range $1–10$ pWHz$^{1/2}$. 
4.4. RECEIVER NOISE

Shot-Noise Limit

Consider the opposite limit in which the receiver performance is dominated by shot noise ($\sigma^2_s \gg \sigma^2_T$). Since $\sigma^2_s$ increases linearly with $P_{in}$, the shot-noise limit can be achieved by making the incident power large. The dark current $I_d$ can be neglected in that situation. Equation (4.4.12) then provides the following expression for SNR:

$$\text{SNR} \geq \frac{R P_{in}}{2q\Delta f} = \frac{\eta P_{in}}{2h\nu \Delta f}.$$  (4.4.15)

The SNR increases linearly with $P_{in}$ in the shot-noise limit and depends only on the quantum efficiency $\eta$, the bandwidth $\Delta f$, and the photon energy $h\nu$. It can be written in terms of the number of photons $N_p$ contained in the “1” bit. If we use $E_p = P_{in} \int_{-\infty}^{\infty} h\nu(t) dt = P_{in}/B$ for the pulse energy of a bit of duration $1/B$, where $B$ is the bit rate, and note that $E_p = N_p h\nu$, we can write $P_{in}$ as $P_{in} = N_p h\nu B$. By choosing $\Delta f = B/2$ (a typical value for the bandwidth), the SNR is simply given by $\eta N_p$. In the shot-noise limit, a SNR of 20 dB can be realized if $N_p \approx 100$. By contrast, several thousand photons are required to obtain SNR $= 20$ dB when thermal noise dominates the receiver. As a reference, for a 1.55-µm receiver operating at 10 Gb/s, $N_p = 100$ when $P_{in} \approx 130$ nW.

4.4.3 APD Receivers

Optical receivers that employ an APD generally provide a higher SNR for the same incident optical power. The improvement is due to the internal gain that increases the photocurrent by a multiplication factor $M$ so that

$$I_p = M R P_{in} = R_{APD} P_{in},$$  (4.4.16)

where $R_{APD}$ is the APD responsivity, enhanced by a factor of $M$ compared with that of $p-i-n$ photodiodes ($R_{APD} = MR$). The SNR should improve by a factor of $M^2$ if the receiver noise were unaffected by the internal gain mechanism of APDs. Unfortunately, this is not the case, and the SNR improvement is considerably reduced.

Shot-Noise Enhancement

Thermal noise remains the same for APD receivers, as it originates in the electrical components that are not part of the APD. This is not the case for shot noise. The APD gain results from generation of secondary electron–hole pairs through the process of impact ionization. Since such pairs are generated at random times, an additional contribution is added to the shot noise associated with the generation of primary electron–hole pairs. In effect, the multiplication factor itself is a random variable, and $M$ appearing in Eq. (4.4.16) represents the average APD gain. Total shot noise can be calculated by using Eqs. (4.2.3) and (4.2.4) and treating $i_e$ and $i_b$ as random variables [86]. The result is

$$\sigma^2_s = 2qM^2 F_A (R P_{in} + I_d) \Delta f.$$  (4.4.17)

where $F_A$ is the excess noise factor of the APD and is given by [86]

$$F_A(M) = k_A M + (1 - k_A)(2 - 1/M).$$  (4.4.18)
Figure 4.16: Excess noise factor $F_A$ as a function of the average APD gain $M$ for several values of the ionization-coefficient ratio $k_A$.

The dimensionless parameter $k_A = \alpha_h/\alpha_e$ if $\alpha_h < \alpha_e$ but is defined as $k_A = \alpha_e/\alpha_h$ when $\alpha_h > \alpha_e$. In other words, $k_A$ is in the range $0 < k_A < 1$. Figure 4.16 shows the gain dependence of $F_A$ for several values of $k_A$. In general, $F_A$ increases with $M$. However, although $F_A$ is at most 2 for $k_A = 0$, it keeps on increasing linearly ($F_A = M$) when $k_A = 1$. The ratio $k_A$ should be as small as possible for achieving the best performance from an APD [87].

If the avalanche–gain process were noise free ($F_A = 1$), both $I_p$ and $\sigma_s$ would increase by the same factor $M$, and the SNR would be unaffected as far as the shot-noise contribution is concerned. In practice, the SNR of APD receivers is worse than that of $p-i-n$ receivers when shot noise dominates because of the excess noise generated inside the APD. It is the dominance of thermal noise in practical receivers that makes APDs attractive. In fact, the SNR of APD receivers can be written as

$$\text{SNR} = \frac{I_p^2}{\sigma_s^2 + \sigma_T^2} = \frac{(MRP_{in})^2}{2qM^2F_A(RP_{in} + I_d)\Delta f + 4(k_BT/R_L)F_n\Delta f}$$

(4.4.19)

where Eqs. (4.4.9), (4.4.16), and (4.4.17) were used. In the thermal-noise limit ($\sigma_s \ll \sigma_T$), the SNR becomes

$$\text{SNR} = \frac{RP_{in}}{4k_BT F_n \Delta f} = M^2P_{in}^2$$

(4.4.20)

and is improved, as expected, by a factor of $M^2$ compared with that of $p-i-n$ receivers [see Eq. (4.4.13)]. By contrast, in the shot-noise limit ($\sigma_s \gg \sigma_T$), the SNR is given by

$$\text{SNR} = \frac{RP_{in}}{2qF_A \Delta f} = \frac{\eta P_{in}}{2hvF_A \Delta f}$$

(4.4.21)
and is reduced by the excess noise factor $F_A$ compared with that of $p-i-n$ receivers [see Eq. (4.4.15)].

Optimum APD Gain

Equation (4.4.19) shows that for a given $P_{in}$, the SNR of APD receivers is maximum for an optimum value $M_{opt}$ of the APD gain $M$. It is easy to show that the SNR is maximum when $M_{opt}$ satisfies the following cubic polynomial:

$$kAM_{opt}^3 + (1 - k_A)M_{opt} = \frac{4k_BT F_n}{qR_L(RP_{in} + I_d)}$$

(4.4.22)

The optimum value $M_{opt}$ depends on a large number of the receiver parameters, such as the dark current, the responsivity $R$, and the ionization-coefficient ratio $k_A$. However, it is independent of receiver bandwidth. The most notable feature of Eq. (4.4.22) is that $M_{opt}$ decreases with an increase in $P_{in}$. Figure 4.17 shows the variation of $M_{opt}$ with $P_{in}$ for several values of $k_A$ by using typical parameter values $R_L = 1 \, \text{k}\Omega$, $F_n = 2$, $R = 1 \, \text{A/W}$, and $I_d = 2 \, \text{nA}$ corresponding to a 1.55-µm InGaAs receiver. The optimum APD gain is quite sensitive to the ionization-coefficient ratio $k_A$. For $k_A = 0$, $M_{opt}$ decreases inversely with $P_{in}$, as can readily be inferred from Eq. (4.4.22) by noting that the contribution of $I_d$ is negligible in practice. By contrast, $M_{opt}$ varies as $P_{in}^{-1/3}$ for $k_A = 1$, and this form of dependence appears to hold even for $k_A$ as small as 0.01 as long as $M_{opt} > 10$. In fact, by neglecting the second term in Eq. (4.4.22), $M_{opt}$ is well...
approximated by

\[
M_{\text{opt}} \approx \left[ \frac{4k_A T F_n}{k_A q R_L (R_{\text{in}} + I_d)} \right]^{1/3}
\]

for \( k_A \) in the range 0.01–1. This expression shows the critical role played by the ionization-coefficient ratio \( k_A \). For Si APDs, for which \( k_A \ll 1 \), \( M_{\text{opt}} \) can be as large as 100. By contrast, \( M_{\text{opt}} \) is in the neighborhood of 10 for InGaAs receivers, since \( k_A \approx 0.7 \). InGaAs APD receivers are nonetheless useful for optical communication systems simply because of their higher sensitivity. Receiver sensitivity is an important issue in the design of lightwave systems and is discussed next.

### 4.5 Receiver Sensitivity

Among a group of optical receivers, a receiver is said to be more sensitive if it achieves the same performance with less optical power incident on it. The performance criterion for digital receivers is governed by the *bit-error rate* (BER), defined as the probability of incorrect identification of a bit by the decision circuit of the receiver. Hence, a BER of \( 2 \times 10^{-6} \) corresponds to on average 2 errors per million bits. A commonly used criterion for digital optical receivers requires the BER to be below \( 1 \times 10^{-9} \). The receiver sensitivity is then defined as the minimum average received power \( \bar{P}_{\text{rec}} \) required by the receiver to operate at a BER of \( 10^{-9} \). Since \( \bar{P}_{\text{rec}} \) depends on the BER, let us begin by calculating the BER.

#### 4.5.1 Bit-Error Rate

Figure 4.18(a) shows schematically the fluctuating signal received by the decision circuit, which samples it at the decision instant \( t_D \) determined through clock recovery. The sampled value \( I \) fluctuates from bit to bit around an average value \( I_1 \) or \( I_0 \), depending on whether the bit corresponds to 1 or 0 in the bit stream. The decision circuit compares the sampled value with a threshold value \( I_D \) and calls it bit 1 if \( I > I_D \) or bit 0 if \( I < I_D \). An error occurs if \( I < I_D \) for bit 1 because of receiver noise. An error also occurs if \( I > I_D \) for bit 0. Both sources of errors can be included by defining the *error probability* as

\[
\text{BER} = p(1)P(0/1) + p(0)P(1/0),
\]

where \( p(1) \) and \( p(0) \) are the probabilities of receiving bits 1 and 0, respectively, \( P(0/1) \) is the probability of deciding 0 when 1 is received, and \( P(1/0) \) is the probability of deciding 1 when 0 is received. Since 1 and 0 bits are equally likely to occur, \( p(1) = p(0) = 1/2 \), and the BER becomes

\[
\text{BER} = \frac{1}{2}[P(0/1) + P(1/0)].
\]

Figure 4.18(b) shows how \( P(0/1) \) and \( P(1/0) \) depend on the probability density function \( p(I) \) of the sampled value \( I \). The functional form of \( p(I) \) depends on the statistics of noise sources responsible for current fluctuations. Thermal noise \( i_T \) in Eq. (4.4.6) is well described by Gaussian statistics with zero mean and variance \( \sigma_T^2 \). The
4.5. RECEIVER SENSITIVITY

Figure 4.18: (a) Fluctuating signal generated at the receiver. (b) Gaussian probability densities of 1 and 0 bits. The dashed region shows the probability of incorrect identification.

The statistics of shot-noise contribution $i_j$ in Eq. (4.4.6) is also approximately Gaussian for $p-i-n$ receivers although that is not the case for APDs [86]–[88]. A common approximation treats $i_j$ as a Gaussian random variable for both $p-i-n$ and APD receivers but with different variance $\sigma^2$ given by Eqs. (4.4.5) and (4.4.17), respectively. Since the sum of two Gaussian random variables is also a Gaussian random variable, the sampled value $I$ has a Gaussian probability density function with variance $\sigma^2 = \sigma^2_p + \sigma^2_T$.

However, both the average and the variance are different for 1 and 0 bits since $I_j$ in Eq. (4.4.6) equals $I_1$ or $I_0$, depending on the bit received. If $\sigma^2_1$ and $\sigma^2_0$ are the corresponding variances, the conditional probabilities are given by

$$P(0/1) = \frac{1}{\sigma_1 \sqrt{2\pi}} \int_{-\infty}^{I_0} \exp \left( -\frac{(I-I_1)^2}{2\sigma_1^2} \right) dI = \frac{1}{2} \text{erfc} \left( \frac{I_1 - I_D}{\sigma_1 \sqrt{2}} \right), \quad (4.5.3)$$

$$P(1/0) = \frac{1}{\sigma_0 \sqrt{2\pi}} \int_{I_D}^{\infty} \exp \left( -\frac{(I-I_0)^2}{2\sigma_0^2} \right) dI = \frac{1}{2} \text{erfc} \left( \frac{I_D - I_0}{\sigma_0 \sqrt{2}} \right), \quad (4.5.4)$$

where $\text{erfc}$ stands for the complementary error function, defined as [89]

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} \exp(-y^2) dy. \quad (4.5.5)$$

By substituting Eqs. (4.5.3) and (4.5.4) in Eq. (4.5.2), the BER is given by

$$\text{BER} = \frac{1}{4} \left[ \text{erfc} \left( \frac{I_1 - I_D}{\sigma_1 \sqrt{2}} \right) + \text{erfc} \left( \frac{I_D - I_0}{\sigma_0 \sqrt{2}} \right) \right]. \quad (4.5.6)$$
Chapter 4. Optical Receivers

Equation (4.5.6) shows that the BER depends on the decision threshold $I_D$. In practice, $I_D$ is optimized to minimize the BER. The minimum occurs when $I_D$ is chosen such that

$$\frac{(I_D - I_0)^2}{2\sigma_0^2} = \frac{(I_1 - I_D)^2}{2\sigma_1^2} + \ln\left(\frac{\sigma_1}{\sigma_0}\right).$$

(4.5.7)

The last term in this equation is negligible in most cases of practical interest, and $I_D$ is approximately obtained from

$$\frac{(I_D - I_0)}{\sigma_0} = \frac{(I_1 - I_D)}{\sigma_1} \equiv Q.$$ 

(4.5.8)

An explicit expression for $I_D$ is

$$I_D = \frac{\sigma_1 I_1 + \sigma_0 I_0}{\sigma_1 + \sigma_0}.$$ 

(4.5.9)

When $\sigma_1 = \sigma_0$, $I_D = (I_1 + I_0)/2$, which corresponds to setting the decision threshold in the middle. This is the situation for most p–i–n receivers whose noise is dominated by thermal noise ($\sigma_T \gg \sigma_s$) and is independent of the average current. By contrast, shot noise is larger for bit 1 than for bit 0, since $\sigma_s^2$ varies linearly with the average current. In the case of APD receivers, the BER can be minimized by setting the decision threshold in accordance with Eq. (4.5.9).

The BER with the optimum setting of the decision threshold is obtained by using Eqs. (4.5.6) and (4.5.8) and depends only on the $Q$ parameter as

$$\text{BER} = \frac{1}{2} \text{erfc}\left(\frac{Q}{\sqrt{2}}\right) \approx \frac{\exp(-Q^2/2)}{Q\sqrt{2\pi}},$$

(4.5.10)

where the parameter $Q$ is obtained from Eqs. (4.5.8) and (4.5.9) and is given by

$$Q = \frac{I_1 - I_0}{\sigma_1 + \sigma_0}.$$ 

(4.5.11)

The approximate form of BER is obtained by using the asymptotic expansion [89] of $\text{erfc}(Q/\sqrt{2})$ and is reasonably accurate for $Q > 3$. Figure 4.19 shows how the BER varies with the $Q$ parameter. The BER improves as $Q$ increases and becomes lower than $10^{-12}$ for $Q > 7$. The receiver sensitivity corresponds to the average optical power for which $Q \approx 6$, since $\text{BER} \approx 10^{-9}$ when $Q = 6$. The next subsection provides an explicit expression for the receiver sensitivity.

### 4.5.2 Minimum Received Power

Equation (4.5.10) can be used to calculate the minimum optical power that a receiver needs to operate reliably with a BER below a specified value. For this purpose the $Q$ parameter should be related to the incident optical power. For simplicity, consider the case in which 0 bits carry no optical power so that $P_0 = 0$, and hence $I_0 = 0$. The power $P_1$ in 1 bits is related to $I_1$ as

$$I_1 = \text{MRP}_1 = 2\text{MRP}_{\text{rec}},$$

(4.5.12)
where $\bar{P}_{\text{rec}}$ is the average received power defined as $\bar{P}_{\text{rec}} = (P_1 + P_0)/2$. The APD gain $M$ is included in Eq. (4.5.12) for generality. The case of $p-i-n$ receivers can be considered by setting $M = 1$.

The RMS noise currents $\sigma_1$ and $\sigma_0$ include the contributions of both shot noise and thermal noise and can be written as

$$\sigma_1 = (\sigma_t^2 + \sigma_T^2)^{1/2}$$

$$\sigma_0 = \sigma_T,$$

where $\sigma_t^2$ and $\sigma_T^2$ are given by Eqs. (4.4.17) and (4.4.9), respectively. Neglecting the contribution of dark current, the noise variances become

$$\sigma_t^2 = 2qM^2F_A R(2\bar{P}_{\text{rec}})\Delta f,$$  \hspace{1cm} (4.5.14)

$$\sigma_T^2 = (4kBT/R_L)F_n\Delta f.$$  \hspace{1cm} (4.5.15)

By using Eqs. (4.5.11)–(4.5.13), the $Q$ parameter is given by

$$Q = \frac{I_1}{\sigma_1 + \sigma_0} = \frac{2MR\bar{P}_{\text{rec}}}{(\sigma_t^2 + \sigma_T^2)^{1/2} + \sigma_T}.$$  \hspace{1cm} (4.5.16)

For a specified value of BER, $Q$ is determined from Eq. (4.5.10) and the receiver sensitivity $\bar{P}_{\text{rec}}$ is found from Eq. (4.5.16). A simple analytic expression for $\bar{P}_{\text{rec}}$ is obtained by solving Eq. (4.5.16) for a given value of $Q$ and is given by [3]

$$\bar{P}_{\text{rec}} = \frac{Q}{R} \left( qF_A Q\Delta f + \frac{\sigma_T}{M} \right).$$  \hspace{1cm} (4.5.17)
Equation (4.5.17) shows how \( \bar{P}_{\text{rec}} \) depends on various receiver parameters and how it can be optimized. Consider first the case of a \( p-i-n \) receiver by setting \( M = 1 \). Since thermal noise \( \sigma_T \) generally dominates for such a receiver, \( \bar{P}_{\text{rec}} \) is given by the simple expression

\[
(\bar{P}_{\text{rec}})_{\text{pin}} \approx \frac{Q \sigma_T}{R}.
\]

From Eq. (4.5.15), \( \sigma_T^2 \) depends not only on receiver parameters such as \( R_L \) and \( F_n \) but also on the bit rate through the receiver bandwidth \( \Delta f \) (typically, \( \Delta f = B/2 \)). Thus, \( \bar{P}_{\text{rec}} \) increases as \( \sqrt{B} \) in the thermal-noise limit. As an example, consider a 1.55-\( \mu \)m \( p-i-n \) receiver with \( R = 1 \) A/W. If we use \( \sigma_T = 100 \) nA as a typical value and \( Q = 6 \) corresponding to a BER of \( 10^{-9} \), the receiver sensitivity is given by \( \bar{P}_{\text{rec}} = 0.6 \) \( \mu \)W or \(-32.2 \) dBm.

Equation (4.5.17) shows how receiver sensitivity improves with the use of APD receivers. If thermal noise remains dominant, \( \bar{P}_{\text{rec}} \) is reduced by a factor of \( M \), and the received sensitivity is improved by the same factor. However, shot noise increases considerably for APD, and Eq. (4.5.17) should be used in the general case in which shot-noise and thermal-noise contributions are comparable. Similar to the case of SNR discussed in Section 4.4.3, the receiver sensitivity can be optimized by adjusting the APD gain \( M \). By using \( F_A \) from Eq. (4.4.18) in Eq. (4.5.17), it is easy to verify that \( \bar{P}_{\text{rec}} \) is minimum for an optimum value of \( M \) given by

\[
M_{\text{opt}} = \frac{k_A}{\sqrt{2}} \left( \frac{\sigma_T}{qQ\Delta f} + k_A - 1 \right) \approx \left( \frac{\sigma_T}{k_A Qq\Delta f} \right)^{1/2},
\]

(4.5.19)

The improvement in receiver sensitivity obtained by the use of an APD can be estimated by comparing Eqs. (4.5.18) and (4.5.20). It depends on the ionization-coefficient ratio \( k_A \) and is larger for APDs with a smaller value of \( k_A \). For InGaAs APD receivers, the sensitivity is typically improved by 6–8 dB; such an improvement is sometimes called the APD advantage. Note that \( \bar{P}_{\text{rec}} \) for APD receivers increases linearly with the bit rate \( B \) (\( \Delta f \approx B/2 \)), in contrast with its \( \sqrt{B} \) dependence for \( p-i-n \) receivers. The linear dependence of \( \bar{P}_{\text{rec}} \) on \( B \) is a general feature of shot-noise-limited receivers. For an ideal receiver for which \( \sigma_T = 0 \), the receiver sensitivity is obtained by setting \( M = 1 \) in Eq. (4.5.17) and is given by

\[
(\bar{P}_{\text{rec}})_{\text{ideal}} = \left( \frac{q\Delta f}{R} \right) Q^2.
\]

(4.5.21)

A comparison of Eqs. (4.5.20) and (4.5.21) shows sensitivity degradation caused by the excess-noise factor in APD receivers.

Alternative measures of receiver sensitivity are sometimes used. For example, the BER can be related to the SNR and to the average number of photons \( N_p \) contained within the “1” bit. In the thermal-noise limit \( \sigma_0 \approx \sigma_1 \). By using \( I_0 = 0 \), Eq. (4.5.11) provides \( Q = I_1/2\sigma_1 \). As SNR = \( I_1^2/\sigma_1^2 \), it is related to \( Q \) by the simple relation SNR = \( 4Q^2 \). Since \( Q = 6 \) for a BER of \( 10^{-9} \), the SNR must be at least 144 or 21.6 dB for achieving BER \( \leq 10^{-9} \). The required value of SNR changes in the shot-noise limit. In
the absence of thermal noise, $\sigma_0 \approx 0$, since shot noise is negligible for the “0” bit if the dark-current contribution is neglected. Since $Q = I_1/\sigma_1 = (\text{SNR})^{1/2}$ in the shot-noise limit, an SNR of 36 or 15.6 dB is enough to obtain $\text{BER} = 1 \times 10^{-9}$. It was shown in Section 4.4.2 that $\text{SNR} \approx \eta N_p$ [see Eq. (4.4.15) and the following discussion] in the shot-noise limit. By using $Q = (\eta N_p)^{1/2}$ in Eq. (4.5.10), the BER is given by

$$\text{BER} = \frac{1}{2} \text{erfc} \left( \sqrt{\eta N_p / 2} \right). \quad (4.5.22)$$

For a receiver with 100% quantum efficiency ($\eta = 1$), $\text{BER} = 1 \times 10^{-9}$ when $N_p = 36$. In practice, most optical receivers require $N_p \sim 1000$ to achieve a BER of $10^{-9}$, as their performance is severely limited by thermal noise.

### 4.5.3 Quantum Limit of Photodetection

The BER expression (4.5.22) obtained in the shot-noise limit is not totally accurate, since its derivation is based on the Gaussian approximation for the receiver noise statistics. For an ideal detector (no thermal noise, no dark current, and 100% quantum efficiency), $\sigma_0 = 0$, as shot noise vanishes in the absence of incident power, and thus the decision threshold can be set quite close to the 0-level signal. Indeed, for such an ideal receiver, 1 bits can be identified without error as long as even one photon is detected. An error is made only if a 1 bit fails to produce even a single electron–hole pair. For such a small number of photons and electrons, shot-noise statistics cannot be approximated by a Gaussian distribution, and the exact Poisson statistics should be used. If $N_p$ is the average number of photons in each 1 bit, the probability of generating $m$ electron–hole pairs is given by the Poisson distribution [90]

$$P_m = \exp(-N_p) N_p^m / m!. \quad (4.5.23)$$

The BER can be calculated by using Eqs. (4.5.2) and (4.5.23). The probability $P(1/0)$ that a 1 is identified when 0 is received is zero since no electron–hole pair is generated when $N_p = 0$. The probability $P(0/1)$ is obtained by setting $m = 0$ in Eq. (4.5.23), since a 0 is decided in that case even though 1 is received. Since $P(0/1) = \exp(-N_p)$, the BER is given by the simple expression

$$\text{BER} = \exp(-N_p) / 2. \quad (4.5.24)$$

For $\text{BER} < 10^{-9}$, $N_p$ must exceed 20. Since this requirement is a direct result of quantum fluctuations associated with the incoming light, it is referred to as the quantum limit. Each 1 bit must contain at least 20 photons to be detected with a BER $< 10^{-9}$. This requirement can be converted into power by using $P_1 = N_p h v B$, where $B$ is the bit rate and $h v$ the photon energy. The receiver sensitivity, defined as $\bar{P}_\text{rec} = (P_1 + P_0) / 2 = P_1 / 2$, is given by

$$\bar{P}_\text{rec} = N_p h v B / 2 = \tilde{N}_p h v B. \quad (4.5.25)$$

The quantity $\tilde{N}_p$ expresses the receiver sensitivity in terms of the average number of photons/bit and is related to $N_p$ as $\tilde{N}_p = N_p / 2$ when 0 bits carry no energy. Its use
as a measure of receiver sensitivity is quite common. In the quantum limit \( \hat{N}_p = 10 \). The power can be calculated from Eq. (4.5.25). For example, for a 1.55-\( \mu \)m receiver (\( h\nu = 0.8 \text{ eV} \)), \( \bar{P}_{\text{rec}} = 13 \text{ nW} \) or \(-48.9 \text{ dBm} \) at \( B = 10 \text{ Gb/s} \). Most receivers operate away from the quantum limit by 20 dB or more. This is equivalent to saying that \( \hat{N}_p \) typically exceeds 1000 photons in practical receivers.

## 4.6 Sensitivity Degradation

The sensitivity analysis in Section 4.5 is based on the consideration of receiver noise only. In particular, the analysis assumes that the optical signal incident on the receiver consists of an ideal bit stream such that 1 bits consist of an optical pulse of constant energy while no energy is contained in 0 bits. In practice, the optical signal emitted by a transmitter deviates from this ideal situation. Moreover, it can be degraded during its transmission through the fiber link. An example of such degradation is provided by the noise added at optical amplifiers. The minimum average optical power required by the receiver increases because of such nonideal conditions. This increase in the average received power is referred to as the \textit{power penalty}. In this section we focus on the sources of power penalties that can lead to sensitivity degradation even without signal transmission through the fiber. The transmission-related power-penalty mechanisms are discussed in Chapter 7.

### 4.6.1 Extinction Ratio

A simple source of a power penalty is related to the energy carried by 0 bits. Some power is emitted by most transmitters even in the off state. In the case of semiconductor lasers, the off-state power \( P_0 \) depends on the bias current \( I_b \) and the threshold current \( I_{th} \). If \( I_b < I_{th} \), the power emitted during 0 bits is due to spontaneous emission, and generally \( P_0 \ll P_1 \), where \( P_1 \) is the on-state power. By contrast, \( P_0 \) can be a significant fraction of \( P_1 \) if the laser is biased close to but above threshold. The \textit{extinction ratio} is defined as

\[
\text{re}_\text{ex} = P_0 / P_1. \tag{4.6.1}
\]

The power penalty can be obtained by using Eq. (4.5.11). For a \( p-i-n \) receiver \( I_1 = RP_1 \) and \( I_0 = RP_0 \), where \( R \) is the responsivity (the APD gain can be included by replacing \( R \) with \( MR \)). By using the definition \( \bar{P}_{\text{rec}} = (P_1 + P_0)/2 \) for the receiver sensitivity, the parameter \( Q \) is given by

\[
Q = \left( \frac{1 - \text{re}_\text{ex}}{1 + \text{re}_\text{ex}} \right) \frac{2\bar{P}_{\text{rec}}}{\sigma_1 + \sigma_0}. \tag{4.6.2}
\]

In general, \( \sigma_1 \) and \( \sigma_0 \) depend on \( \bar{P}_{\text{rec}} \) because of the dependence of the shot-noise contribution on the received optical signal. However, both of them can be approximated by the thermal noise \( \sigma_T \) when receiver performance is dominated by thermal noise. By using \( \sigma_1 \approx \sigma_0 \approx \sigma_T \) in Eq. (4.6.2), \( \bar{P}_{\text{rec}} \) is given by

\[
\bar{P}_{\text{rec}}(\text{re}_\text{ex}) = \left( \frac{1 + \text{re}_\text{ex}}{1 - \text{re}_\text{ex}} \right) \frac{\sigma_T Q}{R}. \tag{4.6.3}
\]
4.6. SENSITIVITY DEGRADATION

This equation shows that $P_{\text{rec}}$ increases when $r_{\text{ex}} \neq 0$. The power penalty is defined as the ratio $\delta_{\text{ex}} = \frac{P_{\text{rec}}(r_{\text{ex}})}{P_{\text{rec}}(0)}$. It is commonly expressed in decibel (dB) units by using

$$\delta_{\text{ex}} = 10 \log_{10} \left( \frac{P_{\text{rec}}(r_{\text{ex}})}{P_{\text{rec}}(0)} \right) = 10 \log_{10} \left( \frac{1 + r_{\text{ex}}}{1 - r_{\text{ex}}} \right).$$  (4.6.4)

Figure 4.20 shows how the power penalty increases with $r_{\text{ex}}$. A 1-dB penalty occurs for $r_{\text{ex}} = 0.12$ and increases to 4.8 dB for $r_{\text{ex}} = 0.5$. In practice, for lasers biased below threshold, $r_{\text{ex}}$ is typically below 0.05, and the corresponding power penalty (<0.4 dB) is negligible. Nonetheless, it can become significant if the semiconductor laser is biased above threshold. An expression for $P_{\text{rec}}(r_{\text{ex}})$ can be obtained [3] for APD receivers by including the APD gain and the shot-noise contribution to $\sigma_0$ and $\sigma_1$ in Eq. (4.6.2). The optimum APD gain is lower than that in Eq. (4.5.19) when $r_{\text{ex}} \neq 0$. The sensitivity is also reduced because of the lower optimum gain. Normally, the power penalty for an APD receiver is larger by about a factor of 2 for the same value of $r_{\text{ex}}$.

4.6.2 Intensity Noise

The noise analysis of Section 4.4 is based on the assumption that the optical power incident on the receiver does not fluctuate. In practice, light emitted by any transmitter exhibits power fluctuations. Such fluctuations, called intensity noise, were discussed in Section 3.3.8 in the context of semiconductor lasers. The optical receiver converts power fluctuations into current fluctuations which add to those resulting from shot noise and thermal noise. As a result, the receiver SNR is degraded and is lower than that given by Eq. (4.4.19). An exact analysis is complicated, as it involves the calculation
of photocurrent statistics [91]. A simple approach consists of adding a third term to the current variance given by Eq. (4.4.10), so that

$$\sigma^2 = \sigma^2_s + \sigma^2_T + \sigma^2_I,$$

(4.6.5)

where

$$\sigma_I = R\langle(\Delta P_{in}^2)\rangle^{1/2} = RP_{in}r_I.$$  

(4.6.6)

The parameter $r_I$, defined as

$$r_I = \langle(\Delta P_{in}^2)\rangle^{1/2}/P_{in},$$

(4.6.7)

where $RIN(\omega)$ is given by Eq. (3.5.32). As discussed in Section 3.5.4, $r_I$ is simply the inverse of the SNR of light emitted by the transmitter. Typically, the transmitter SNR is better than 20 dB, and $r_I < 0.01$.

As a result of the dependence of $\sigma_0$ and $\sigma_1$ on the parameter $r_I$, the parameter $Q$ in Eq. (4.5.11) is reduced in the presence of intensity noise. Since $Q$ should be maintained to the same value to maintain the BER, it is necessary to increase the received power. This is the origin of the power penalty induced by intensity noise. To simplify the following analysis, the extinction ratio is assumed to be zero, so that $I_0 = 0$ and $\sigma_0 = \sigma_T$. By using $I_1 = RP_1 = 2RP_{rec}$ and Eq. (4.6.5) for $\sigma_1$, $Q$ is given by

$$Q = \frac{2RP_{rec}}{(\sigma_T^2 + \sigma_s^2 + \sigma_I^2)^{1/2} + \sigma_T},$$

(4.6.8)

where

$$\sigma_s = (4qRP_{rec}\Delta f)^{1/2}, \quad \sigma_I = 2r_I R\bar{P}_{rec},$$

(4.6.9)

and $\sigma_T$ is given by Eq. (4.4.9). Equation (4.6.8) is easily solved to obtain the following expression for the receiver sensitivity:

$$P_{rec}(r_I) = \frac{Q\sigma_T + Q^2q\Delta f}{R(1 - r_I^2Q^2)}.$$  

(4.6.10)

The power penalty, defined as the increase in $P_{rec}$ when $r_I \neq 0$, is given by

$$\delta_I = 10 \log_{10}[P_{rec}(r_I)/P_{rec}(0)] = -10 \log_{10}(1 - r_I^2Q^2).$$

(4.6.11)

Figure 4.21 shows the power penalty as a function of $r_I$ for maintaining $Q = 6$ corresponding to a BER of $10^{-9}$. The penalty is negligible for $r_I < 0.01$ as $\delta_I$ is below 0.02 dB. Since this is the case for most optical transmitters, the effect of transmitter noise is negligible in practice. The power penalty is almost 2 dB for $r_I = 0.1$ and becomes infinite when $r_I = Q^{-1} = 0.167$. An infinite power penalty implies that the receiver cannot operate at the specific BER even if the received optical power is increased indefinitely. In the BER diagram shown in Fig. 4.19, an infinite power penalty corresponds to a saturation of the BER curve above the $10^{-9}$ level, a feature referred to as the BER floor. In this respect, the effect of intensity noise is qualitatively different
4.6. SENSITIVITY DEGRADATION

than the extinction ratio, for which the power penalty remains finite for all values of \( r_{ex} \) such that \( r_{ex} < 1 \).

The preceding analysis assumes that the intensity noise at the receiver is the same as at the transmitter. This is not typically the case when the optical signal propagates through a fiber link. The intensity noise added by in-line optical amplifiers often becomes a limiting factor for most long-haul lightwave systems (see Chapter 5). When a multimode semiconductor laser is used, fiber dispersion can lead to degradation of the receiver sensitivity through the mode-partition noise. Another phenomenon that can enhance intensity noise is optical feedback from parasitic reflections occurring all along the fiber link. Such transmission-induced power-penalty mechanisms are considered in Chapter 7.

4.6.3 Timing Jitter

The calculation of receiver sensitivity in Section 4.5 is based on the assumption that the signal is sampled at the peak of the voltage pulse. In practice, the decision instant is determined by the clock-recovery circuit (see Fig. 4.11). Because of the noisy nature of the input to the clock-recovery circuit, the sampling time fluctuates from bit to bit. Such fluctuations are called timing jitter \([92]–[95]\). The SNR is degraded because fluctuations in the sampling time lead to additional fluctuations in the signal. This can be understood by noting that if the bit is not sampled at the bit center, the sampled value is reduced by an amount that depends on the timing jitter \( \Delta t \). Since \( \Delta t \) is a random
variable, the reduction in the sampled value is also random. The SNR is reduced as a
result of such additional fluctuations, and the receiver performance is degraded. The
SNR can be maintained by increasing the received optical power. This increase is the
power penalty induced by timing jitter.

To simplify the following analysis, let us consider a \textit{p–i–n} receiver dominated by
thermal noise $\sigma_T$ and assume a zero extinction ratio. By using $I_0 = 0$ in Eq. (4.5.11),
the parameter $Q$ is given by

$$Q = \frac{I_1 - \langle \Delta i_j \rangle}{(\sigma_T^2 + \sigma_j^2)^{1/2} + \sigma_T},$$  \hspace{1cm} (4.6.12)

where $\langle \Delta i_j \rangle$ is the average value and $\sigma_j$ is the RMS value of the current fluctuation $\Delta i_j$
induced by timing jitter $\Delta t$. If $h_{\text{out}}(t)$ governs the shape of the current pulse,

$$\Delta i_j = I_1 [h_{\text{out}}(0) - h_{\text{out}}(\Delta t)],$$  \hspace{1cm} (4.6.13)

where the ideal sampling instant is taken to be $t = 0$.

Clearly, $\sigma_j$ depends on the shape of the signal pulse at the decision current. A sim-
ple choice [92] corresponds to $h_{\text{out}}(t) = \cos^2(\pi B t / 2)$, where $B$ is the bit rate. Here Eq.
(4.3.6) is used as many optical receivers are designed to provide that pulse shape. Since
$\Delta t$ is likely to be much smaller than the bit period $T_B = 1 / B$, it can be approximated as

$$\Delta i_j = (2\pi^2 / 3 - 4) (B \Delta t)^2 I_1$$  \hspace{1cm} (4.6.14)

by assuming that $B \Delta t \ll 1$. This approximation provides a reasonable estimate of the
power penalty as long as the penalty is not too large [92]. This is expected to be the case in practice. To calculate $\sigma_j$, the probability density function of the timing jitter $\Delta t$
is assumed to be Gaussian, so that

$$p(\Delta t) = \frac{1}{\tau_j \sqrt{2\pi}} \exp \left( - \frac{\Delta t^2}{2 \tau_j^2} \right),$$  \hspace{1cm} (4.6.15)

where $\tau_j$ is the RMS value (standard deviation) of $\Delta t$. The probability density of $\Delta i_j$
can be obtained by using Eqs. (4.6.14) and (4.6.15) and noting that $\Delta i_j$ is proportional
to $(\Delta t)^2$. The result is

$$p(\Delta i_j) = \frac{1}{\sqrt{\pi b \Delta i_j I_1}} \exp \left( - \frac{\Delta i_j}{b I_1} \right),$$  \hspace{1cm} (4.6.16)

where

$$b = (4\pi^2 / 3 - 8)(B \tau_j)^2.$$  \hspace{1cm} (4.6.17)

Equation (4.6.16) is used to calculate $\langle \Delta i_j \rangle$ and $\sigma_j = \langle (\Delta i_j)^2 \rangle^{1/2}$. The integration
over $\Delta i_j$ is easily done to obtain

$$\langle \Delta i_j \rangle = b I_1 / 2, \quad \sigma_j = b I_1 / \sqrt{2}.$$  \hspace{1cm} (4.6.18)
4.6. SENSITIVITY DEGRADATION

By using Eqs. (4.6.12) and (4.6.18) and noting that \( I_1 = 2RP_{\text{rec}} \), where \( R \) is the respon-
sivity, the receiver sensitivity is given by

\[
\bar{P}_{\text{rec}}(b) = \left( \frac{\sigma_T Q}{R} \right) \frac{1 - b/2}{(1 - b/2)^2 - b^2 Q^2/2}.
\]  

(4.6.19)

The power penalty, defined as the increase in \( \bar{P}_{\text{rec}} \), is given by

\[
\delta_j = 10 \log_{10} \left( \frac{\bar{P}_{\text{rec}}(b)}{\bar{P}_{\text{rec}}(0)} \right) = 10 \log_{10} \left( \frac{1 - b/2}{(1 - b/2)^2 - b^2 Q^2/2} \right).
\]  

(4.6.20)

Figure 4.22: Power penalty versus the timing jitter parameter \( B\tau_j \),

![Figure 4.22: Power penalty versus the timing jitter parameter \( B\tau_j \).](image)

Figure 4.22 shows how the power penalty varies with the parameter \( B\tau_j \), which has the physical significance of the fraction of the bit period over which the decision time fluctuates (one standard deviation). The power penalty is negligible for \( B\tau_j < 0.1 \) but increases rapidly beyond \( B\tau_j = 0.1 \). A 2-dB penalty occurs for \( B\tau_j = 0.16 \). Similar to the case of intensity noise, the jitter-induced penalty becomes infinite beyond \( B\tau_j = 0.2 \). The exact value of \( B\tau_j \) at which the penalty becomes infinite depends on the model used to calculate the jitter-induced power penalty. Equation (4.6.20) is obtained by using a specific pulse shape and a specific jitter distribution. It is also based on the use of Eqs. (4.5.10) and (4.6.12), which assumes Gaussian statistics for the receiver current. As evident from Eq. (4.6.16), jitter-induced current fluctuations are not Gaussian in nature. A more accurate calculation shows that Eq. (4.6.20) underestimates the power penalty [94]. The qualitative behavior, however, remains the same. In general, the RMS value of the timing jitter should be below 10% of the bit period for a negligible power penalty. A similar conclusion holds for APD receivers, for which the penalty is generally larger [95].
4.7 Receiver Performance

The receiver performance is characterized by measuring the BER as a function of the average optical power received. The average optical power corresponding to a BER of $10^{-9}$ is a measure of receiver sensitivity. Figure 4.23 shows the receiver sensitivity measured in various transmission experiments \[96\]–\[107\] by sending a long sequence of pseudorandom bits (typical sequence length $2^{15} - 1$) over a single-mode fiber and then detecting it by using either a $p$–$i$–$n$ or an APD receiver. The experiments were performed at the 1.3- or 1.55-μm wavelength, and the bit rate varied from 100 MHz to 10 GHz. The theoretical quantum limit at these two wavelengths is also shown in Fig. 4.23 by using Eq. (4.5.25). A direct comparison shows that the measured receiver sensitivities are worse by 20 dB or more compared with the quantum limit. Most of the degradation is due to the thermal noise that is unavoidable at room temperature and generally dominates the shot noise. Some degradation is due to fiber dispersion, which leads to power penalties; sources of such penalties are discussed in the following chapter.

The dispersion-induced sensitivity degradation depends on both the bit rate $B$ and the fiber length $L$ and increases with $BL$. This is the reason why the sensitivity degradation from the quantum limit is larger (25–30 dB) for systems operating at high bit rates. The receiver sensitivity at 10 Gb/s is typically worse than $-25$ dBm [107]. It can be improved by 5–6 dB by using APD receivers. In terms of the number of photons/bit, APD receivers require nearly 1000 photons/bit compared with the quantum
limit of 10 photons/bit. The receiver performance is generally better for shorter wavelengths in the region near 0.85 µm, where silicon APDs can be used; they perform satisfactorily with about 400 photons/bit; an experiment in 1976 achieved a sensitivity of only 187 photons/bit [108]. It is possible to improve the receiver sensitivity by using coding schemes. A sensitivity of 180 photons/bit was realized in a 1.55-µm system experiment [109] after 305 km of transmission at 140 Mb/s.

It is possible to isolate the extent of sensitivity degradation occurring as a result of signal propagation inside the optical fiber. The common procedure is to perform a separate measurement of the receiver sensitivity by connecting the transmitter and receiver directly, without the intermediate fiber. Figure 4.24 shows the results of such a measurement for a 1.55-µm field experiment in which the RZ-format signal consisting of a pseudorandom bit stream in the form of solitons (sequence length $2^{23} - 1$) was propagated over more than 2000 km of fiber [110]. In the absence of fiber (0-km curve), a BER of $10^{-9}$ is realized for $-29.5$ dBm of received power. However, the launched signal is degraded considerably during transmission, resulting in about a 3-dB penalty for a 2040-km fiber link. The power penalty increases rapidly with further
propagation. In fact, the increasing curvature of BER curves indicates that the BER of $10^{-9}$ would be unreachable after a distance of 2600 km. This behavior is typical of most lightwave systems. The eye diagram seen in Fig. 4.24 is qualitatively different than that appearing in Fig. 4.13. This difference is related to the use of the RZ format.

The performance of an optical receiver in actual lightwave systems may change with time. Since it is not possible to measure the BER directly for a system in operation, an alternative is needed to monitor system performance. As discussed in Section 4.3.3, the eye diagram is best suited for this purpose; closing of the eye is a measure of degradation in receiver performance and is associated with a corresponding increase in the BER. Figures 4.13 and 4.24 show examples of the eye diagrams for lightwave systems making use of the NRZ and RZ formats, respectively. The eye is wide open in the absence of optical fiber but becomes partially closed when the signal is transmitted through a long fiber link. Closing of the eye is due to amplifier noise, fiber dispersion, and various nonlinear effects, all of which lead to considerable distortion of optical pulses as they propagate through the fiber. The continuous monitoring of the eye pattern is common in actual systems as a measure of receiver performance.

The performance of optical receivers operating in the wavelength range 1.3–1.6 µm is severely limited by thermal noise, as seen clearly from the data in Fig. 4.23. The use of APD receivers improves the situation, but to a limited extent only, because of the excess noise factor associated with InGaAs APDs. Most receivers operate away from the quantum limit by 20 dB or more. The effect of thermal noise can be considerably reduced by using coherent-detection techniques in which the received signal is mixed coherently with the output of a narrow-linewidth laser. The receiver performance can also be improved by amplifying the optical signal before it is incident on the photodetector. We turn to optical amplifiers in the next chapter.

**Problems**

4.1 Calculate the responsivity of a p–i–n photodiode at 1.3 and 1.55 µm if the quantum efficiency is 80%. Why is the photodiode more responsive at 1.55 µm?

4.2 Photons at a rate of $10^{10}$/s are incident on an APD with responsivity of 6 A/W. Calculate the quantum efficiency and the photocurrent at the operating wavelength of 1.5 µm for an APD gain of 10.

4.3 Show by solving Eqs. (4.2.3) and (4.2.4) that the multiplication factor $M$ is given by Eq. (4.2.7) for an APD in which electrons initiate the avalanche process. Treat $\alpha_\text{e}$ and $\alpha_\text{h}$ as constants.

4.4 Draw a block diagram of a digital optical receiver showing its various components. Explain the function of each component. How is the signal used by the decision circuit related to the incident optical power?

4.5 The raised-cosine pulse shape of Eq. (4.3.6) can be generalized to generate a family of such pulses by defining

$$h_{\text{out}}(t) = \frac{\sin(\pi Bt)}{\pi Bt} \frac{\cos(\pi \beta Bt)}{1 - (2\beta Bt)^2},$$
where the parameter $\beta$ varies between 0 and 1. Derive an expression for the transfer function $H_{\text{out}}(f)$ given by the Fourier transform of $h_{\text{out}}(t)$. Plot $h_{\text{out}}(t)$ and $H_{\text{out}}(f)$ for $\beta = 0, 0.5, \text{and } 1$.

4.6 Consider a 0.8-$\mu$m receiver with a silicon $p-i-n$ photodiode. Assume 20 MHz bandwidth, 65% quantum efficiency, 1 nA dark current, 8 pF junction capacitance, and 3 dB amplifier noise figure. The receiver is illuminated with 5 $\mu$W of optical power. Determine the RMS noise currents due to shot noise, thermal noise, and amplifier noise. Also calculate the SNR.

4.7 The receiver of Problem 4.6 is used in a digital communication system that requires a SNR of at least 20 dB for satisfactory performance. What is the minimum received power when the detection is limited by (a) shot noise and (b) thermal noise? Also calculate the noise-equivalent power in the two cases.

4.8 The excess noise factor of avalanche photodiodes is often approximated by $M^x$ instead of Eq. (4.4.18). Find the range of $M$ for which Eq. (4.4.18) can be approximated within 10% by $F_A(M) = M^x$ by choosing $x = 0.3$ for Si, 0.7 for InGaAs, and 1.0 for Ge. Use $k_A = 0.02$ for Si, 0.35 for InGaAs, and 1.0 for Ge.

4.9 Derive Eq. (4.4.22). Plot $M_{\text{opt}}$ versus $k_A$ by solving the cubic polynomial on a computer by using $R_L = 1 \text{k}\Omega$, $F_n = 2$, $R = 1 \text{A/W}$, $P_{\text{in}} = 1 \mu$W, and $I_d = 2 \text{nA}$. Compare the results with the approximate analytic solution given by Eq. (4.4.23) and comment on its validity.

4.10 Derive an expression for the optimum value of $M$ for which the SNR becomes maximum by using $F_A(M) = M^x$ in Eq. (4.4.19).

4.11 Prove that the bit-error rate (BER) given by Eq. (4.5.6) is minimum when the decision threshold is set close to a value given by Eq. (4.5.9).

4.12 A 1.3-$\mu$m digital receiver is operating at 100 Mb/s and has an effective noise bandwidth of 60 MHz. The $p-i-n$ photodiode has negligible dark current and 90% quantum efficiency. The load resistance is 100 $\Omega$ and the amplifier noise figure is 3 dB. Calculate the receiver sensitivity corresponding to a BER of $10^{-9}$. How much does it change if the receiver is designed to operate reliably up to a BER of $10^{-12}$?

4.13 Calculate the receiver sensitivity (at a BER of $10^{-9}$) for the receiver in Problem 4.12 in the shot-noise and thermal-noise limits. How many photons are incident during bit 1 in the two limits if the optical pulse can be approximated by a square pulse?

4.14 Derive an expression for the optimum gain $M_{\text{opt}}$ of an APD receiver that would maximize the receiver sensitivity by taking the excess-noise factor as $M^x$. Plot $M_{\text{opt}}$ as a function of $x$ for $\sigma_T = 0.2 \text{mA}$ and $\Delta f = 1 \text{GHz}$ and estimate its value for InGaAs APDs (see Problem 4.8).

4.15 Derive an expression for the sensitivity of an APD receiver by taking into account a finite extinction ratio for the general case in which both shot noise and thermal noise contribute to the receiver sensitivity. You can neglect the dark current.
4.16 Derive an expression for the intensity-noise-induced power penalty of a p–i–n receiver by taking into account a finite extinction ratio. Shot-noise and intensity-noise contributions can both be neglected compared with the thermal noise in the off state but not in the on state.

4.17 Use the result of Problem 4.16 to plot the power penalty as a function of the intensity-noise parameter $r_I$ [see Eq. (4.6.6) for its definition] for several values of the extinction ratio. When does the power penalty become infinite? Explain the meaning of an infinite power penalty.

4.18 Derive an expression for the timing-jitter-induced power penalty by assuming a parabolic pulse shape $I(t) = I_p(1 - B^2 t^2)$ and a Gaussian jitter distribution with a standard deviation $\tau$ (RMS value). You can assume that the receiver performance is dominated by thermal noise. Calculate the tolerable value of $B\tau$ that would keep the power penalty below 1 dB.

References

REFERENCES


REFERENCES