Elements of Photonics, Volume II: For Fiber and Integrated Optics. Keigo Iizuka

ISBNs: 0-471-40815-8 (Hardback); 0-471-22137-6 (Electronic)

16

COMMUNICATING BY FIBER OPTICS

The prime purpose of communication systems is to transfer data from point A to point B at a remote site. In the past this was accomplished primarily by using copper wires as a conduit for communication. In the present day, however, fibers are now much more often installed than copper wire. This is a result of many significant advantages of fiber-optic communications over its copper counterpart. Some main advantages include lower transmission loss and greater information capacity. These concepts are briefly summarized as follows:

1. Lower Transmission Loss. Fiber and copper wires are used to transmit telephone, television, and computer signals. For both fiber and copper wire communications, this signal level reduces with distance. Before the signal level is attenuated to a value that is lower than the noise level of the detector, the data line is terminated. The signal is then detected and amplified or just amplified in order to be retransmitted to the next repeater station.

The spacing between the repeater stations is one of the parameters used to characterize transmission systems. Long repeater spacings are desirable. While a typical repeater distance for copper wire transmission is every 10–50 kilometers, the optical fiber requires repeater stations at a distance of greater than a few hundred kilometers.

2. Greater Information Capacity. The speed at which data can be transferred (the bit rate) for copper coaxial cables, over a distance of 1 km, ranges only up to a few hundred megabits/second (Mb/s). In comparison, light pulses sent through optical fibers have bit rates of up to tens of gigabits/second (Gb/s). This is an improvement of about one hundredfold over copper transmission.

Furthermore, by making use of a special method of lightwave propagation called soliton waves in the optical fiber (see Section 15.5) combined with optical amplifiers, the light pulses can be sent over a distance as far as a few thousand kilometers with minimal distortion of the transmitted light pulses.

16.1 OVERVIEW OF FIBER-OPTIC COMMUNICATION SYSTEMS

Before going into detail, a brief overview of the entire fiber-optic communication system is presented here [1–11]. It is hoped that this section will be useful in orienting the reader and providing a simple perspective on the subject.

A pictorial block diagram of a fiber-optic communication system is shown in Fig. 16.1. The operation of the system is based simply on three major functions:

1. Conversion of electrical signals into modulated light by the transmitter.
2. Transmission of the modulated light signal through optical fiber networks.
3. Detection of the transmitted light by the receiver in order to recover the original electrical signal.

16.1.1 Transmitters

Semiconductor light sources such as the laser diode (LD) and light-emitting diode (LED) are most frequently used as light transmitters. Their popularity is largely due to such advantages as longevity, reliability, high efficiency, and compactness. The wavelengths of light used are predominantly 1.3 and 1.55 μm. These wavelengths were chosen from the attractive properties of optical fibers at these values. For example, light transmission is distorted least at a wavelength of 1.3 μm. On the other hand, light transmission loss is minimum at 1.55 μm.

16.1.2 Modulation of Light

In order to encode the signal information, the transmitted light must be altered or modulated using modulating circuitry. There are two major methods of

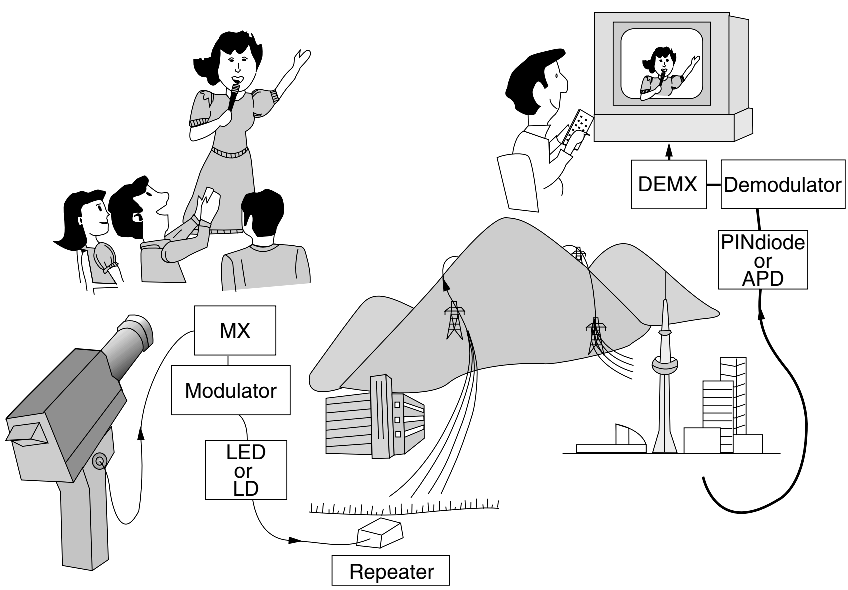


Figure 16.1 Pictorial block diagram of fiber-optic communication system. MX: Multiplexer. DEMX: Demultiplexer. LED: Light emitting diode. LD: Laser diode. APD: Avalanche photodiode.

modulation — analog and digital. Analog modulation is, in general, simpler than digital modulation, but the quality of the transmitted signal has some limitations. Digital modulation provides better quality of the transmitted signal but requires more complicated circuitry and a wider frequency bandwidth. The choice of modulation depends on the type of information being sent, the desired quality of signal, and how signals are combined during transmission.

For instance, transmitting speech through a voice channel does not require a high quality of transmission. Hence, analog modulation is usually used. However, if it is combined with other digitized signals, digital modulation is used to send the voice channel as well.

The intensity of the light from a laser diode can be modulated by varying its bias current, as shown in Fig. 16.2a. The operating characteristics of the laser diode in Fig. 16.2b demonstrate that the power of the light output follows the current of the electrical input.

In contrast, one may want to convert the amplitude of an electrical signal into its binary equivalent. For instance, a TV signal, which is an amplitude modulated signal,

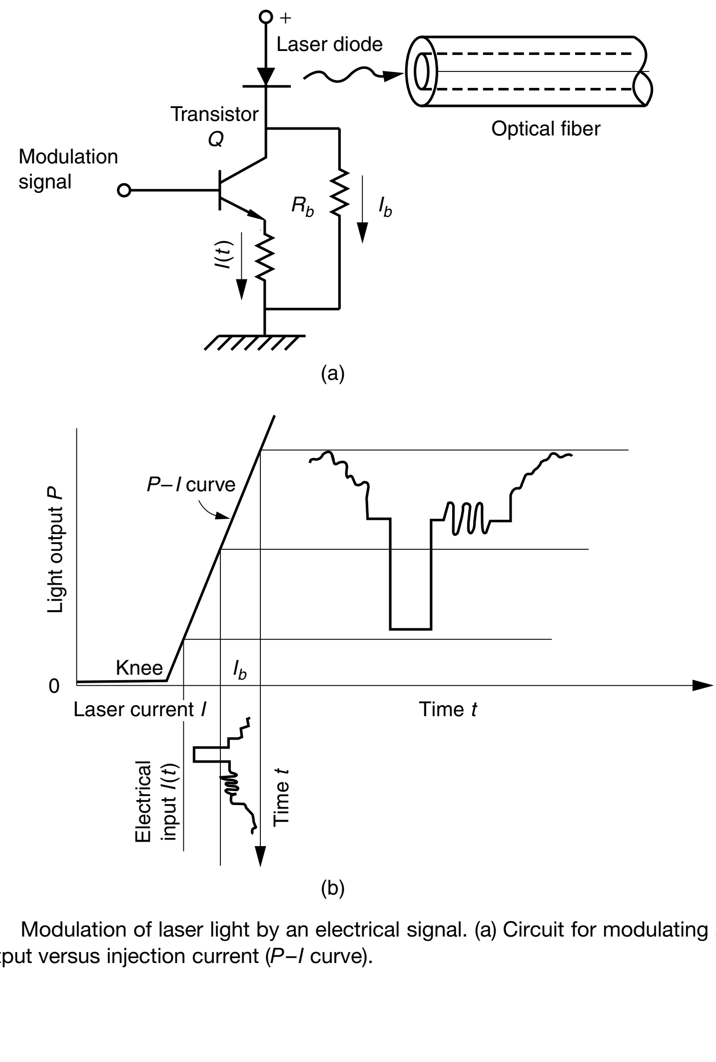


Figure 16.2 Modulation of laser light by an electrical signal. (a) Circuit for modulating a laser diode. (b) Light output versus injection current (P–I curve).

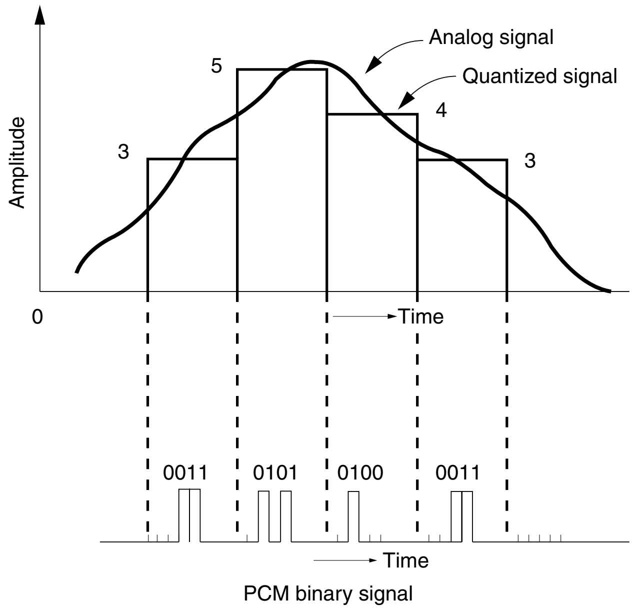


Figure 16.3 Principle of pulse code modulation (PCM).

must be coded in digital form. The most common method of digital modulation is the pulse code modulation (PCM) method, as illustrated in Fig. 16.3. In this figure, the value of the amplitude of the signal is sampled periodically. The sampled values are then converted into a binary form. For instance, if the first four sampled values were 3, 5, 4, and 3, they would be converted into an array of the following binary 0 and 1 combinations of 0011, 0101, 0100, 0011. Such an array of zeros and ones are then used to bias the laser diode to convert into light pulses. The light pulses propagate through the optical fiber network and reach the receiver, where the light pulses are converted back into an electrical signal of zeros and ones. These signals are then converted into analog signals by means of a digital-to-analog converter. Transmission with digital modulation requires much greater bandwidth than analog modulation. However, since the fiber-optic communication system inherently has an abundance of bandwidth, it can cope nicely with this situation.

16.1.3 Transmission Through the Optical Fiber

Next, the transmission of the light signal through the fiber is introduced. There are two types of fibers. One is the multimode fiber and the other is the single-mode fiber. The standard core diameter of the multimode fiber is 50 μm, while that of the single-mode fiber is 9.5 μm. For the multimode fiber, coupling light from the transmitter into the fiber's core is easier than coupling light into a smaller diameter single-mode fiber. However, the amount of the spread or distortion of a light pulse is greater in the multimode fiber than in the single-mode fiber.

The spread of the pulse shape is dealt with by increasing the spacing between the adjacent pulses so as to avoid overlapping and hence corrupting the individual signals. Since the spread is greater for multimode fibers, the spacing in time between adjacent pulses is greater than that for single-mode fibers. Thus, the maximum bit rate that can be sent through multimode fibers is less than through single-mode fibers. Due to this difference, multimode fibers are used for low-bit-rate communication or in applications that require only short-distance communication. In contrast, single-mode fibers are used for high-bit-rate, long-haul communication systems.

An extremely short distance line-of-sight communication can use space-optic communication. Such systems are convenient for high-rise business buildings within a one kilometer radius. This scheme saves both cost and time of laying fiber optic cables underground of already congested business districts. The transmission condition of the laser beam, however, is subjected to atmospheric conditions, such as, fog, heavy rain, or clear air turbulence. Reliability of the communication system has to be sacrificed unless redundancy or some other countermeasure is incorporated [12].

16.1.4 Received Signal

The light reaching the receiver is first converted into an electrical signal by means of either a PIN photodiode or an avalanche photodiode (APD).

The detector is then followed by an electronic amplifier. Since the internal impedance of the PIN photodiode or APD is very high (of the order of megohms), special impedance transducers are necessary for optimizing the signal transfer from the detector to the preamplifier without sacrificing the frequency bandwidth.

In the event that the transmission fiber is long, optical amplifiers are installed. Erbium-doped fiber amplifiers are used to amplify light signals for 1.55-μm wavelengths. For 1.3-μm systems, semiconductor laser amplifiers (SLAs) are used.

16.1.5 Multiplexing Hierarchies

In reality, optical communication systems never send just one voice channel at a time; instead, they combine as many as over several thousand voice channels together. This combination is called *multiplexing* (MX). Currently, there are three main methods of multiplexing the electrical signal. They are frequency division multiplexing (FDM), time division multiplexing (TDM), and code division multiplexing (CDM). Separate from electrical signal multiplexing, there is optical signal multiplexing of different wavelengths of light, which is called wavelength division multiplexing (WDM). Most frequently, FDM, TDM, or CDM are combined with WDM to make the most of the wideband capability of fiber-optic communication systems.

The following sections will provide more detailed knowledge of the theory and practical implementations of the basic components of an optical communication system.

16.2 MODULATION

Paul Revere, who successfully defended the New England coast of America from a massive British invasion, used a flare signal from an observation post on Beacon Hill, signaling “one if by land, two if by sea.” One flare meant invasion through land; two flares meant invasion through Boston harbor. Modern communication theory would describe Revere’s warning light as being amplitude modulated by the baseband signal of one pulse meaning by land, and two pulses meaning by sea. Information can only be sent after being altered or modulated by a baseband signal to which a meaning is assigned. For instance, in communications, our voice is considered the baseband signal that modulates light around a center frequency called a carrier frequency. In this section, methods of modulation [13–16] by the baseband signal and methods of demodulation to recover the baseband signal will be summarized.

Methods of modulation are broadly categorized into analog and digital modulations. Digital modulation is further subdivided into binary pulse modulation and multiple state pulse modulation. Each has advantages and disadvantages. Selection is made based on criteria such as frequency bandwidth, robustness against distortion and noise, and economy of light power. A summary of the various types of modulation and their features is presented in Table 16.1. Let's start with analog modulation that includes amplitude, frequency, and phase modulations.

16.2.1 Amplitude Modulation

Amplitude, frequency, and phase are the three parameters that can be used to modulate a lightwave. Amplitude modulation (AM) is the simplest of all. From the top row of Table 16.1, amplitude modulation with a baseband signal $s(t)$ is expressed as

$$E(t) = E_0[1 + ms(t)] \cos 2\pi f_c t \tag{16.1}$$

where f_c is the carrier frequency of the light and m is the modulation index. The frequency spectrum obtained by Fourier transforming this equation ranges from $f_c - B$ to $f_c + B$ and is centered at f_c, where B is the frequency bandwidth of the baseband signal. Thus, the frequency spectrum of the AM stretches over a bandwidth of $2B$. There are two ways of implementing AM in connection with a semiconductor laser diode source. One is internal and the other is external modulation.

With internal modulation, the baseband signal current combined with the proper bias current is directly injected into a laser diode to modulate its output light. In this approach, it is not the amplitude but the intensity of the light that is linearly modulated by the baseband signal current. Luckily, it is not the amplitude but the intensity of the received light that is proportional to the output electrical current from the photodiode detection. Such a method of detection is called *direct detection*.

The problem with this method of internal modulation is that the frequency of the output light is also modulated according to the injection current, resulting in unwanted frequency modulation (FM). This inherent FM of the light signal introduces the problem of additional spreading of the signal pulse within the fiber due to material and waveguide dispersion (see Section 11.1.4).

With external modulation, the laser diode is driven by a constant injection current, and its output light amplitude is externally modulated by devices such as the electrooptic amplitude modulator described in Section 5.2.1. This system is more involved but is free from the unwanted FM. Amplitude as well as intensity modulations can be achieved by the external modulation method.

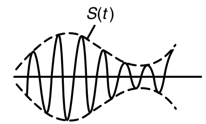
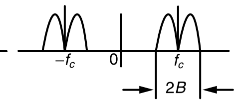
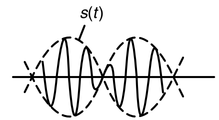
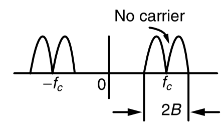
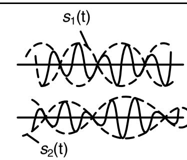
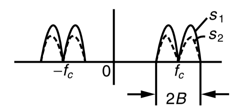
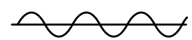
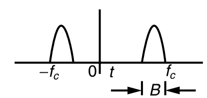
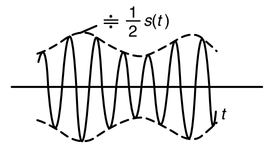
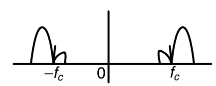
16.2.2 Variations of Amplitude Modulation

There are a few variations in the AM as also shown in Table 16.1. The first is double sideband (DSB) modulation in which the carrier is suppressed and is expressed as

$$E(t) = s(t) \cos \omega_c t \tag{16.2}$$

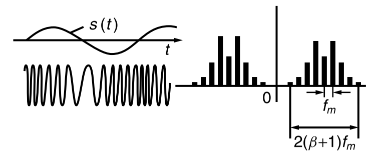
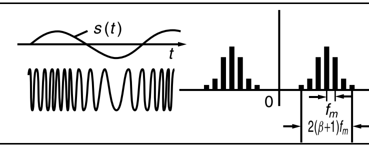
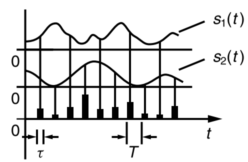
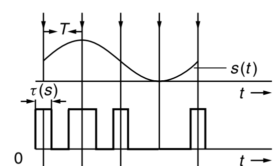
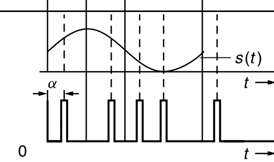
The carrier does not carry information, and its removal saves light energy. However, it requires a more elaborate demodulation scheme.

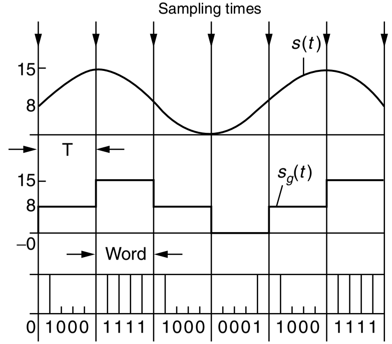
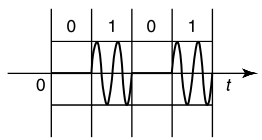
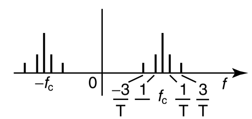
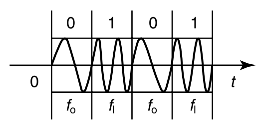
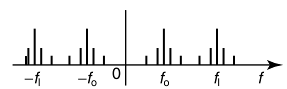
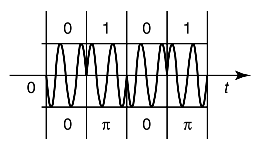
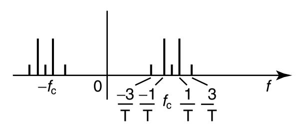
Table 16.1 Table of various types of modulation

Category	Acronym	Full Name	Expression	Spectrum: Time Domain	Spectrum: Frequency Domain	Features
Carrier Amplitude Modulation	AM	Amplitude modulation	$A[1 + ms(t)] \cos \omega_c t$ Spectrum of $s(t)$ is B Hertz			Simplest to modulate and demodulate. Bandwidth is not economized.
	IM	Intensity modulation	$I[1 + ms(t)] \cos \omega_c t$			
	DSB	Double sideband modulation	$s(t) \cos \omega_c t$			Easy to modulate but difficult to demodulate. No energy for carrier frequency. (Required power is 50% less than AM).
	DSBC	Double sideband suppressed carrier modulation				
	QAM	Quadrature amplitude modulation	$s_l(t) \cos \omega_c t + s_2(t) \sin \omega_c t$			Bandwidth is conserved.
	SSB	Single sideband modulation	$\frac{1}{2}[s(t) \cos \omega_c t + \hat{s}(t) \sin \omega_c t]$ $\hat{s}(t)$: Hilbert transform of $s(t)$	 For the special case of $s(t) = \cos 2\pi f_m t$		Bandwidth is conserved. Difficult to demodulate.
	VSB	Vestigial sideband modulation	$\frac{1}{2}[s(t) \cos \omega_c t - js(t) \sin \omega_c t]$			Bandwidth is conserved. Difficult to demodulate. One and a fraction of sideband is used. No such abrupt filter as used for SSB is needed.

(*continued next page*)

Table 16.1 (*Continued*)

Category	Acronym	Full Name	Expression	Spectrum: Time Domain	Spectrum: Frequency Domain	Features
Angle Modulation	PM	Phase modulation	$A\cos[\omega_c t + s(t)]$			Implementation is complicated. Stringent requirement of LO laser. Robust against distortion and noise.
	FM	Frequency modulation	$A\cos[\omega_c t + \int_{t_o}^{t} s(\lambda)d\lambda]$			Implementation is complicated. Robust against distortion and noise.
Pulse Modulation	PAM	Pulse amplitude modulation	$\sum_{n=-\infty}^{\infty} [as(nT)+K]\Pi\left(\frac{t-nT}{\tau}\right)$			Both modulation and demodulation are simple. The narrow pulse width is susceptible to noise. The example shown is used for the time division multiplexing.
	PDM	Pulse duration modulation	$\sum_{n=-\infty}^{\infty} \Pi\left(\frac{t-nT}{\tau(s)}\right)$ $\tau(s) = as(nT) + K$			The amplitude is not susceptible to noise. The positions of the zero crossing point of both sides of the pulse are susceptible to noise.
	PPM	Pulse position modulation	$\sum_{n=-\infty}^{\infty} \Pi\left(\frac{t-nT-\alpha}{\tau(s)}\right)$ $\alpha = as(nT)$			The pulses are all the same height and width. Only one side of the zero crossing point is susceptible to noise.

Analog Signal in Digital Form	PCM	Pulse code modulation			Bandwidth = 2nB n = Nunber of quantization B = Bandwidth of the analog signal	Commonly used. Robust against distortion and noise. Moderate cost for high-quality transmission.
Binary Modulation (Two Level Modulation)	ASK	Amplitude shift keying	$h(t)\cos\omega_c t$ $h(t) = \begin{cases} 0 \\ 1 \end{cases}$			Simple to modulate and demodulate. Robust against distortion and noise
	FSK	Frequency shift keying	$\cos[\omega_h(t)t]$ $\omega_h(t) = \begin{cases} 2\pi f_0 \\ 2\pi f_1 \end{cases}$			Three-decibel higher energy than ASK. Complicated in implementation. Robust against distortion and noise.
	PSK	Phase shift keying	$\cos[\omega t + \phi(t)]$ $\phi(t) = \begin{cases} 0 \\ \pi \end{cases}$			Three-decibel higher energy than ASK. Stringent requirement on the reference wave to detect.

Next, there is quadratic amplitude modulation (QAM), which can transmit two signals without increasing the frequency bandwidth. The light is split into two parts. The first part (in-phase component) is multiplied by the baseband signal $s_1(t)$ to give

$$E_I(t) = s_1(t)\cos\omega_c t \tag{16.3}$$

The phase of the other part (quadrature component) is shifted from that of the first part by $\pi/2$ radians. The quadrature component is multiplied by the other baseband signal $s_2(t)$ to give

$$E_Q(t) = s_2(t)\sin\omega_c t \tag{16.4}$$

E_I and E_Q are transmitted simultaneously. Without increasing the frequency bandwidth, both $s_1(t)$ and $s_2(t)$ can be transmitted. In the receiver, the signals are selectively demodulated by multiplying $E_I(t)$ by $\cos\omega_c t$, and by multiplying $E_Q(t)$ by $\sin\omega_c t$, and then filtering with a low-pass filter.

In DSB modulation, the two sideband signals contain essentially the same information. Hence, this redundant signal wastes power and bandwidth during its transmission. To economize in both of these areas, single sideband (SSB) modulation was introduced, where only a single sideband is transmitted. A single-sideband signal is mathematically expressed as [14]

$$E(t) = \tfrac{1}{2}[s(t)\cos\omega_c t + \hat{s}(t)\sin\omega_c t] \tag{16.5}$$

where $\hat{s}(t)$ is called the Hilbert transform of $s(t)$, which is realized by passing the baseband signal through a 90° phase shifter. The above expression can be verified by using the example of a sinusoidal signal, namely,

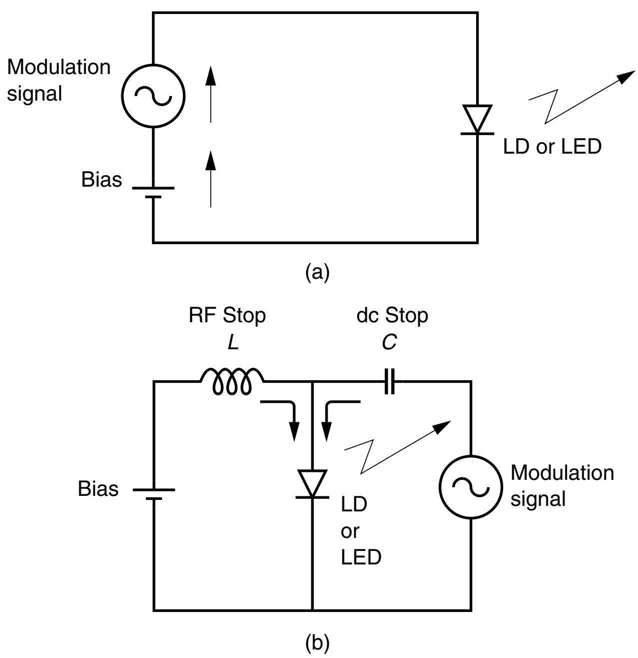


Figure 16.4 Amplitude modulation of a radiant diode. (a) Series circuit modulation. (b) Shunt circuit modulation.

$$\begin{aligned} s(t) &= a\cos\omega_m t \\ \hat{s}(t) &= a\sin\omega_m t \end{aligned} \tag{16.6}$$

This results in a modulated signal of

$$E(t) = \tfrac{1}{2}a\cos(\omega_c - \omega_m)t \tag{16.7}$$

which contains only the lower sideband $(\omega_c - \omega_m)$ but not the higher sideband $(\omega_c + \omega_m)$.

The drawback of SSB modulation is that it requires additional circuitry to remove one of the sidebands. It also requires a more elaborate demodulation.

Finally, in vestigial sideband (VSB) modulation, only a portion of the amplitude modulated signal is removed. No abrupt cutoff filters are needed, and the remaining vestigial fraction of the carrier frequency is used for demodulating the signal.

In the design of the internal AM circuit, one should consider (1) the ease of modulation, (2) the linearization of the characteristic curve, and (3) matching to a low input impedance of typically 1.5–2.0 Ω. The methods of internal AM are classified into series and shunt connections, as shown in Fig. 16.4. Samples of both types of connections for modulating an LED or LD are shown in Fig. 16.5.

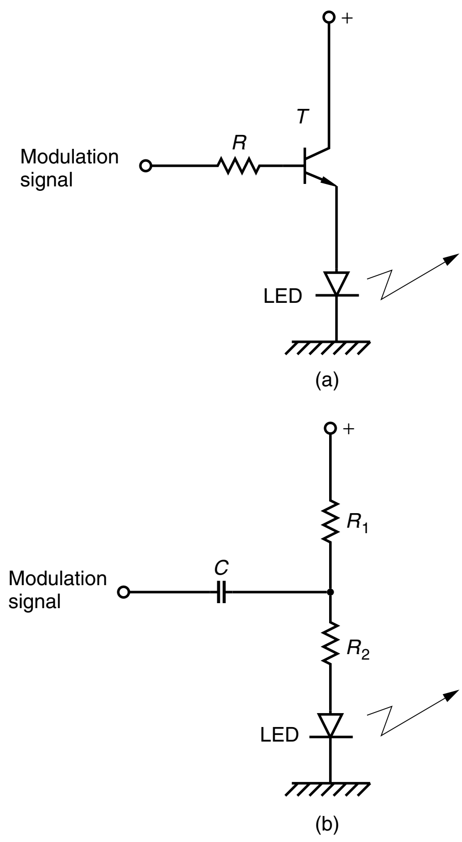


Figure 16.5 Simple circuits to modulate an LED. (a) Emitter follower (series connection). (b) *RC* coupling (shunt connection).

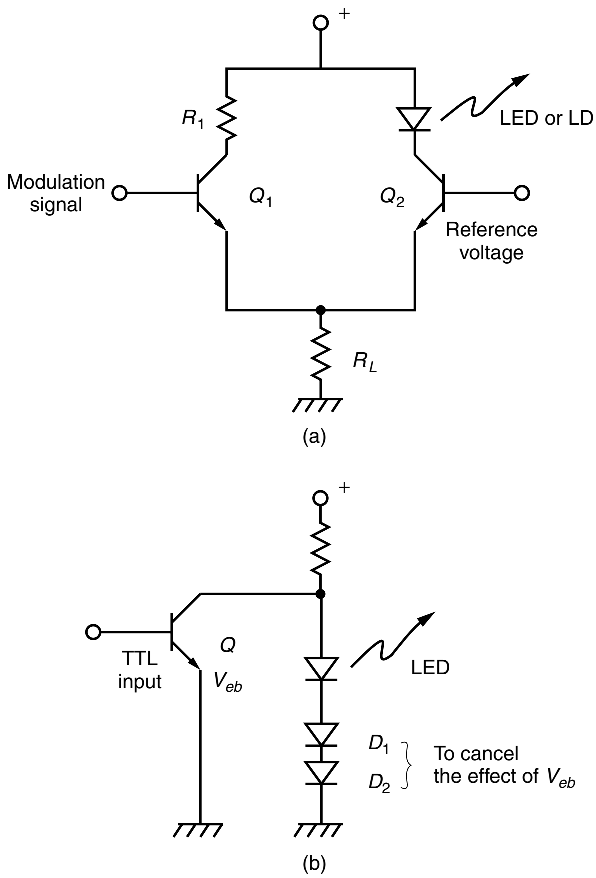


Figure 16.6 Circuits for modulating an LED. (a) Emitter coupled circuit. (b) Parallel switching.

Figure 16.6 shows some additional circuits used in modulating an LD or LED. The circuit shown in Fig. 16.6a is an emitter coupled circuit. By pairing the same type of transistors, the drift of the operating point due to the ambient temperature is suppressed. Another modulation method is the parallel switching shown in Fig. 16.6b. As the transistor Q is turned on and off, the shunt circuit to the LED is also switched on and off. The diodes D_1 and D_2 are to assure the off state of the LED.

Figure 16.7 shows driver circuits with pre-enhancement. In order to operate in the linear region of the characteristic curve of the light power versus bias current of the LD, the driver shown in Fig. 16.7a is prebiased by shunting the transistor Q with R_b. The value of R_b controls the bias point, which is necessary to operate in the linear region. Figure 16.7b shows a driver that compensates for the drop of output light power from an LED at higher modulating frequencies. It uses a capacitor across the resistor R. With this driver, the frequency band of modulation can be stretched.

16.2.3 Angle Modulation

Another means of modulation besides amplitude modulation is angle modulation. Phase modulation (PM) and frequency modulation (FM) are two types of angle modulation.

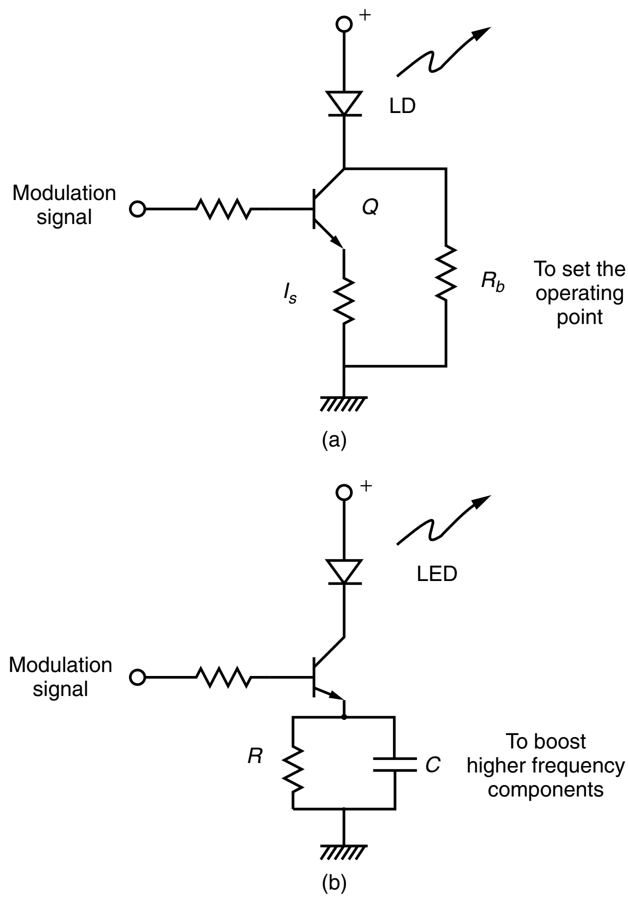


Figure 16.7 Drivers with pre-enhancement. (a) Driver with pre-biasing. (b) Driver with pre-emphasis.

The baseband signal modulates the phase of the signal to give a phase modulated expression of

$$E(t) = A\cos[\omega_c t + s(t)] \tag{16.8}$$

whereas frequency modulation can be expressed as

$$E(t) = A\cos\left(\omega_c t + \int_{t_0}^{t} s(\lambda)d\lambda\right) \tag{16.9}$$

Note that the required frequency bandwidth of the AM signal is $2B$, but in the case of FM, the required bandwidth is much wider, as shown in Table 16.1.

The benefit of PM and FM over AM is significantly improved discrimination against noise and interference. For both PM and FM, the method of demodulation utilizes phase and frequency as a means of retrieving the baseband signal. Thus, they do not depend on the amplitude of the light signal where interference and noise primarily show up.

The phase or frequency modulated signal, however, cannot be demodulated by the direct detection method, which can only detect the light power and cannot discriminate either the frequency or phase of the light carrier. A different detection system called a *coherent detection system* has to be used to demodulate phase or frequency modulated signals. With coherent detection, the received signal is first fed to a mixer diode where the received light is superimposed onto a local oscillator light to generate an intermediate frequency (IF) signal,

$$f_{\mathrm{IF}} = |f_c - f_{\mathrm{LO}}|$$

where f_{LO} is the carrier frequency of the local oscillator light. The FM signal is demodulated from the IF signal using a frequency discriminator, and the PM signal is demodulated from the IF signal by a phase discriminator.

16.2.4 Pulse Modulation

Pulse modulation conveys information in pulse form rather than in a continuous wave form. Various kinds of pulse modulation are outlined in this section, the first of which is called pulse amplitude modulation (PAM). In PAM, the height of the pulse is modulated according to the amplitude of the baseband signal.

The most popular use of PAM is in time division multiplexing (TDM). With TDM, several different baseband signals can be transmitted in a single channel. The graph in Table 16.1 explains how TDM is used for multiplexing two signals $s_1(t)$ and $s_2(t)$ in one channel. Signal 1 is sampled for a short time of τ seconds, and then signal 2 is sampled for another τ seconds. Samples are alternated between signals 1 and 2. The composite signal is a pulse amplitude modulated signal that is sent through a single channel. At the receiver, the two signals are separated by a properly timed switch.

At this point, it is worth briefly mentioning three other multiplexing techniques besides TDM. Frequency division multiplexing (FDM) uses different carrier frequencies for each signal and transmits them simultaneously into a single channel. Code division multiplexing (CDM) uses different codes for each signal, and at the receiver, the signals are separated using autocorrelation according to the difference in codes.

Returning now to the subject of pulse modulation, pulse duration modulation (PDM) and pulse position modulation (PPM) are two other types of pulse modulation and are shown below PAM in Table 16.1.

PDM modulates the pulse duration τ of each pulse in the pulse period T according to the amplitude of the baseband signal. The advantage of this modulation is the constant pulse amplitude. Hence, as with FM, amplitude variation due to either noise or interference can easily be removed.

PPM modulates the time of occurrence of the pulse α after the beginning of the pulse period T. PPM possesses similar advantages to those of PDM.

16.2.5 Pulse Code Modulation

In pulse code modulation (PCM), as previously introduced in Fig. 16.3, the baseband signal is first quantized into discrete amplitude levels that can be represented by binary coded pulses. The chain of binary coded pulses thus generated is used to modulate the transmitter light. The simplest binary code light modulation is to represent “1” by the presence of light and “0” by the absence. At the receiver, the light is detected to recover the binary coded pulses. The binary signal is then converted into an analog signal by a digital-to-analog converter.

The binary code modulation is special in the sense that it requires two states representing “0” and “1.” Since this type of modulation is commonly used, the next section expands on this type of modulation.

16.2.6 Binary Modulation (Two-State Modulation)

Binary modulation is often referred to as “shift keying.” In earlier days, communication officers were trained to send messages using Morse code. In Morse code, each letter of the alphabet is represented by a unique set of dots and dashes. The dots and dashes are generated by turning an electrical signal on and off. The switching device for sending Morse code was called a keying device, and the terminology has been carried over to modern communication. The keying device can generate only two states. The electrical signal is either “on” or “off” and there is no in between state.

In communication theory, binary modulation is called shift keying. If the two states are generated by the difference of two levels of amplitude, the modulation is called amplitude shift keying (ASK). If the two states are obtained using two different frequencies of the light carrier, the modulation is called frequency shift keying (FSK), and if the two states are created using two different phases of the carrier, the modulation is called phase shift keying (PSK).

16.2.7 Amplitude Shift Keying (ASK)

The two levels of ASK modulation in optical communication are provided by turning the injection current of the semiconductor laser diode on and off. Since the level of injection current of the “on” state is the same at every pulse, there is no concern about unwanted FM as in the case of direct AM of the semiconductor laser diode.

At the receiver, the ASK modulated light is detected and reduced to the original chain of electrical pulses. The “0” or “1” state is determined by comparing the level of the received signal to that of a comparator. If the signal level is higher than the comparator level, then the state is a “1” state; and if lower, the state is a “0” state.

Thus, with ASK modulation, even when distortion due to noise is almost half as large as the regular pulse height, the pulse can still be correctly decoded as long as the extent of the distortion is less than half the height of the undistorted pulses. Not only is ASK modulation robust against signal deterioration due to distortion and noise, but it is also simple to implement and analyze theoretically.

16.2.8 Frequency Shift Keying (FSK) and Phase Shift Keying (PSK)

FSK shifts the carrier frequency of the light between two different frequencies corresponding to the “1” and “0” states of the binary code. The FSK modulated light can be generated either by using a frequency tunable semiconductor laser diode or by switching between two semiconductor laser diodes of two different wavelengths.

The implementation of FSK modulation is more complicated than that of ASK, and the frequency spectrum bandwidth is $2(\Delta f + B)$, where $2\Delta f$ is the spacing between the two frequencies, and $2B$ is due to the bandwidth of the baseband signal. Compared to ASK, the required bandwidth of FSK is $2\Delta f$ wider. PSK shifts the phase of the light carrier between two different phases using an electrooptical phase modulator.

One of the great advantages of FSK and PSK, however, is greater robustness against distortion and noise than ASK. The reason for this is that ASK uses amplitude information, which is most easily influenced by noise; whereas FSK uses information about the frequency to determine the state.

The ASK modulated light is in the off state about 50% of the time. The FSK modulated light is always on; thus, the amount of light energy that reaches the receiver is 3 dB higher than that of ASK, which leads to a higher signal to noise ratio for an FSK system.

Compared to FSK, PSK provides an improved signal to noise ratio, but its method of detection is more complicated. The detection method is homodyne detection. It involves creating a local light with a frequency and phase identical to or in the vicinity of those of the received light.

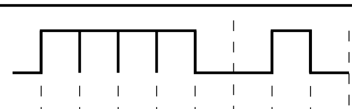
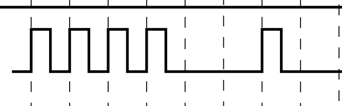
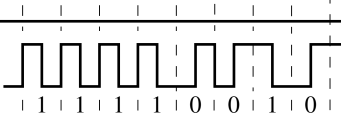
16.2.9 Representation of Bits

The manner of representing the logic states "1" and "0" by the states of an electrical signal will next be illustrated. The state of the electrical signal is represented by either the amplitude of the pulse in ASK, the frequency of the pulse in FSK, or the phase of the pulse carrier in PSK. The coding characteristics are listed in Table 16.2.

NRZ (Nonreturn to Zero) Code The logic "1" is represented by maintaining the electrical signal in the "1" state throughout the bit period. Similarly, the logic "0" is represented by the state "0" of the electrical signal maintained throughout the bit period. Even though coding is simple, there are some drawbacks. When a capacitor coupled circuit is used in the input of a signal amplifier, an uninterrupted chain of continuous "1" bits is seen as a dc signal and the signal is blocked by the capacitor and the dc level drifts. This phenomenon is called *baseline wander*. In an effort to remove this drawback, the next code was invented.

RZ (Return to Zero) Code The logic "1" state is represented by the state "1" of the electrical signal maintained only during the first half of the bit period, and during

Table 16.2 Representation of bits

Acronyn	Full Name	Description	Example
NRZ	Nonreturn to zero code	Logic "1": Electrical state 1 during entire period Logic "0": Electrical state 0 during entire period	
RZ	Return to zero code	Logic "1": Electrical state 1 during half period Logic "0": Electrical state 0 during entire period	
Manchester	Manchester code	Logic "1": Transition from electrical state 1 to 0 Logic "0": Transition from electrical state 0 to 1	

the second half of the period, the signal is turned to the "0" state. Thus, the pulse of the "1" state lasts only one-half of the bit period. The logic "0", on the other hand, is represented by the state "0" of the electrical signal during the entire bit period.

With this coding, even when "1" states appear in a continuous sequence, the signal level changes every midbit and can be amplified by a capacitor coupled amplifier. The required frequency bandwidth of the amplifier, however, doubles that of NRZ.

In a long-distance transmission, however, the noise of the optical amplifiers and the dispersion of the optical fiber affect the performance. A better overall quality is reported with NRZ code than with RZ code [16].

Manchester Code With Manchester code, logic is represented by the transitions of the states of the electrical signal rather than by the state itself.

The logic "1" is represented by the transition of the state of the electric signal from "1" to "0", while the logic "0" is represented by the transition in the opposite direction which is from "0" to "1." The simple rule is that logic "1" starts from the state of "1" and the logic "0" starts from the state of "0." Since Manchester code always compares the levels of two adjacent pulses, it is not affected by the long-term variation of the signal level. Moreover, transitions exist at every midbit, and a clock pulse can easily be generated inside the receiver.

There is a price to be paid for the robustness of binary coding for all three methods and that price is increased frequency bandwidth. Let us calculate quickly the frequency bandwidth needed to send one 4-kHz voice channel using PCM coding. In order to send a 4-kHz voice channel, the amplitude of the voice signal must be sampled at least twice each period of the highest frequency component of the voice channel (Nyquist criterion). This means that the instantaneous value of the voice signal has to be sampled at 8×10^3 times a second. Each sampled value must immediately be converted into 8-bit words or 256 gray scales to represent the sampled value. Thus, the bit rate, equal to the number of bits generated per second to send one voice channel, is $8 \times 8 \times 10^3 =$ 64 kb/s.

Each pulse is a rectangular shape. A rule of thumb for estimating the required frequency bandwidth for an almost perfect recovery of the pulse shape is 10 times the bit rate; but in digital communication, perfect shape recovery is not essential since it is only necessary to distinguish between the presence and absence of a pulse. In order to discern between the presence and absence of a pulse, a frequency bandwidth B equal to the bit rate B_t for RZ code and one-half of the bit rate B_t for NRZ code of the pulses will suffice; namely,

$$B = B_t \quad \text{for RZ}$$

$$B = \tfrac{1}{2}B_t \quad \text{for NRZ}$$

This is explained further in Section 16.6.5. Thus, the PCM of one voice channel of 4 kHz needs a frequency bandwidth of at least 32 kHz.

16.3 MULTIPLEXING

A fiber-optic communication system has such an abundant frequency bandwidth that multiplexing schemes for sending all different channels in the same optical fiber have been implemented.

16.3.1 Wavelength Division Multiplexing (WDM)

The first layer of the hierarchy of multiplexing methods is wavelength division multiplexing (WDM). Figure 16.8a shows an example of a small scale WDM used for sending several TV channels in one fiber [17]. The electrical signal of each TV camera intensity modulates a laser diode emitting at its own specific wavelength. The spacing between adjacent carrier frequencies is maintained at 8 GHz (see Section 3.2.4). The light outputs from the array of laser diodes are then fed into a scrambler, which

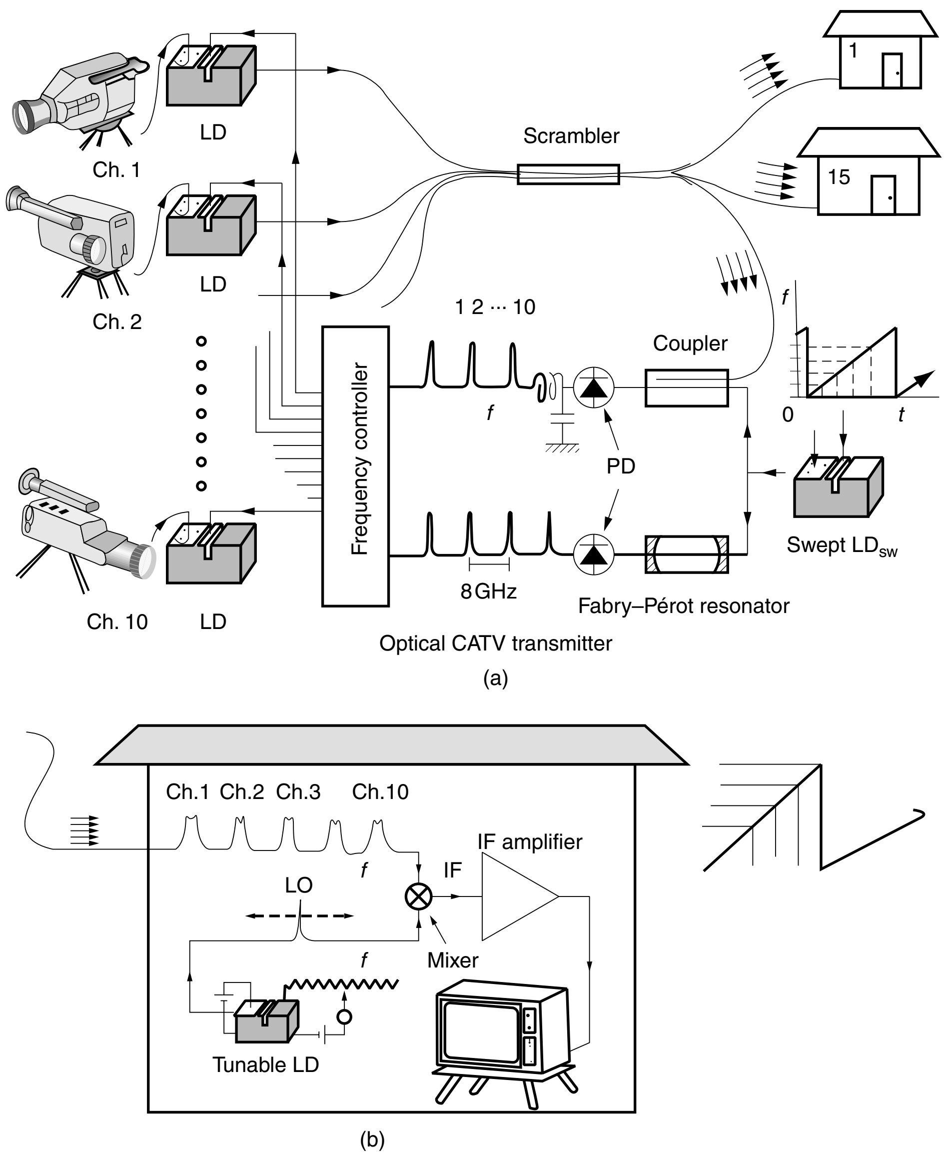


Figure 16.8 An example of wavelength division multiplexing (WDM) used for sending TV channels. (a) Channel space locking by the reference pulse method. (After S. Yamazaki et al. [17].) (b) Optical heterodyne receiver at home.

combines all the wavelengths of the light, producing a WDM system. The combined signal is then shared and sent to its subscribers via fiber.

From there, the combined signal is received by the subscriber's TV. The TV shown in Fig. 16.8b is equipped with a coherent detector. As you may recall (in Section 12.6), a coherent detector uses a local oscillator that can be tuned in order to select a particular frequency of the incoming scrambled light, thereby allowing the subscriber to view a desired channel.

16.3.2 Frequency Division Multiplexing (FDM)

Figure 16.9a illustrates how FDM is combined with the above-mentioned WDM and applied to telecommunications. Each voice channel (baseband) is used to amplitude modulate the output of a fixed frequency oscillator. The oscillator frequency is called the subcarrier frequency. The ith voice channel modulates the ith subcarrier frequency f_i. The outputs of the subcarrier frequencies

$$f_1, f_2, f_3, \ldots, f_i$$

are combined by a combiner T_1 that superimposes the various signals. Typically, a wide frequency band transformer is used to combine the signal for a lower frequency range of carrier frequencies, as shown in Fig. 16.9a. Alternatively, an RC circuit combiner can be used for a higher subcarrier frequency range. This is the first stage of frequency division multiplexing.

The frequency spacing between the subcarriers f_i is 8 kHz for the FDM of a voice channel because the frequency spectrum of $\cos 2\pi f_c t$, where f_c is 4 kHz, has two spectra at ± 4 kHz, occupying a frequency spectrum of 8 kHz centered at the subcarrier frequency. In general, the subcarrier frequency f_i has a spacing of $2B$ for multiplexing B frequency band signals.

Similar arrangements of FDM are made with transformers: $T_2, T_3, T_4, \ldots, T_k$. Let us take T_k as an example. The output from transformer T_k modulates the laser diode LD_k whose carrier wavelength is λ_k. The light outputs of k different wavelengths are focused into a single optical fiber to perform WDM. Thus, only one optical fiber transports $i \times k$ voice channels all at once.

If the multiplexed signals are again FD multiplexed by n times before modulating the laser diode, $i \times n \times k$ voice channels can be multiplexed.

Next, a method of demultiplexing the signal after reaching the receiver will be explained by referring to Fig. 16.9b. The received light first illuminates an optical demultiplexer such as a dispersive prism or a grating. The received light components are diffracted toward their respective locations according to wavelength into a waiting photodiode detector.

The output of each detector contains the frequency division multiplexed electrical signals that are further demultiplexed according to the subcarrier frequencies f_i by a bank of band-pass filters with center frequencies $f_1, f_2, f_3, \ldots, f_i$ with bandwidth $2B$. The demultiplexed signals are detected by each of their own detectors and an array of the original baseband signals is recovered. Figure 16.10 shows a similar but bidirectional case. An arrayed-waveguide grating (AWG) is used as a demultiplexer (DEMX). The AWG is a grating of arrayed waveguides deposited on a slab guide (see Section 10.6.1). It disperses the angle of diffraction in accordance with of the wavelengths [18].

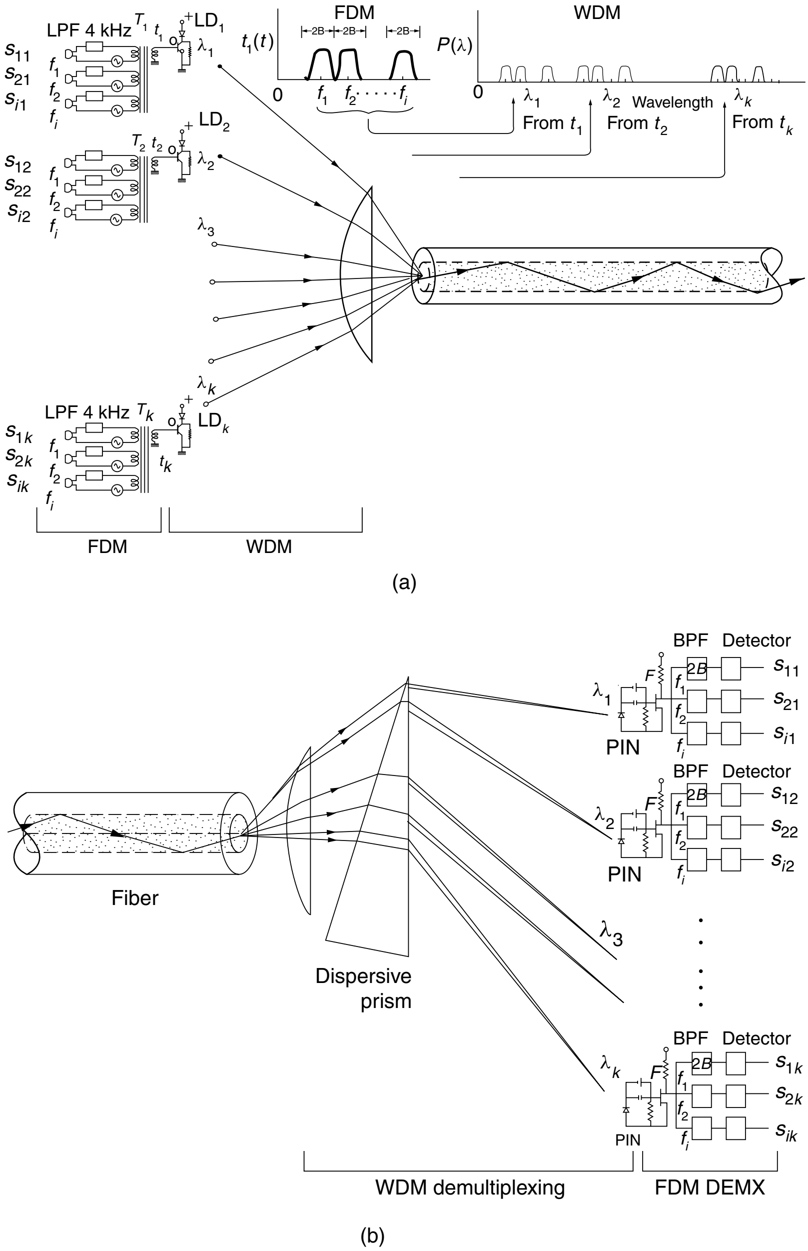


Figure 16.9 (a) WDM combined with FDM. (b) Demultiplexing of WDM combined with FDM.

16.3.3 Time Division Multiplexing (TDM)

Next, time division multiplexing (TDM), which is used for digital modulation, is explained using an example of multiplexing two digital voice channels.

In Fig. 16.11, output pulses from two channels are switched back and forth and two chains of light pulses are interleaved into one. Such a multiplexed signal is sent through

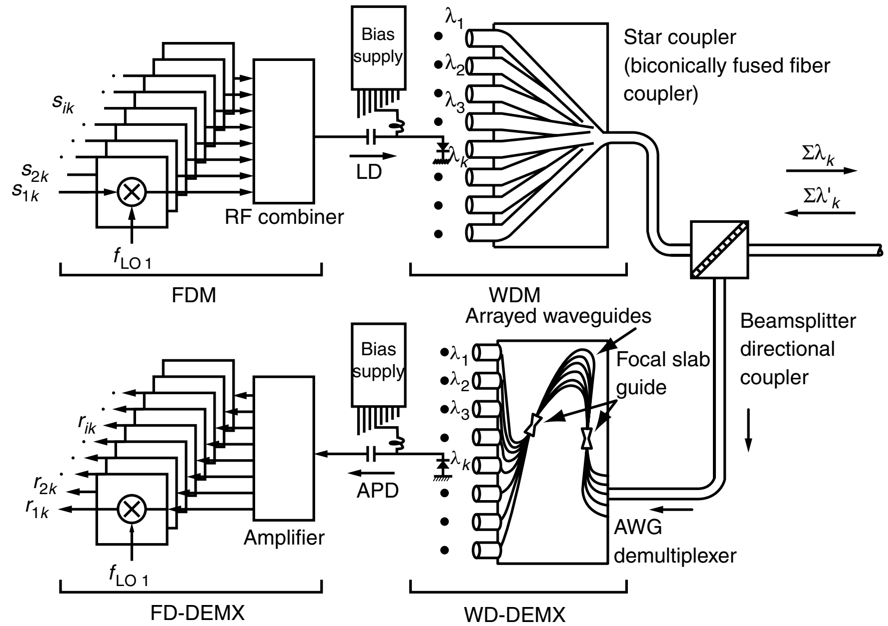


Figure 16.10 Bidirectional WDM combined with FDM.

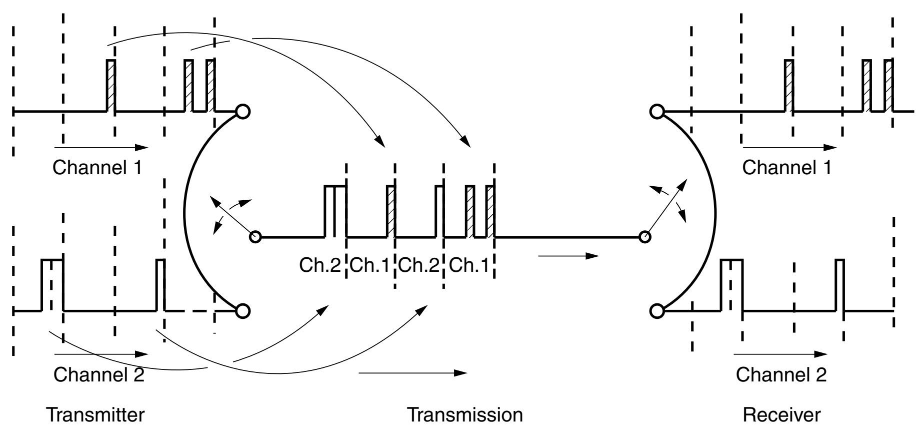


Figure 16.11 Time division multiplexing (TDM) and demultiplexing.

the optical communication system. At the receiver, the signal is demultiplexed into the two original separate channels by a similar switching system. Provisions for proper synchronization of the switches at the transmitter and the receiver are necessary and are derived from the periods of the received signal averaged over time. The illustrated example demonstrates the case of multiplexing two channels but it is not unusual to multiplex over a thousand channels per fiber.

A system that can send 24 voice channels is considered the most primitive (basic) communication system and is called digital system 1 (DS-1) in the American telephone

Table 16.3 Designation of PCM telephone bit rate

System Name	Number of Voice Channels	Data Rate (Mb/s)
	In the United States	
DS-1	24	1.544
DS-1C	48	3.152
DS-2	96	6.312
DS-3	672	44.736
DS-3C	1344	91.053
DS-4	4032	274.175
	In Japan	
F-6M	96	6.312
F-32M	480	32.064
F-100M	1440	97.728
F-200M	2880	198.584
F-400M	5760	397.200

system. Such a system allocates 1.544 Mb/s for the pulse rate as shown in Table 16.3. Let us confirm this allocated bit rate. The required pulse rate is 4 kHz (voice frequency bandwidth) $\times$ 2 (number of samples in one cycle) $\times$ 8 (each sampled value is gray scaled by 8 bits) $\times$ 24 (number of TDM circuits) = 1.536 Mb/s, which is (1.544 − 1.536) = 8 kHz less than the allocated bit rate. This gap is used for synchronizing the switches and signaling the telephone calls. Table 16.3 summarizes telephone bit rates for the standardized U.S. system, as well as the standardized Japanese system.

In the next section, detection of the modulated light in the receiver will be discussed.

16.4 LIGHT DETECTION SYSTEMS

The principles involved in light detection have already been mentioned in Chapter 12. In this chapter, these principles are explored further to describe how the performance of the detector influences the overall performance of the fiber-optic communication system. Such factors as the system's transmission distance, frequency bandwidth, signal to noise ratio for the analog signal, and bit error rate in the digital system all critically depend on the quality and capabilities of the receiver.

16.4.1 Equivalent Circuit of the PIN Photodiode

The PIN photodiode can be considered as a combination of a p-n junction diode and a constant current generator driven by the incident light. The current i_d from the p-n junction diode for the case without incident light (see Appendix A of Vol. II) is

$$i_d = I_{so}\left(e^{eV/kT} - 1\right) \tag{16.10}$$

As shown in Fig. 16.12, the current i_d increases exponentially for positive V. For large negative V, the current approaches the value I_{so}. As seen from Eq. (16.10), I_{so} is the current when the p-n junction is deeply back-biased. It is called the *saturated back-biased current*.

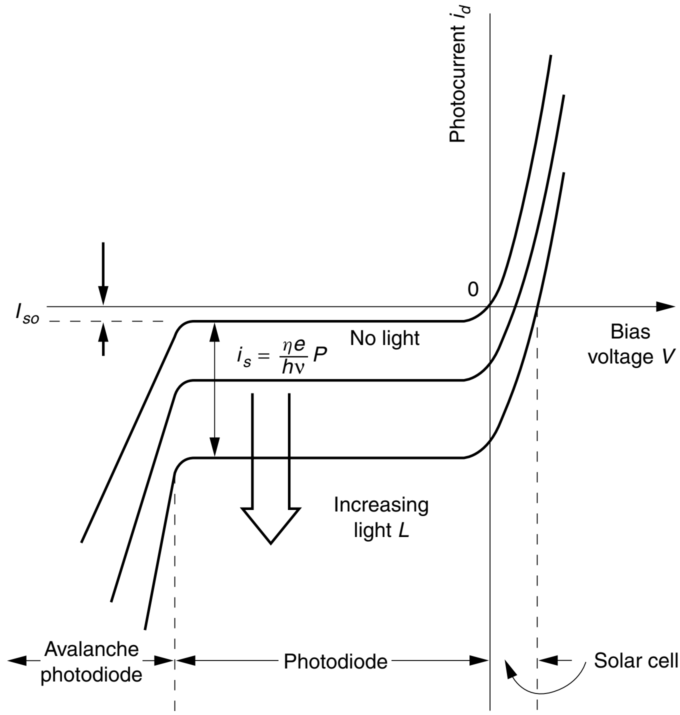


Figure 16.12 Current–voltage characteristic of a p-n junction.

When light whose quantaum $h\nu$ is larger than the bandgap is incident on the junction, there is a chance that an electron in the valence band will absorb the photon energy and move into the conduction band, leaving a hole behind. This process is called *photon-induced pair production*. The current due to the incident photons or photocurrent has to be added to Eq. (16.10). The characteristic curve including the photocurrent is shown in Fig. 16.12. Increasing the number of incident photons causes a downward translation of the I–V curve. In the negative bias region, also referred to as the back-biased region, the n layer is positive and the p layer is negative. The negative bias region is used as a photodetector because the dark current (current in absence of incident light) is at its minimum, having reached the saturated back-biased current. The output current is then primarily due to the photocurrent. The shot noise, which is proportional to the current through the junction, can be minimized. Moreover, in this region, the incremental impedance of the p-n junction is practically independent of the back-bias voltage.

The region of extreme negative bias is that of the avalanche effect. With a large bias voltage, the accelerated electrons start creating additional electron–hole pairs by colliding with the orbital electrons. The photocurrent can be multiplied up to a thousand times depending on the bias current. This is the avalanche photodiode (APD) (see Section 12.4).

Figure 16.13a shows a direct detection circuit using a PIN photodiode such as the one shown in Fig. 12.4. Figure 16.13b shows the simplified equivalent circuit including noise sources and the current generator driven by the incident light. The source and the current therefore consist of the diode current expressed in Eq. (16.10) and the photocurrent expressed in Eq. (12.8). Quite often, the dark current, which is the current

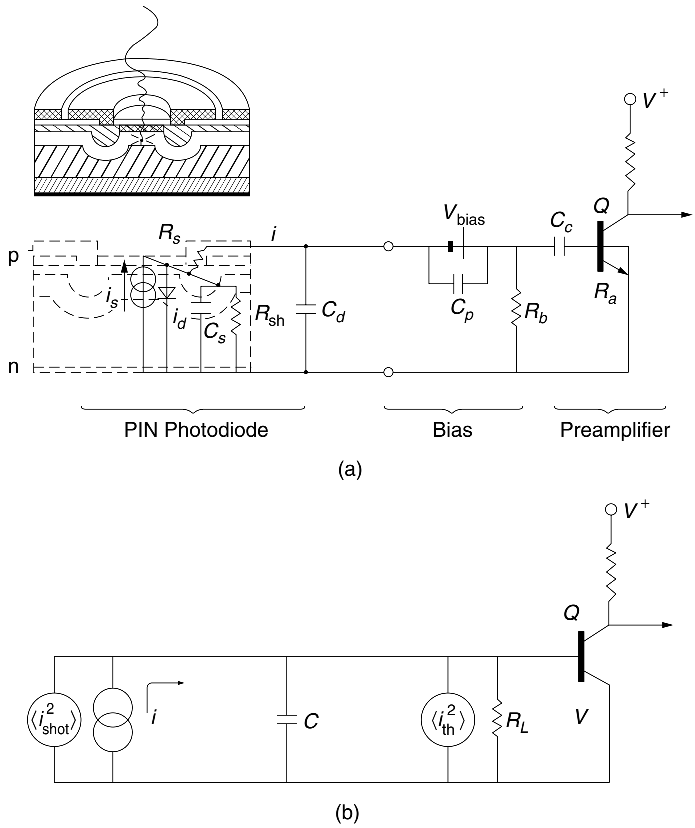


Figure 16.13 Equivalent circuits of a PIN diode. (a) Connection diagram of PIN photodiode. (b) Simplified equivalent circuit including noise sources.

without incident light, is more than the saturated back-biased current. Such excessive dark current is caused by impurities and lattice defects inside, as well as on the surface, of the crystal. This extra current is accounted for by the shunt resistance R_{sh}. The capacitive effect of the immobile charge in the depletion region is represented by C_s; the distributed capacitance associated with the connecting wire, by C_d; the contact resistance of the electrode, by R_s; the input resistance of the amplifier, by R_a; the bias resistor, by R_b; the RF bypass capacitor, by C_p; and the coupling capacitor, by C_c.

16.4.2 Frequency Response of the PIN Diode

The upper limit on the amplitude modulation frequency of the signal light that can be detected by a PIN diode is determined by the following factors:

1. The transit time for an electron generated in the intrinsic-type layer to reach the electrode.
2. The time for the carriers generated in either p-type or n-type layers to reach the electrodes.
3. The RC time constant of the signal circuit.

Ways to raise the upper limit on the frequency response are outlined as follows:

1. The transit time in the intrinsic layer can be shortened by raising the back-bias voltage. With a minimum doping level in the intrinsic layer, the drift velocity is accelerated and the resistivity that is needed to maintain the high electric field can be achieved. The higher electric field means a shorter transit time. Another way of shortening the transit time is to shorten the length of the intrinsic region, but the penalty for this is a reduction in the efficiency of converting the incident photons into electrons.

2. Referring to Fig. 12.5, there is a chance that an incident photon will create an electron–hole pair inside the p^+-type layer before reaching the i-type layer. For such carriers inside the p^+-type layer, the electric field is weak due to the higher conductivity. The carriers must then depend on diffusion to reach the electrodes, which is a much slower process than electron drift in the presence of an external field. These carriers not only slow down the response of the photodiode but also become a source of detector noise. Such problems are minimized by designing the p^+-type layer as thin as possible.

3. The time constant RC associated with the external circuit of the photodiode also limits the frequency response of the detector. Referring to the equivalent circuit in Fig. 16.13a, this time constant will be calculated. The potential drop across R_s, and the current through R_{sh}, and the impedance of the coupling capacitor C_c are assumed negligibly small.

The photocurrent i from Fig. 16.13a is

$$i = i_s - i_d \tag{16.11}$$

where i_s is the signal current, and i_d is the diode current as given by Eq. (16.10).

With the simplified equivalent circuit in Fig. 16.13b, the photocurrent is divided between the capacitor C and resistor R_L, where

$$\begin{aligned} C &= C_s + C_d \\ \frac{1}{R_L} &= \frac{1}{R_a} + \frac{1}{R_b} \end{aligned} \tag{16.12}$$

The output voltage is calculated as

$$\begin{aligned} V &= \frac{R_L/j\omega C}{R_L + 1/j\omega C} i \\ &= \frac{R_L}{\sqrt{1 + (\omega R_L C)^2}} e^{j\phi} \cdot i \end{aligned} \tag{16.13}$$

where

$$\tan\phi = -\omega R_L C \tag{16.14}$$

The cutoff frequency is

$$f_c = \frac{1}{2\pi R_L C} \tag{16.15}$$

The cutoff frequency determines the upper limit of the modulation frequency. One can increase this upper limit by reducing the value of $R_L C$. This can be accomplished by selecting a detector diode with a small value of C, or reducing the load R_L to the diode. However, in the latter case, one should note that reduction of R_L results also in a reduction of the output voltage V as shown by Eq. (16.13).

The above description of the PIN diode characteristics also applies to the APD except for the higher gain.

Next, the coupling of current converted from the incident light to the preamplifier will be explained.

16.4.3 Coupling Circuits to a Preamplifier

Special attention is needed in the design of the circuit that couples the detector with the preamplifier. Like any other circuit design, the following three considerations are important:

1. Minimization of mismatch.
2. Wide frequency band operation.
3. Optimized signal to noise ratio.

Depending on the region of the operating frequency, the circuit design is broadly divided into two approaches. One is the lumped element approach for frequencies below a gigahertz, and the other is the distributed element approach for frequencies above this level. For midrange frequencies, hybrid circuits of the two are used.

16.4.3.1 Coupling Circuits to a Preamplifier at Subgigahertz

Examples of the lumped element approach are described next.

A 50-Ω Circuit with Mismatch Figure 16.14 shows a PIN detector circuit [19]. The photodiode is loaded with a 50-Ω resistor, and its output is fed directly to a 50-Ω input impedance preamplifier. R_b prevents the photocurrent going through the battery as well as prevents damage to the preamplifier in the event that the photodiode becomes short-circuited. An advantage of such a circuit is that any length of 50-Ω coaxial cable can be used to connect the photodiode to a standard 50-Ω amplifier, making it convenient for putting instrumentation together quickly. A disadvantage is the low sensitivity of the circuit to the input voltage due to the low value of the 50-Ω load in comparison to the high internal impedance of the photodiode. Furthermore, at higher frequencies, the reflection at the junction of the diode to the 50-Ω lead wire severely degrades the performance. Another disadvantage of this circuit is the high average square thermal noise current. As will be shown in Section 16.5.2, the average square thermal noise current is inversely proportional to the load resistance (refer to Eq. (16.27)).

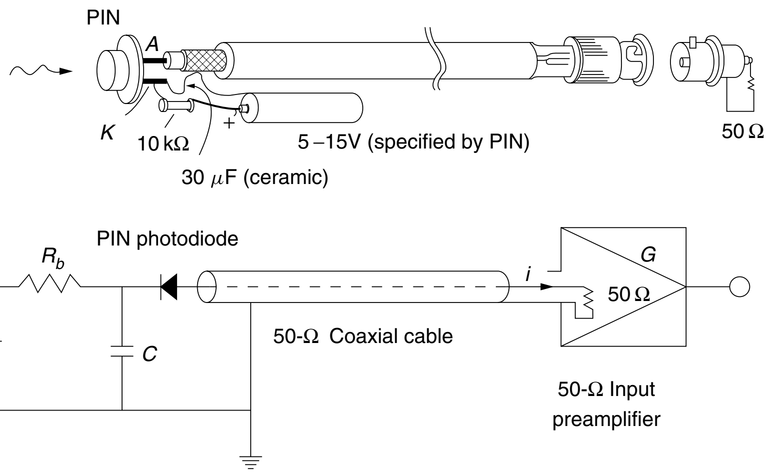


Figure 16.14 PIN photodiode connected to a 50-Ω input preamplifier.

High Impedance with a Compensating Circuit (HZ Circuit) Another circuit design involves using a high load impedance. An advantage of a high load impedance is a low average square thermal noise current, as detailed later. The problem with this design, however, is a low cutoff of the frequency response given by Eq. (16.15). One means of dealing with this problem is to combine the high impedance load with a compensator circuit that raises the cutoff frequency.

Figure 16.15 shows an example of such a circuit. The input to the operational amplifier is essentially a high input impedance low-pass filter. The output circuit of the operational amplifier is made a high-pass filter to compensate the input low-pass filter. The output of the operational amplifier is connected to the R_1C_1 parallel circuit, which is then connected in series with R_L. The output voltage from the operational amplifier is divided between the impedance of the R_1C_1 parallel circuit and R_L. As the frequency is raised beyond a critical frequency f_{c_1} the portion of the potential drop across R_L becomes large compared to that of the R_1C_1 parallel circuit. The critical frequency is the frequency above which the reactance $1/\omega_1 C_1$ is smaller than R_1, and

$$f_{c_1} = \frac{1}{2\pi R_1 C_1}$$

Next, let us design an output circuit that neither overcompensates nor undercompensates the input low-pass filter. If the input impedance of the operational amplifier is assumed infinity, the input voltage V to the operational amplifier becomes the same as the value already given by Eq. (16.13).

The output voltage V_O is

$$\begin{aligned} V_O &= VG\frac{R_L}{R_L + \dfrac{R_1}{1 + j\omega C_1 R_1}} \\ &= \frac{iGRR_L}{(1 + j\omega CR)} \cdot \frac{(1 + j\omega C_1 R_1)}{R_1 + R_L + j\omega C_1 R_1 R_L} \end{aligned} \tag{16.16}$$

where G is the gain of the amplifier and i is the input current.

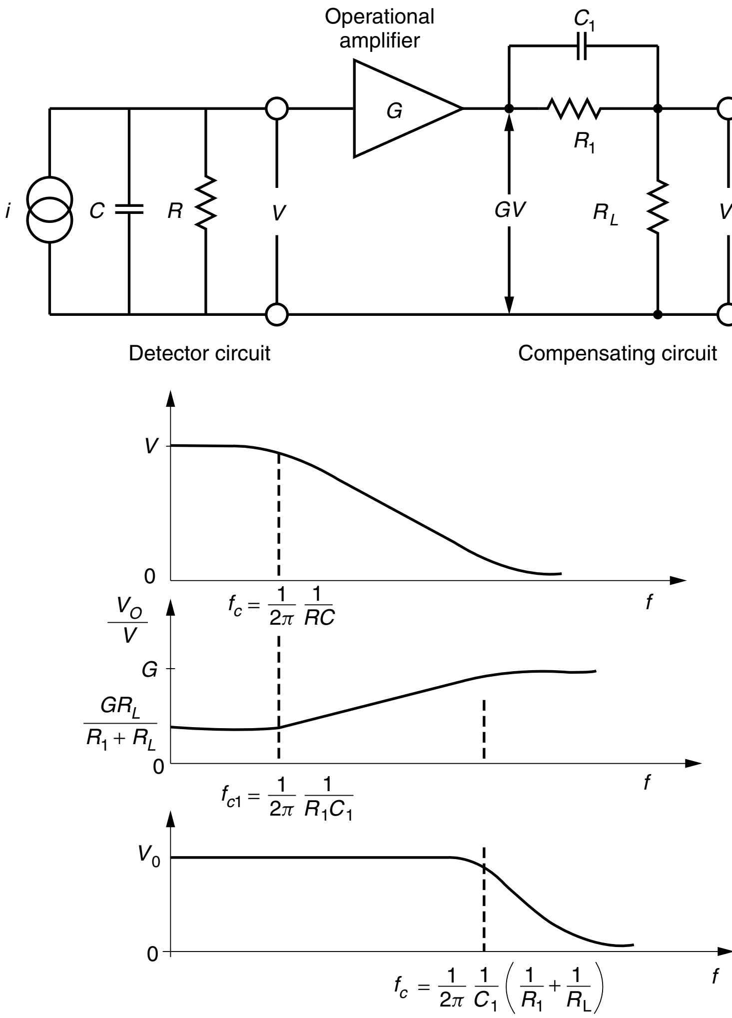


Figure 16.15 High impedance with a compensation circuit.

Proper compensation is achieved if the denominator of the first factor cancels the numerator of the second factor, namely,

$$R_1C_1 = RC \tag{16.17}$$

The condition, therefore, is that the cutoff frequency of the high-pass filter matches that of the low-pass filter. It follows from Eq. (16.17) that the cutoff frequency of the overall circuit is

$$f_c = \frac{1}{2\pi} \cdot \frac{1}{C_1}\left(\frac{1}{R_1} + \frac{1}{R_L}\right) \tag{16.18}$$

The advantages of such a preamplifier circuit are high sensitivity in a wide frequency band and low average square thermal current due to the high input impedance of the operational amplifier. The disadvantage, however, is that if f_c is stretched too far, the signal V fed into the operational amplifier becomes too small and soon reaches the noise level of the operational amplifier, causing the minimum detectable power to suffer.

Transimpedance Circuit (TZ) Another well-established method for widening the frequency band of a high input impedance amplifier is the use of a negative feedback circuit, such as the one shown in Fig. 16.16. The current i_f and the voltages V_0 and V are given by

$$i_f = \frac{V_O - \mathrm{V}}{R_f} \tag{16.19}$$

$$V = \frac{i + i_f}{1/R + j\omega C} \tag{16.20}$$

$$V_O = -GV \tag{16.21}$$

Observe that the input current to the operational amplifier is essentially zero.

Briefly explaining the operation, an increase in V_O increases i_f as shown by Eq. (16.19). An increase in i_f increases V across R and C. The increase in V then

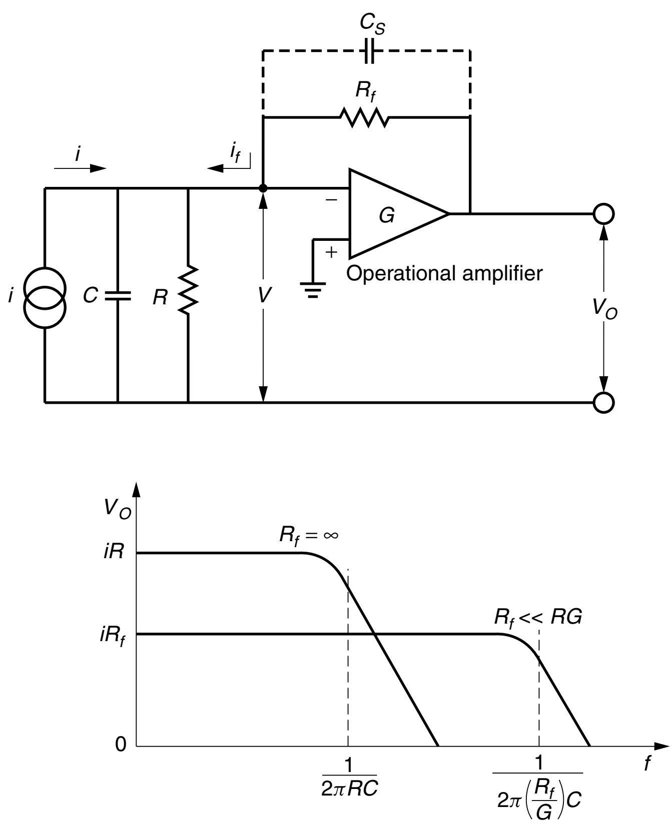


Figure 16.16 Transimpedance coupling.

decreases V_O. First, eliminating i_f from Eqs. (16.19) and (16.20) and then using Eq. (16.21) gives

$$V_O = -\frac{R_f i}{1 + \frac{1}{G} + \frac{R_f}{RG} + j\omega C \frac{R_f}{G}}$$

If one chooses

$$G \gg 1 \quad \text{and} \quad RG \gg R_f \tag{16.22}$$

then V_O becomes

$$V_O = -\frac{R_f i}{1 + j\omega C(R_f/G)} \tag{16.23}$$

Thus, the cutoff frequency is

$$f_c = \frac{1}{2\pi(R_f/G)C} \tag{16.24}$$

Inserting $R_f \ll RG$ into Eq. (16.24) gives

$$f_c \gg \frac{1}{2\pi RC}$$

This cutoff frequency is much higher than the cutoff frequency of Eq. (16.15) without compensation. One can raise f_c further by lowering R_f. The lowering of R_f, however, means lowering the output voltage, which is essentially $R_f i$ from Eq. (16.23).

As far as the output voltage is concerned, compared to the previously mentioned 50-Ω circuit with mismatch in Fig. 16.14, the transimpedance circuit gives a much larger output. If both circuits use an operational amplifier of gain G, the output from the 50-Ω circuit with mismatch is $50iG$. The output from the transimpedance circuit in Fig. 16.16 gives approximately iR_f and if $R_f > 50G$, then the transimpedance circuit gives a higher output.

Any stray capacitance C_s around the feedback resistance R_f severely affects the frequency bandwidth (Problem 16.7).

16.4.3.2 Coupling Circuits to a Preamplifier Above a Gigahertz

The high impedance (HZ) [20,21] and transimpedance (TZ) [22,23] approaches are still used up to few gigahertz, provided that ultrawide frequency band elements are used, and provided that sensor capacitance and stray capacitance are minimized. Above several gigahertz the distributed amplifier approach is used.

HZ and TZ with Special Considerations The stray capacitance or series resistance of the bonding wire or inductor wire should be minimized.

In order to increase the cutoff frequency as given by either Eq. (16.18) or (16.24), it is important to reduce C. For instance, by reducing the diameter of the sensor area of the photodiode to 30 μm, C can be reduced to 0.1 pF.

Operational amplifiers whose cutoff frequencies are normally subgigahertz can no longer be used in the gigahertz frequency range. These have to be replaced by ultrawide frequency band elements such as the GaAs metal–semiconductor field effect transistor

(GaAs MESFET), the high electron mobility transistor (HEMT), or the heterojunction bipolar transistor (HBT). Their cutoff frequencies are several tens of gigahertz. When an ultrawide frequency band element is used, it is important to install an adequately designed low-pass filter in order to prevent the noise spectra outside the required bandwidth from deteriorating the signal.

Distributed Amplifier Approach When the required frequency bandwidth exceeds several gigahertz, another approach such as the distributed preamplifier [24–26] has to be employed. Figure 16.17 shows a schematic diagram of distributed preamplifiers. The distributed FET amplifiers form two artificial transmission lines. The inductors L_g, which are connected to the gate terminals of the FETs and the gate–source capacitance C_{gs} of the FETs, form one transmission line with characteristic impedance $Z_g = \sqrt{L_g/C_{gs}}$, while the inductors L_d, which are connected to the drain terminals and the drain–source capacitors C_{ds}, form another transmission line with characteristic impedance $Z_d = \sqrt{L_d/C_{ds}}$.

A photodiode with the proper value of the load inductor L_p can smoothly be amalgamated into the artificial transmission line as an additional line element.

Each transmission line is loaded with impedances that match the characteristic impedance of the line in order to eliminate reflection for all signal frequencies. Thus, a wide frequency band operation is accomplished.

Referring to Fig. 16.17, there are two possible directions for the output of each FET to propagate along the drain transmission line; one is forward, and the other backward. First, consider only the backward wave in the vicinity of Q_1. There are two ways of reaching point d_1 from point g_1. The signal at g_1 is split into two components. One takes the direct route through Q_1 and the other takes an indirect route through Q_2 via g_1–g_2–Q_2–d_2–d_1. Both components go through the FET amplifier once, but their path lengths are different. By setting this path difference to be one-half wavelength, these two components cancel each other, and the backward wave is suppressed. As for the forward wave from g_1 to d_2, there is no difference between the path lengths

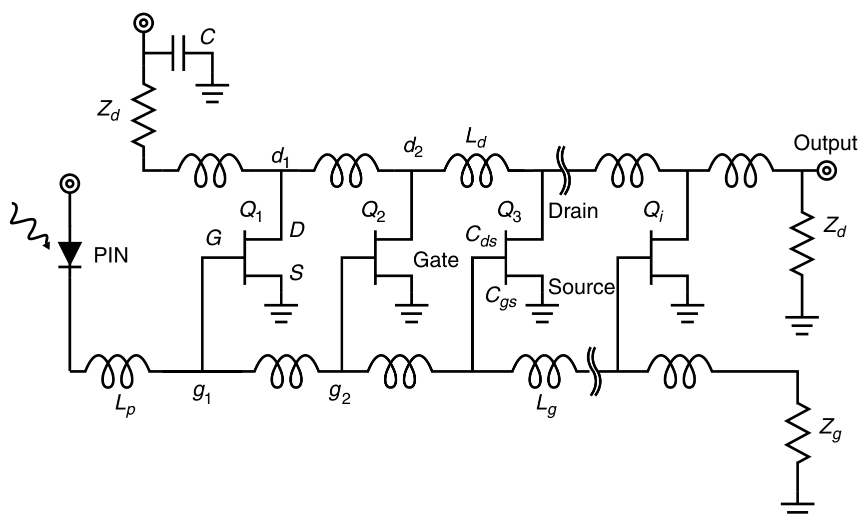


Figure 16.17 Diagram of distributed preamplifiers. (After A. P. Freundorfer and P. Lionais [25].)

of $g_1–g_2–d_2$ and $g_1 - d_1 - d_2$ and the two components add. The forward wave is enhanced.

The transmission line being artificial, it behaves like a low-pass filter and there is a cutoff frequency of

$$f_{cg} = \frac{1}{\pi\sqrt{L_g C_{gs}}} = \frac{1}{\pi Z_g C_{gs}} \tag{16.25}$$

and f_{cg} limits the frequency bandwidth of the receiver. The cutoff frequency f_{cg} can be increased by reducing Z_g but the thermal noise i_{th} (to be explained in Section 16.5.2) also increases with a reduction of Z_g. The thermal noise generated in the system is predominantly determined by Z_g and is approximately

$$i_{\text{th}}^2 = 4kTB/Z_g$$

Consequently, Z_g has to be set at a compromised value.

Figure 16.18 shows a photograph of a receiver consisting of a PIN photodiode and the distributed FET amplifiers. Such a distributed amplifier has a frequency bandwidth of 23 GHz. The gain of the amplifier is expressed in terms of the transimpedance

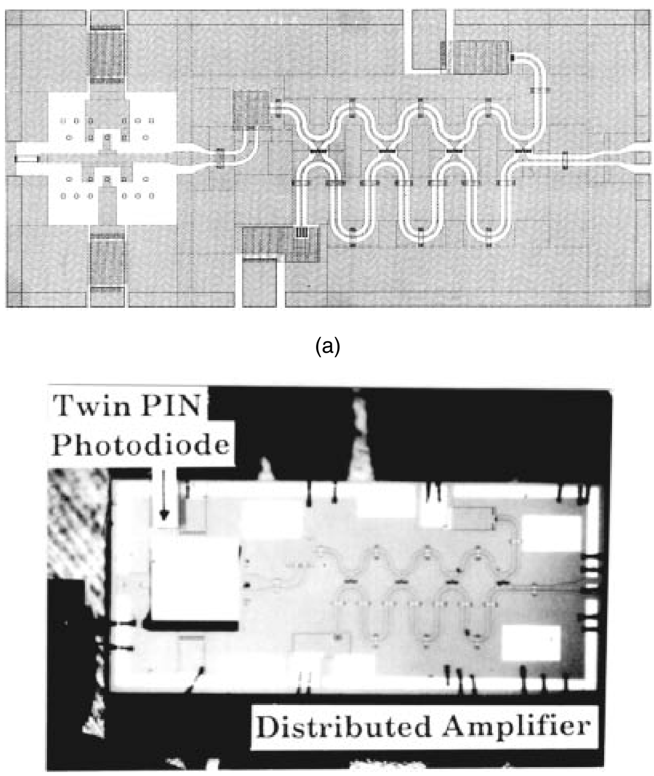


Figure 16.18 Balanced optical receiver. (a) Preamplifier pattern layout. (b) Realized optical receiver. (Courtesy of N. Takachio et al. [27].)

$Z_T = V_O/i$, where V_O is the output voltage and i is the photocurrent from the photodiode. The transimpedance is 46 dB $\cdot$ Ω (or approximately 40,000 Ω) [27].

16.5 NOISE IN THE DETECTOR SYSTEM

The minimum detectable power of the receiver is determined from noise considerations. The two major sources of noise are shot noise and thermal noise.

16.5.1 Shot Noise

The flow of electrons across the p-n junction is like pouring a bucket of beans into another bucket. Each bean arrives discretely and randomly. Each discrete electron is equivalent to a δ function of current whose Fourier transform has a constant spectrum in the frequency domain, giving a white noise characteristic. The quantity used to characterize the shot noise is the mean square value of the noise current

$$\langle i_{\text{shot}}^2 \rangle = 2eIB \tag{16.26}$$

where B is the bandwidth of the receiver and I is the average current through the detector. The shot noise current is $\sqrt{2eIB}$ and increases with $\sqrt{B}$. The shot noise current per $\sqrt{\text{Hz}}$ is $\sqrt{2eI}$.

Flow of electron ''beans'' generates shot noise.

Not only the discreteness but also the randomness of the arrival time is important. The number of electrons in I amperes of flow for 1 second is $N = I/e$. If these electrons were precisely equally spaced, then a periodic flow of current would be generated with period (e/I) s that can be expressed as $\Pi\Pi(t/(e/I))$ (see Section 1.4.6). The fundamental frequency component of such a hypothetical current would be $f_0 = I/e$. The fundamental frequency f_0 is already a big number. For instance, if $I = 10^{-6}$ A, then $e = 1.59 \times 10^{-19}$ coulomb gives $f_0 = 6300$ GHz. There would not be any spectrum between dc and 6300 GHz!

16.5.2 Thermal Noise

Any heated substance emits radiation. Even at room temperature, there will be some thermal radiation, and hence an associated thermal noise. The only way to shut off thermal noise completely is to cool the substance to absolute zero on the Kelvin scale.

The thermal noise power per unit spectrum is kT, and the mean square of the noise current generated inside the resistor is

$$\langle i_{th}^2 \rangle = 4kTB/R \tag{16.27}$$

where R is the resistance of the resistor considered as the noise source, k is Boltzmann's constant 1.38×10^{-23} J·K, T is the temperature of the resistor on the Kelvin scale, and B is the bandwidth. We will sidetrack briefly to explain how Eq. (16.27) was derived.

When a piece of resistor is heated, it becomes an electromagnetic wave radiator. For that matter, any material becomes a good radiator when heated. The radiating power spectrum contained in the frequency interval between ν and $\nu + \Delta\nu$ is given by Planck's factor

$$p(\nu) = \frac{h\nu}{e^{h\nu/kT} - 1} \tag{16.28}$$

whose maximum value shifts with temperature.

Now the heated resistor is connected to a load resistor, which extracts energy from the heated source resistor. When the resistor is piping hot, the mechanism of energy transfer is primarily due to blackbody radiation in free space, which radiates and also heats up the load resistor. The amount of energy transfer does not depend on the value of the resistance of the load; it primarily depends on the effective cross-sectional area and the emitted infared wavelength of the heated resistor. Even though the amount of energy transfer to the load can be optimized by shaping and "coloring" the load resistor, the one invariant parameter that cannot be changed is the maximum available power from the source. This is set by Planck's factor $p(\nu)\,d\nu$.

The main difference between the cooled down and piping hot case is the means of energy transfer to the load resistor. Next, the case when the source resistor is cooled down to room temperature is considered. The resistor still continues to radiate but the amount of energy radiated is reduced and its radiated spectrum is shifted toward a lower frequency. The shifted frequency spectrum is now too low to radiate effectively. The noise power, however, can conduct very effectively through the connecting copper wire to the load as a noise current. In this case, the amount of transfer of energy depends on the resistance of the load while the maximum available power is again determined by Planck's factor, Eq. (16.28).

Next, a mathematical model for an equivalent circuit of the resistor as a noise source is developed. The model is a resistor with the same resistance but paralleled with a constant noise current source as encircled in Fig. 16.19. The value of the noise generator current i_N is determined by setting the maximum available power to a fictitious load R_1 equal to the invariant value of $p(\nu)\,d\nu$ as follows.

With the equivalent circuit in Fig. 16.19, the transferred power P_1 to the load is

$$P_1 = \left\langle \left(i_N \frac{R}{R + R_1} \right)^2 \right\rangle R_1 \tag{16.29}$$

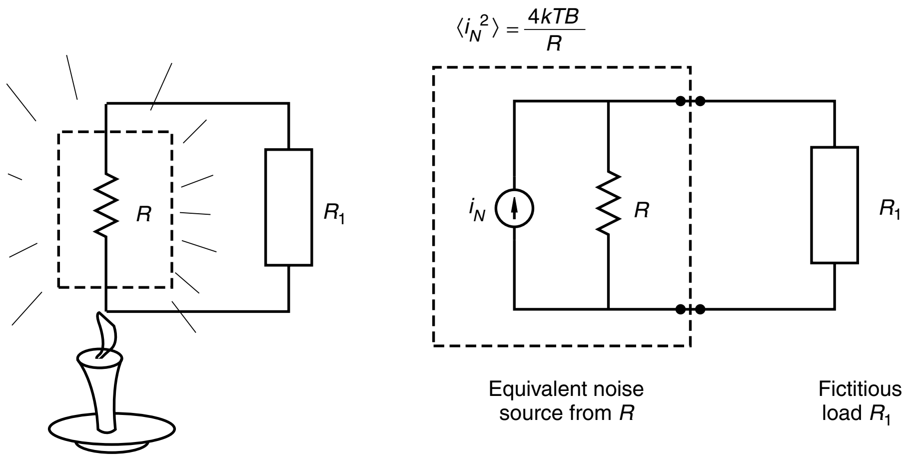


Figure 16.19 The equivalent circuit of noise from resistor R. $\langle i_N^2 \rangle$ is determined such that the noise source delivers a noise power of kTB when R_1 is selected to be R.

The maximum available power is found from $dP_1/dR_1 = 0$ and occurs when $R_1 = R$. Therefore, the maximum available power becomes

$$\frac{\langle i_N^2 \rangle}{4} R = p(\nu)\, d\nu \tag{16.30}$$

When

$$\frac{h\nu}{kT} \ll 1$$

Eq. (16.28) can then be approximated as

$$p(\nu)\, d\nu = kT\, d\nu \tag{16.31}$$

At room temperature, this condition is satisfied for

$$\nu \ll 6.2 \text{ THz} \tag{16.32}$$

Finally, inserting Eq. (16.31) into (16.30) gives the noise current generated from the resistor R in bandwidth B expressed by Eq. (16.27).

16.5.3 Signal to Noise Ratio

The signal to noise ratio S/N is used to quantitatively represent the quality of the detecting system. Initially, the incident light is assumed to be unmodulated. The detection system is a direct detection system such as shown in Fig. 16.13a with its equivalent circuit in Fig. 16.13b.

With incident light power P_s of the PIN diode, the output electrical signal current i_s is, from Eq. (12.8),

$$i_s = \frac{\eta e}{h\nu} P_s \tag{16.33}$$

The electrical signal power S delivered to the load R_L is $S = i_s^2 R_L$ and

$$S = \left(\frac{\eta e}{h\nu}\right)^2 P_s^2 R_L \tag{16.34}$$

The shot noise power N_{shot} is, from Eq. (16.26),

$$N_{\text{shot}} = 2e\left(\frac{\eta e}{h\nu} P_s + I_d\right) R_L B \tag{16.35}$$

where I_d is the dark current and is essentially equal to I_{so} in Eq. (16.10). The thermal noise power is generated in the load resistor. The noise power N_{th} generated from resistance R_L is

$$N_{\text{th}} = \frac{4kTB}{R_L} R_L = 4kTB \tag{16.36}$$

The signal to noise ratio of the PIN diode taking both shot and thermal noise into consideration, therefore, is

$$\frac{\text{S}}{\text{N}} = \frac{\left(\dfrac{\eta e}{h\nu}\right)^2 P_s^2}{2e\left(\dfrac{\eta e}{h\nu} P_s + I_d\right) B + \dfrac{4kTB}{R_L}} \tag{16.37}$$

As seen from Eq. (16.37), the behavior of the S/N with regard to P_s depends on whether shot noise or thermal noise is predominant. When the shot noise contribution is predominant, the S/N is said to be *shot noise limited or quantum limited.* On the other hand, when the thermal noise is predominant, it is said to be *thermal noise limited*.

For most PIN diodes, $\eta e/h\nu = 0.5$ A/W and $I_d = 2$ nA, and as long as P_s is larger than 4 nW, the contribution of I_d in the parentheses in the denominator of Eq. (16.37) is not important. When P_s is smaller than 4 nW, a more expensive low dark current photodiode is necessary to detect the light.

In the quantum-limited case, Eq. (16.37) becomes

$$\frac{\text{S}}{\text{N}} = \frac{\eta}{2h\nu B} P_s \tag{16.38}$$

where the assumption was made that

$$P_s \gg 4 \text{ nW} \tag{16.39}$$

Note that the S/N increases linearly with P_s and decreases linearly with B. In order to maintain the same S/N for a wider frequency of operation, P_s must be raised accordingly.

Next, in the thermal-noise-limited case, Eq. (16.37) becomes

$$\frac{\text{S}}{\text{N}} = \frac{\left(\dfrac{\eta e}{h\nu}\right)^2}{\dfrac{4kTB}{R_L}} P_s^2 \tag{16.40}$$

This time the signal to noise ratio increases quadratically with P_s and is more sensitive to P_s than in the quantum-limited case, and it increases with R_L while maintaining the same dependence on B. In the thermal-noise-limited case, a low dark current photodiode is completely unnecessary.

16.5.4 Excess Noise in the APD

As shown in Fig. 12.5b, the APD consists of a photoelectric converter layer and a multiplier layer. The noise from the APD is normally characterized by a combination of the noise in these two layers.

When the APD is back-biased, the internal resistance is on the order of megohms. The shot noise current generated in the photoelectric converter layer, $i_{\text{shot}} = \sqrt{2eIB}$, is also multiplied by the multiplication factor M in the multiplier layer; thus, the mean square value of the shot noise from the APD becomes

$$i_{\text{shot}}^2 = 2eB\left(\frac{\eta e}{h\nu}P_s + I_d\right)M^{2+x} \tag{16.41}$$

Note that Eq. (16.41) contains an extra factor M^x, which is called the excess noise (x is called the excess noise index). The excess noise is due primarily to the noise generated during the avalanche effect. The value of x is determined by the constituent materials of the APD. For example, $x \approx 0.3$–0.5 for silicon and $x \approx 0.9$–1.0 for germanium. Sometimes, a high-quality PIN diode is more desirable than an APD in the presence of a high excess noise index.

The signal to noise ratio of an APD is

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\left(\dfrac{\eta e}{h\nu}P_s M\right)^2}{2eB(I_s + I_d)M^{(2+x)} + 4kTB/R_L} \tag{16.42}$$

where $I_s = \eta e/h\nu P_s$. Note that I_s is the average signal current before avalanche multiplication. From this equation we see that although the output power increases with M^2, the shot noise increases more rapidly with M^{2+x}; thus, the S/N starts to decrease beyond a certain value of M as shown in Fig. 16.20. A larger M is not necessarily better, and the back-biased voltage should be adjusted for the proper value of M, which is normally a few hundred.

16.5.5 Noise Equivalent Power (NEP)

The minimum detectable light power is a quantity that is often used for designing optical communication systems. The minimum detectable light power $P_{s\,\text{min}}$ is defined as the unmodulated light level for which the output electrical signal can barely be detected. This value can be obtained by setting the S/N in Eq. (16.42) equal to unity, giving

$$P_{s\,\text{min}} = \frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}}\sqrt{B} \tag{16.43}$$

where

$$I_s \ll I_d$$

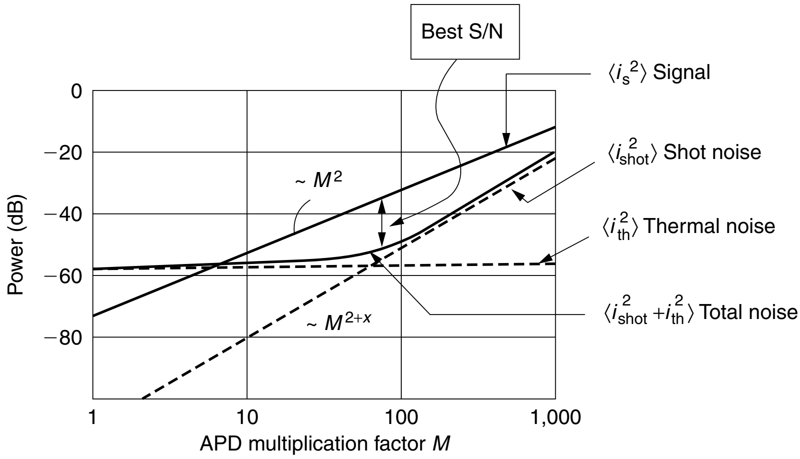


Figure 16.20 Dependence of signal and noise powers of an APD on the multiplication factor M.

was assumed. The NEP is defined as

$$P_{s\,\min} = \text{NEP} \cdot \sqrt{B} \tag{16.44}$$

hence,

$$\text{NEP} = \frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}} \quad \text{W}/\sqrt{\text{Hz}} \tag{16.45}$$

The NEP is used to quantify the sensitivity of the detector and is the minimum detectable power for a 1-Hz signal bandwidth. The strange unit of W/$\sqrt{\text{Hz}}$ is used because the minimum detectable power is proportional to the square root of the frequency band.

Example 16.1 A detection system employs a silicon avalanche photodiode Type S2383 listed in the Hamamatsu Photonics catalogue. The characteristics of the APD are given in Fig. 16.21. The operational parameters are as follows. The input light power P_s (unmodulated) is 10 nW, and the wavelength λ is 0.83 μm. The cutoff frequency f_c is 50 MHz, and the temperature is 20°C.

Assume that the APD characteristics given in Fig. 16.21 for 25°C are the same for 20°C. Answer the following questions.

(a) What is the responsivity R? Responsivity was defined in Section 12.4.2 as

$$R = \frac{\eta e}{h\nu} M$$

(b) What is the relevant loaded resistance R_L of the APD?

(c) What is the electrical signal power, the shot noise power, and the thermal noise power?

(d) What is the S/N?

(e) Is the S/N quantum limited or thermal noise limited?

(f) What is the value of the NEP?

Si APDs (Low-bias operation types, for 800 nm range)

Type No.	Spectral response range λ (nm)	Peak sensitivity wavelength λp (nm)	Photo *4 sensitivity M = 1 λ = 800 nm (A/W)	Quantum Efficiency M = 1 λ = 800 nm (%)	Dark current *4 I_D (nA) Typ.	Max.	Cut-off *4 frequency f_c $R_L = 50\ \Omega$ (MHz)	Terminal *4 capaci-tance Ct (pF)	Excess *4 noice figure x (pF)	Gain M λ = 800nm	Type No.
S2381	400 to 1000	800	0.5	75	0.05	0.5	100	1.5	0.3	100	S2381
S2382					0.1	1	900	3			S2382
S5139											S5139
S2383					0.2	2	600	6			S2383
S2383-10 *2											S2383-10
S3884					0.5	5	400	10			S3884
S2384					1	10	120	40		60	S2384
S2385					3	30	40	95		40	S2385

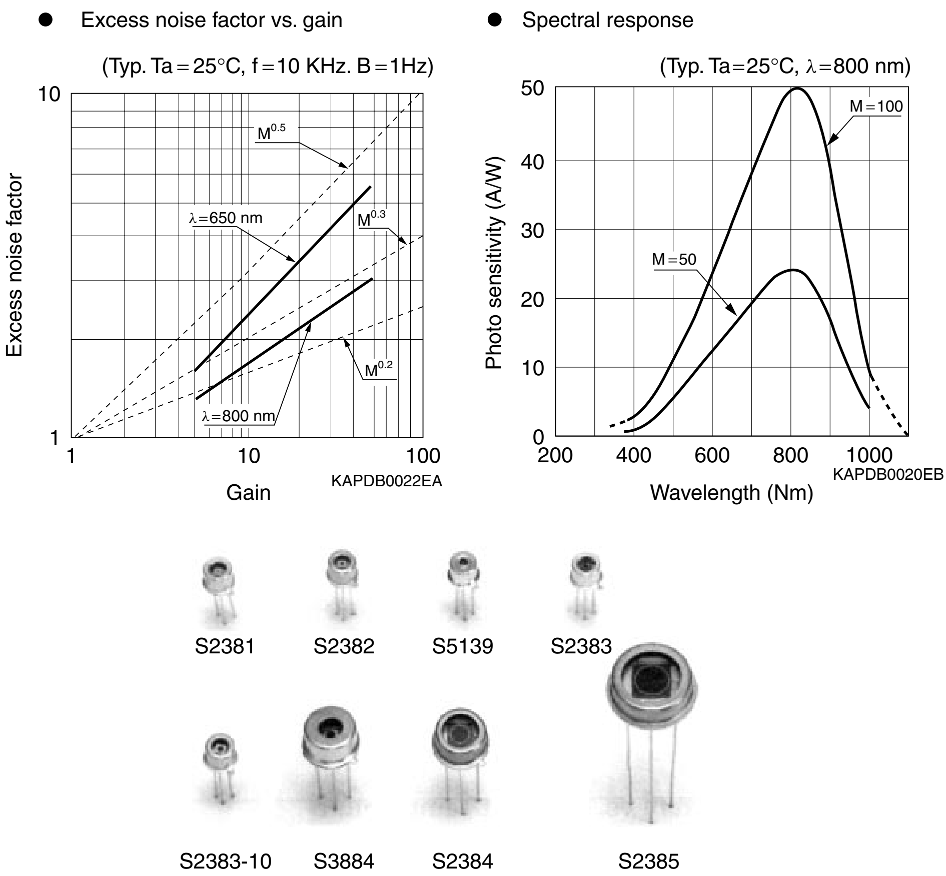


Figure 16.21 APD characteristics. (From Hamamatsu Photonics catalogue [19].)

Solution The relevant parameters in Fig. 16.21 are:

$$P_s = 10\ \text{nW}$$
$$\eta = 0.75$$
$$I_d = 0.2\ \text{nA}$$
$$C_f = 6\ \text{pF}$$

$x = 0.3$
$M = 100$
$T = 293$ K

(a) The responsivity of the APD is

$$R = \frac{\eta e}{h\nu} M = \frac{(0.75)(1.6 \times 10^{-19})(100)}{(6.63 \times 10^{-34})(3.6 \times 10^{14})} = 50.2 \text{ A/W}$$

(b) The load impedance for $f_c = 50$ MHz is, from Eq. (16.15),

$$R_L = \frac{1}{2\pi f_c C_f} = \frac{1}{2\pi(5 \times 10^7)(6 \times 10^{-12})} = 531\ \Omega$$

(c) The electrical signal power is

$$\begin{aligned} S &= \left(\frac{\eta e}{h\nu} M P_s\right)^2 R_L = (RP_s)^2 R_L \\ &= [(50.2)(10^{-8})]^2(531) = 1.34 \times 10^{-10}\ W \end{aligned}$$

From Eq. (16.41), the shot noise power is

$$\begin{aligned} N_{\text{shot}} &= 2eB\left(\frac{\eta e}{h\nu} P_s + I_d\right) M^{2+x} R_L \\ &= 2(1.6 \times 10^{-19})(5 \times 10^7)\{(0.502)(10^{-8}) + 0.2 \times 10^{-9}\}(100^{(2+0.3)})(531) \\ &= 1.77 \times 10^{-12}\ W \end{aligned}$$

From Eq. (16.36), the thermal noise power is

$$\begin{aligned} N_{\text{th}} &= 4kTB \\ &= 4(1.38 \times 10^{-23})(293)(5 \times 10^7) \\ &= 8.09 \times 10^{-13}\ W \end{aligned}$$

(d) S/N $= S/(N_{\text{shot}} + N_{\text{th}}) = 1.34 \times 10^{-10}/1.77 \times 10^{-12} + 8.09 \times 10^{-13} = 52.0$
(e) It is quantum limited.
(f) From Eq. (16.45), the NEP is

$$\begin{aligned} \text{NEP} &= \frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}} \\ &= \frac{1}{0.502}\sqrt{2(1.6 \times 10^{-19})(0.2 \times 10^{-9})(100)^{0.3} + \frac{4(1.38 \times 10^{-23})(293)}{100^2(531)}} \\ &= 1.15 \times 10^{-13}\ \text{W}/\sqrt{Hz} \end{aligned}$$

□

Example 16.2

(a) The same system as given in Example 16.1 is to be used at much lower input light power. What is the minimum detectable light power?

(b) The cutoff frequency of the system described in Example 16.1 was lowered to 5 MHz and the load impedance R_L was raised accordingly. What is the new value of S/N?

Solution

(a) From Eq. (16.44), the minimum detectable light power is

$$P_{\min} = \text{NEP}\sqrt{B} = (1.15 \times 10^{-13})(\sqrt{5 \times 10^7}) = 0.81 \text{ nW}$$

(b) By reducing $f_c (= B)$ to $f_c/10$, not only is the noise bandwidth reduced to $1/10$, but R_L is also increased by 10 times. From result (d) of Example 16.1,

$$\frac{\text{S}}{\text{N}} = \frac{1.34 \times 10^{-10}(10)}{1.77 \times 10^{-12}\left(\frac{1}{10}\right)(10) + 8.09 \times 10^{-13}\left(\frac{1}{10}\right)} = 724$$

and the S/N is increased about 24 times. □

16.5.6 Signal to Noise Ratio for ASK Modulation

What modifications should be made if the intensity of the incident light is modulated? The effect of modulation on signal to noise ratio is examined for amplitude shift keying (ASK) mentioned in Section 16.2.7. The current from the detector Eq. (12.8) has to be modified as

$$i_s(t) = \frac{\eta e}{h\nu} M P_s A(t) \tag{16.46}$$

where $A(t)$ is the ASK modulation and is either 1 or 0.

The average square signal current, therefore, is

$$\langle i_s^2 \rangle = \frac{1}{2}\left(\frac{\eta e}{h\nu} M P_s\right)^2 \tag{16.47}$$

where use was made of the relationship

$$[A(t)]^2 = A(t) \tag{16.48}$$

and the average of the "0"s and the "1"s of $A(t)$ is $1/2$.

It is important to remember that the signal electrical power $\langle i_s^2 \rangle R_L$ is proportional to the square of the signal optical power P_s, where P_s is the power for the "1" bit.

The signal to noise ratio of an APD is

$$\frac{\text{S}}{\text{N}} = \frac{\frac{1}{2}\left(\frac{\eta e}{h\nu} P_s M\right)^2}{2eB(I_s + I_d)M^{(2+x)} + 4kTB/R_L} \tag{16.49}$$

where $I_s = \frac{1}{2}\eta e/h\nu P_s$ and I_d is the amplitude of the dark current.

Another point is that I_s is the average signal current before avalanche multiplication. For the ASK modulation it has a factor of $\frac{1}{2}$ because the light is off half the time.

The minimum light power to establish S/N = 1 for the ASK modulated light is

$$P_{s\,\min}^{\mathrm{ASK}} = \sqrt{2}\frac{h\nu}{\eta e}\sqrt{2eI_d M^x + \frac{4kT}{M^2 R_L}}\sqrt{B} \tag{16.50}$$

which is $\sqrt{2}$ times that of the unmodulated incident light given by Eq. (16.43). It might be added that the NEP is always defined with unmodulated light.

16.5.7 Signal to Noise Ratio of Homodyne Detection

The output signal current from homodyne detection ($f_{\mathrm{IF}} = 0$) without modulation is given by Eq. (12.22). If the signal is ASK modulated, the output electrical signal power is

$$S = \left(2\frac{\eta e}{h\nu}\right)^2 P_S P_L M^2 \overline{A(t)^2} \tag{16.51}$$

and the signal to noise ratio of homodyne detection for ASK coded signal light is

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\left(\dfrac{2\eta e}{h\nu}\right)^2 P_s P_L M^2 \overline{A(t)^2}}{2eB_{IF}\left(\dfrac{\eta e}{h\nu}P_L + I_d\right)M^{2+x} + \dfrac{4kTB_{IF}}{R_L} + N_{\mathrm{LO}}} \tag{16.52}$$

N_{LO} is the local oscillator intensity noise (LOIN) power, which is the noise due to the fluctuation of the local oscillator intensity. This, however, can be removed by using the balanced mixer, as mentioned in Section 12.7. For a large local oscillator laser power P_L, the thermal noise power can be ignored and the detector becomes quantum limited, almost regardless of R_L, and can provide the ultimate limit in sensitivity. Equation (16.52) can be approximated as

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\eta}{M^x}\left(\frac{P_s}{h\nu}\right)\frac{1}{B_{\mathrm{IF}}} \tag{16.53}$$

where use was made of $\overline{A^2(t)} = \overline{A(t)} = \frac{1}{2}$. Note that the IF bandwidth B_{IF} is twice the bandwidth B of the signal, as mentioned in Section 16.3.2.

Comparing the signal to noise ratios of direct detection by an APD of Eq. (16.49) to that of homodyne detection of Eq. (16.53), one realizes that the S/N of the direct detection is proportional to P_s^2 in the lower limit that the I_s term in the denominator of Eq. (16.49) is small, whereas the signal to noise ratio of the homodyne detection is proportional to P_s. This difference is significant for the commonly small values of P_s.

Let us examine Eq. (16.53) more closely. The factor in the parentheses is the number of photons coming in per unit time. As mentioned above, $B_{\mathrm{IF}} = 2B$. From Section 16.6.5, the bit rate B_t for the NRZ code is $B = B_t/2$ and $B_{\mathrm{IF}} = B_t$. The number of electrons contained in one pulse is $(P_s/h\nu)(1/B_t)$. If η/M^x is assumed to be a typical value such as 0.4, then we can conclude that only 2.2 photons in one pulse is good enough to achieve a signal to noise ratio of 1!

16.5.8 Borderline Between the Quantum-Limited and Thermal-Noise-Limited S/N

It is useful to know the borderline between the regions of the quantum-limited and thermal-noise-limited S/N to facilitate the calculation of the required light power for a given S/N. The borderline is calculated from the denominator of Eq. (16.42) by setting $N_{\text{shot}} = N_{\text{th}}$:

$$2e\frac{\eta e}{h\nu}M^2 P_s^a = \frac{4kT}{R_L} \tag{16.54}$$

where the contributions of the excess noise and dark current are ignored.

Note that in Eq. (16.42), I_s is the time average current. A time average is necessary because the encoded light is off some of the time, and during the off time, no shot noise is generated. Thus, P_s^a (which is the average light power received by the PIN diode or APD) rather than P_s (which is the peak light power or power for the "on" state) has to be used.

Inserting the physical constants and $\eta e/h\nu = 0.5$, $x = 0$, and $T = 293$ K, we can say that the separation between the quantum-limited and thermal-noise-limited regions is roughly as follows:

$$\begin{aligned} P_s^a M^2 &> \frac{0.1}{R_L} \quad \text{quantum limited} \\ P_s^a M^2 &< \frac{0.1}{R_L} \quad \text{thermal noise limited} \end{aligned} \tag{16.55}$$

Thus, the regions are determined by the relative values of P_s^a and $1/R_L$. Generally, analog systems are more often quantum limited and digital systems are usually thermal noise limited. The reason for this is that analog systems need a large P_a^s to meet the requirement of a large S/N, whereas digital systems do not need a large S/N or P_a^s, but rather have small R_L for a higher bit rate. In Fig. 16.22, the borderline given by Eq. (16.55) is plotted with average received power P_s^a as a function of load resistance R_L. The solid lines represent a PIN diode with $M = 1$, and the dashed line represents an APD with $M = 10$, $x = 0$ and with $M = 100$, $x = 0$. The region of lower R_L with lower P_s^a is the thermal-noise-limited S/N and that of higher R_L with higher P_s^a is the quantum-limited S/N. The quantum-limited region expands with increasing M.

The choice of the load resistance R_L is made according to the cutoff frequency f_c of the detection system. When the system needs a wide frequency bandwidth B of operation, the $R_L C$ time constant given by Eq. (16.15) has to be reduced so that the cutoff frequency $f_c = 1/2\pi R_L C$ is larger than B. But with a smaller R_L, the gain of the system suffers, as shown by Eq. (16.13). The value of $f_c = 1/2\pi R_L C$ for $C = 1.6$ pF is included in the bottom horizontal axis in Fig. 16.22.

16.5.9 Relationship Between Bit Error Rate (BER) and Signal to Noise Ratio

In the field of digital signal communication, the quality of the digital signal is evaluated by the number of error bits over the total number of bits transmitted. This relationship

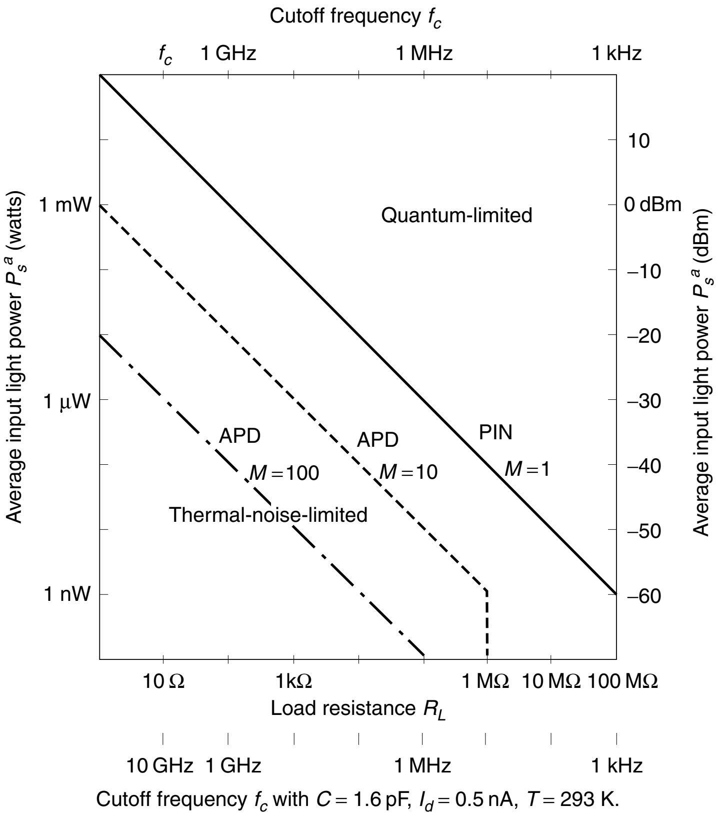


Figure 16.22 Borderline between quantum-limited and thermal-noise-limited cases of PIN and APD photodiodes.

is called the bit error rate (BER). The noisier the received signal is, the more the detection is subject to error, increasing the BER of the system. The digital system uses BER while the analog system uses S/N. The relationship that exists between BER and S/N will now be presented [28,29].

A pulse that contains noise in the "1" state is shown in Fig. 16.23. (Noise in the "0" state is not drawn for clarity.) If the probability density function $p(v)$ that the voltage is between v and $v + dv$ is Gaussian, then

$$p(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_1} e^{-1/2((v-s)/\sigma_1)^2} \tag{16.56}$$

where s is the voltage representing "1," for example, 5 V for TTL (transistor−transistor logic).

The probability P_{10} of mistaking "1" as "0" is represented by the area of the shaded portion in Fig. 16.23 and is equal to

$$P_{10} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_1} \int_{-\infty}^{V_{\text{th}}} e^{-1/2((v-s)/\sigma_1)^2} dv \tag{16.57}$$

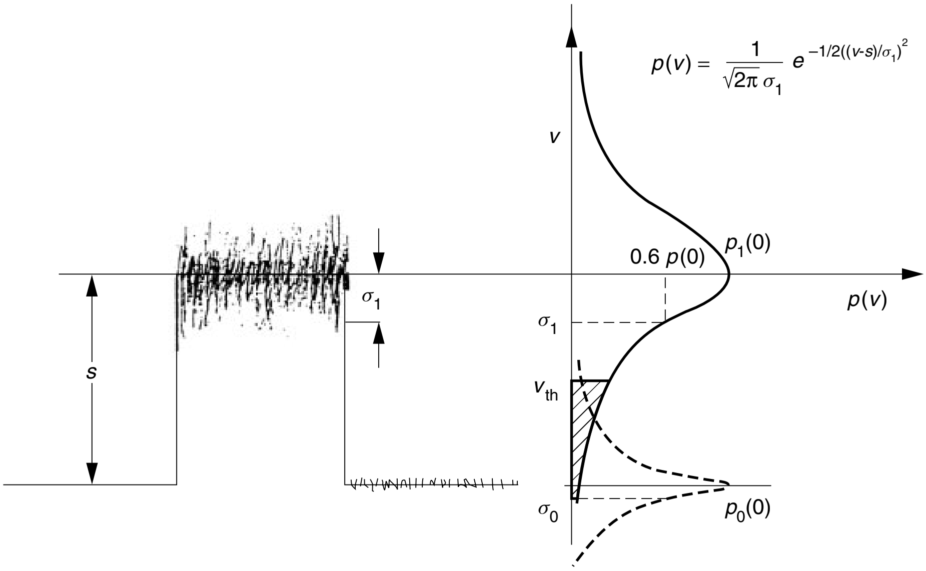


Figure 16.23 Pulse with Gaussian noise in "1." The noise in "0" is suppressed for clarity.

Setting

$$\lambda = \frac{s - v}{\sqrt{2}\sigma_1} \tag{16.58}$$

Eq. (16.57) is rewritten as

$$P_{10} = \frac{1}{\sqrt{\pi}} \int_{(s-v_{\text{th}})/\sqrt{2}\sigma_1}^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.59}$$

Similarly, the probability P_{01} of mistaking "0" as "1" is represented by

$$P_{01} = \frac{1}{\sqrt{\pi}} \int_{v_{\text{th}}/\sqrt{2}\sigma_0}^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.60}$$

where σ_0 is the rms of the noise superimposed on "0." If one-half of the signal is "1" and the other half is "0," then

$$\text{BER} = \tfrac{1}{2}P_{10} + \tfrac{1}{2}P_{01} \tag{16.61}$$

The skew rate is defined as the number of errors caused by reading "1" as "0" divided by the total number of errors. By changing the threshold value v_{th}, the skew rate changes. When v_{th} is set too high, the skew rate becomes larger than 0.5. A skew rate of 0.5 approximately provides the condition for the minimum error rate. For the condition that provides exactly the minimum BER, see Problem 16.8. If the skew rate is set to 0.5, then

$$P_{10} = P_{01} \tag{16.62}$$

It follows that

$$\frac{s - v_{\text{th}}}{\sigma_1} = \frac{v_{\text{th}}}{\sigma_0} \tag{16.63}$$

and

$$v_{\text{th}} = \frac{\sigma_0}{\sigma_1 + \sigma_0} s \tag{16.64}$$

and from Eqs. (16.60)–(16.64)

$$\text{BER} = \frac{1}{\sqrt{\pi}} \int_{s/\sqrt{2}(\sigma_1+\sigma_0)}^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.65}$$

Equation (16.65) can be rewritten using the definition of the complementary error function

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-\lambda^2}\, d\lambda \tag{16.66}$$

Thus,

$$\text{BER} = \tfrac{1}{2}\text{erfc}(Q) \tag{16.67}$$

where

$$Q = \frac{s}{\sqrt{2}(\sigma_0 + \sigma_1)} \tag{16.68}$$

There is no closed form of integration of Eq. (16.66), but the calculated curve is available in Fig. 16.24.

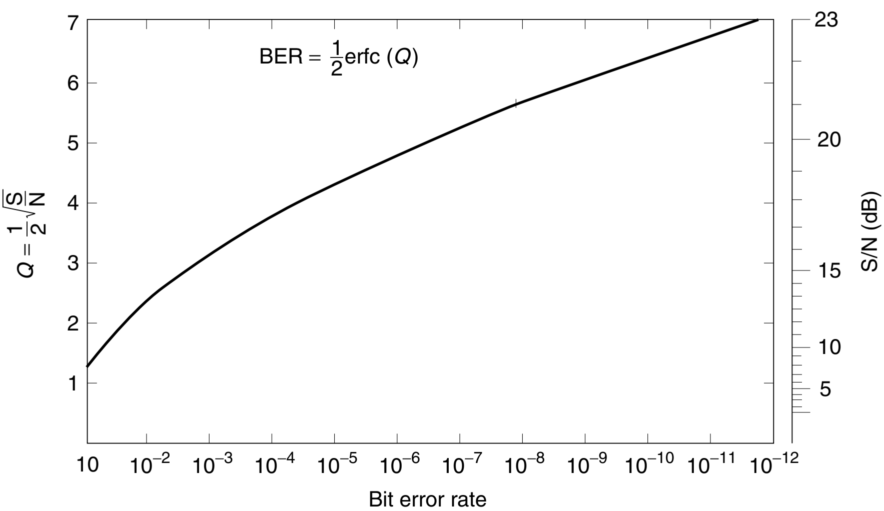


Figure 16.24 Bit error rate versus Q and S/N in dB.

An approximate expression for $\mathrm{erfc}(x)$, which is good for the entire range of x, is [30]

$$\tfrac{1}{2}\,\mathrm{erfc}(x) = \frac{1}{(1-a)x + a\sqrt{x^2+b}}\,\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \tag{16.69}$$

where

$$a = 1/\pi$$

$$b = 2\pi$$

$$0 < x < \infty$$

For $x^2 \gg 2\pi$, Eq. (16.69) reduces to

$$\tfrac{1}{2}\,\mathrm{erfc}(x) = \frac{1}{\sqrt{2\pi}x}e^{-x^2/2}$$

It should be noted that σ_0 and σ_1 are not necessarily the same. For instance, in the case of direct detection, shot noise due to the signal current is present for “1” but absent during “0.”

With some approximation, the value of Q will be represented in terms of the S/N. Let us assume that

$$\sigma_0 = \sigma_1 = \sigma \tag{16.70}$$

and that the amplitude of the ASK signal for “1” is s volt. For a 50% duty cycle, the signal power S is

$$S = \tfrac{1}{2}s^2/R_L \tag{16.71}$$

σ^2 represents the average value of $\overline{(v-s)^2}$ and the noise power is

$$N = \sigma^2/R_L \tag{16.72}$$

From Eqs. (16.68) to (16.72) this means that

$$Q = \tfrac{1}{2}\sqrt{\frac{\mathrm{S}}{\mathrm{N}}} \tag{16.73}$$

For instance, from Fig. 16.24, in order to achieve BER $= 10^{-9}$, $Q = 6$ is required. The corresponding S/N from Eq. (16.73) is S/N $= 144$ or 21.6 dB.

The required BER depends on the kind of information being sent. Roughly speaking, BER $= 10^{-2}$ is sufficient for a voice channel. BER $= 10^{-4}$ is required for data transmission and BER $= 10^{-9}$ is necessary for computer communication.

Even a relatively noisy transmission system of S/N $= 144$ (S/N $= 5000$ is a typical requirement of an analog signal transmission) can transmit the signal at a BER of 10^{-9} if the signal is first digitized and transmitted. Thus, by using digital modulation, even a poor performance transmission system can be used. The drawback of the digital system is the demand for wider frequency bandwidth, which is, of course, the allure of the large information-carrying capacity of the optical fiber. Another advantage of digital systems is that the effects of not only the noise but also the pulse shape distortion are not as critical as for analog systems.

As a final note, it should be recognized that the expression for the BER depends on the discriminator mechanism that differentiates between "0" and "1" signals. The ASK discriminator uses the threshold voltage of the receiver output, while the FSK and PSK discriminators use the outputs between the two frequencies and the outputs between the two states of phases of the light, respectively. Rigorously speaking, the expressions are slightly different.

Example 16.3 A PIN diode whose NEP is 3.9×10^{-13} W/$\sqrt{\text{Hz}}$ and whose sensitivity is thermal noise limited is used as a receiver. Find the required incident light power to the PIN diode to maintain the BER $= 10^{-9}$. The frequency bandwidth of the PIN diode is 300 MHz.

Solution

Method I. From the BER curve in Fig. 16.24, the required S/N for BER $= 10^{-9}$ is S/N $= 144$.

For unmodulated light, the light power $P_{s\,\min}$ needed for S/N $= 1$ is

$$\begin{aligned} P_{s\,\min} &= NEP\sqrt{B} \\ &= (3.9 \times 10^{-13})(\sqrt{3 \times 10^{8}}) \\ &= 6.76 \times 10^{-9}\ \text{W} \\ &= -51.7\ \text{dBm} \end{aligned}$$

For the ASK modulated light, the corresponding necessary light power is $\sqrt{2}$ times larger:

$$\begin{aligned} P_{s\,\min}^{\text{ASK}} = \sqrt{2}P_{s\,\min} &= 9.56 \times 10^{-9}\ \text{W} \\ &= -50.2\ \text{dBm} \end{aligned}$$

The output electrical signal power S is proportional to the square of the optical light power P_s^{ASK}. In order to increase the electrical power S by 144 times, the light power P_s^{ASK} has to be increased by $\sqrt{144}$.

$$\begin{aligned} P_s^{\text{ASK}} &= \sqrt{144} \times 9.56 \times 10^{-9} \\ &= 115\ \text{nW} \\ &= -39.4\ \text{dBm} \end{aligned}$$

Method II. From the definition of NEP

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right)_1 = \frac{(RP_{s\,\min})^2 R_L}{N_1} = 1$$

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right)_2 = \frac{\frac{1}{2}(RP_s^{\mathrm{ASK}})^2 R_L}{N_2} = 144$$

where N_1 is the noise power at the minimum detectable power and N_2 is the noise power at 144 times the minimum detectable power.

$$P_{s\,\min} = \mathrm{NEP}\sqrt{B}$$

If the diode is thermal noise limited, then

$$N_1 = N_2 = N$$

Solving these simultaneous equations gives

$$\begin{aligned} P_s^{ASK} &= \sqrt{2}\sqrt{144}\ NEP\sqrt{B} \\ &= \sqrt{2}\sqrt{144}(3.9 \times 10^{-13})\sqrt{3 \times 10^8} \\ &= 115\mathrm{nW} = -39.4\ \mathrm{dBm} \end{aligned}$$

Method III. The electrical noise current i_N for a given NEP and responsivity R is

$$\begin{aligned} i_N &= R\ \mathrm{NEP}\sqrt{B} \\ N &= \langle i_N^2 \rangle R_L = R^2(\mathrm{NEP})^2 B R_L \end{aligned}$$

The required electrical signal power to establish (S/N) = 144 is

$$S = \left(\frac{\mathrm{S}}{\mathrm{N}}\right) N = 144\ R^2(\mathrm{NEP})^2 B R_L$$

whereas the light power P_s^{ASK} needed to generate this electrical signal is obtained by setting

$$S = \tfrac{1}{2}(RP_s^{\mathrm{ASK}})^2 R_L$$

Hence,

$$P_s^{\mathrm{ASK}} = \sqrt{2}\sqrt{144}\ \mathrm{NEP}\sqrt{B}$$ □

16.6 DESIGNING FIBER-OPTIC COMMUNICATION SYSTEMS

When designing a fiber-optic communication system, one must keep two major considerations in mind. The first consideration is power. There must be sufficient light power to maintain the desired levels of S/N or BER. The second consideration is the response time of the system, which must be fast enough to respond to the variation of the signal. Achieving adequate response time ensures that information can be transported with the required accuracy.

16.6.1 System Loss

A fiber-optic communication system consists of three components: an optical fiber, a transmitter, and a receiver. The optical fiber is terminated by the transmitter at one end and the receiver at the other. While there is a minimum light power requirement for a receiver to interrogate information, there is also a limit to the light power that the laser diode can put out. The allowed loss of light power is calculated by the difference between the light power of the transmitter and the receiver. If decibel scales are used, the formula for the power requirement is

$$P_S = P_T - P_A \tag{16.74}$$

where P_S is the power received by the PIN or APD detector in dBm, P_T is the transmitter light power from the LD or LED in dBm, and P_A is the attenuation of the light incurred in the path from the transmitter to the receiver in dB.

Let us first deal with the attenuation P_A. The transmission loss of light in the fiber is one of several mechanisms for attenuation in the system. Fiber loss ranges from 0.2 dB/km to a few dB/km depending on the quality of the fiber and the wavelength of operation.

The next item contributing to loss is the light power lost in coupling the output light from the LD or LED into the end face of the fiber. Coupling loss is several dB, but quite often the LD and LED are pigtailed to the fiber by the manufacturers. A short piece of optical fiber prealigned and glued to the LD or LED by the manufacturer is called a pigtailed LD, and the coupling loss of such an LD is already accounted for in the light output power rating.

Splicing losses are due to fiber connections that are needed during installation for joining spools of fiber or making repairs at the time of accidental breakages. A splice is intended to be a permanent connection, whereas a connector is designed so that it can easily be disconnected and reconnected. The splice loss is less than 0.1 dB per splice.

Optical fiber connectors are needed at both transmitter and receiver ends. Connector loss is less than 1 dB each. Time degradation of the output power of the LD or LED should also be accounted for by adding a few dB to the net loss.

After taking these power losses into account, generally a margin of safety is also included. A sample calculation is shown in Table 16.4 for a 50-km fiber with a pigtailed source.

The procedures for calculating the power requirement for analog and digital modulated signals are different and they will be treated separately.

16.6.2 Power Requirement for Analog Modulation

Calculating the power restriction normally starts from the receiver side. First, the light power P_s, which is required for a given frequency bandwidth B to maintain a

Table 16.4 Loss budget for a 50-km span

Item	Unit Loss	Quantity	Total
Fiber loss	0.3 dB/km	50	15.0
Connector loss	1 dB	2	2.0
Splicing loss	0.1 dB	24	2.4
Degradation	3 dB	1	3.0
Margin of safety	5 dB	1	5.0
			$P_A = 27.4$ dB

specified S/N, will be calculated. An increase in the frequency bandwidth B significantly increases noise, which raises the required light power. The influence of B on noise is twofold. First, shot as well as thermal noise contributions are proportional to B. Second, in order to increase B, R_L in the detector circuit must be decreased. Reducing R_L, however, further raises the thermal noise $(4kT/R_L)B$ contribution.

The case of direct detection is considered. If both shot noise as well as thermal noise are taken into consideration, the S/N is a quadratic function of P_s as shown by Eq. (16.42). Solving for P_s for a given S/N is somewhat involved. The following trial and error method may be simpler.

Let us first make the assumption that the S/N is thermal noise limited in order to calculate the required power P_s for a given S/N, and then reexamine the validity of this assumption from the obtained P_s.

Since the expression for the signal power for ASK modulation has already been derived, for a new challenge, let us now find the signal power S for a sinusoidally modulated analog signal where the light power is

$$P = P_s^a(1 + m\cos\omega_{st}) \tag{16.75}$$

A modulation index $m = 1$ is assumed. The signal power $S' = \langle i_s^2\rangle R_L$ from the APD is, from Eq. (16.34),

$$S = \tfrac{1}{2}\left(\frac{\eta e}{h\nu}P_s^a M\right)^2 R_L \tag{16.76}$$

This turns out to be identical to that of ASK modulation. However, note that P_s^a with the superscript a in Eq. (16.75) is the average light power, while P_s for ASK is the peak power or power for the "on" state.

In the thermal-noise-limited case, Eq. (16.49) for amplitude modulation reduces to

$$\left(\frac{\mathrm{S}}{\mathrm{N}}\right) = \frac{\tfrac{1}{2}\left(\dfrac{\eta e}{h\nu}M\right)^2 P_s^{a^2}}{\dfrac{4kTB}{R_L}} \tag{16.77}$$

The value of R_L is determined by the cutoff frequency f_c of the detector to provide the bandwidth B. From Eq. (16.15) with $f_c = B$, R_L is

$$R_L = \frac{1}{2\pi CB} \tag{16.78}$$

From Eqs. (16.77) and (16.78), the average light power P_s^a in terms of S/N and B is

$$P_s^a = 2\frac{1}{M}\frac{h\nu}{\eta e}\sqrt{4kT\pi C\left(\frac{\mathrm{S}}{\mathrm{N}}\right)B} \tag{16.79}$$

Inserting typical values $\eta e/h\nu = 0.5$ A/W, $T = 293$ K, and $C = 1.6$ pF, we get

$$P_s^a = 1.14 \times 10^{-15}\frac{1}{M}\sqrt{\frac{\mathrm{S}}{\mathrm{N}}}B \tag{16.80}$$

A few values calculated from Eq. (16.80) for S/N $= 10^5$, which is the largest value of S/N considered, are tabulated below.

	P_s^a (dBm)	
B	PIN Diode ($M = 1$)	APD ($M = 100$)
1 MHz	−34.4	−54.4
10 MHz	−24.4	−44.4
1 GHz	−4.4	−24.4

By examining where the above values fall in relation to Fig. 16.22, we can verify that all the points for the PIN diode fall into the assumed thermal-limited region, but those for the APD do not with the given value of S/N $= 10^5$. When the detector is an APD, the assumption of the thermal-noise-limited S/N does not hold for S/N $= 10^5$, and the quantum-limited S/N and Eq. (16.88), which are discussed later on, have to be used. Therefore, Eq. (16.80) can be used safely only for the PIN photodiode. The final results are plotted in Fig. 16.25.

In summary, when a PIN diode is used as the detector, find P_s^a either by Fig. 16.25 or Eq. (16.80) for a given S/N and B, and then adjust the values of P_T and P_A to be within the constraint of Eq. (16.74).

It should be mentioned that the designed frequency bandwidth of the receiver should be estimated to be about 5% wider than the required bandwidth so as to allow for the contribution of the rise times of the fiber and the transmitter $(1.05)B$ rather than B is used in Eq. (16.80) as explained in Section 16.6.4.

16.6.3 Rise-Time Requirement for Analog Modulation

The frequency response of the overall fiber-optic communication system has to account for the frequency responses of the transmitter, fiber, and receiver. There is an empirical formula for the overall frequency response of the entire system. This formula is

$$t = \sqrt{t_1^2 + t_2^2 + t_3^2 + \cdots + t_i^2 + \cdots} \tag{16.81}$$

where t is the rise time of the overall system and t_i is the rise time for an individual component. This is a very handy formula but the performance of each component has to be characterized by the rise time rather than the frequency response. The factor that converts the frequency bandwidth to the rise time will be obtained next.

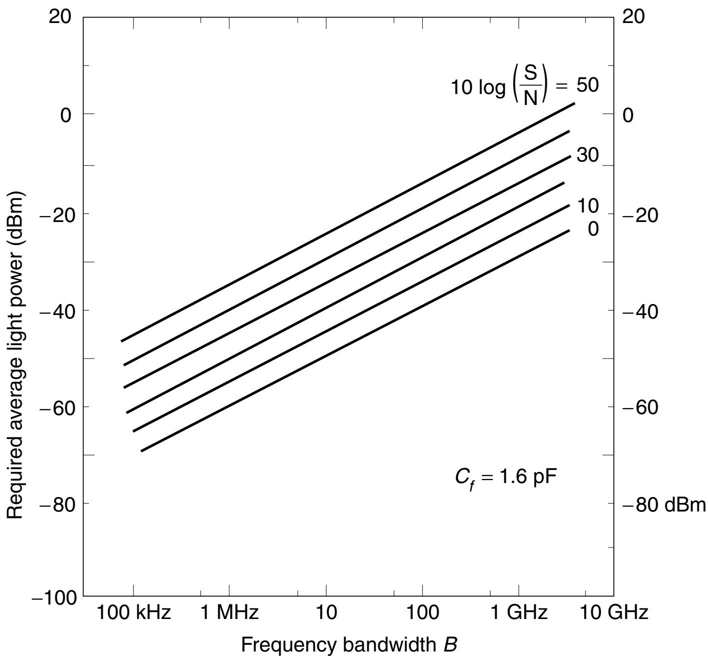


Figure 16.25 Required average light power of a PIN photodiode as a function of frequency bandwidth B with S/N as a parameter. The modulation is analog and the S/N is thermal noise limited. $C_f = 1.6$ pF.

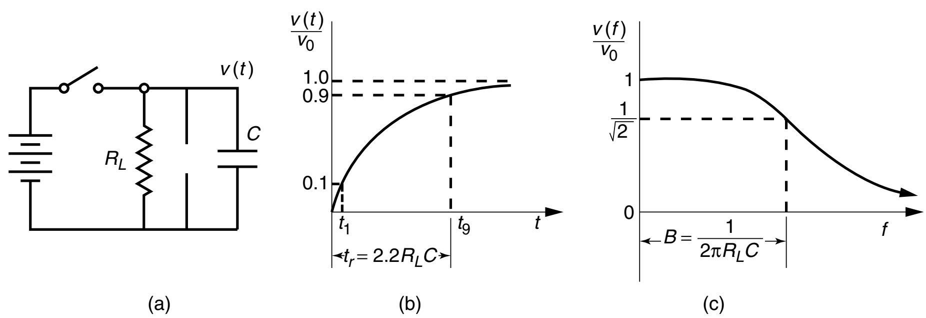


Figure 16.26 Relationship between rise time and frequency response of an RC circuit to show $Bt_r = 0.35$ by eliminating R_LC from the expressions. (a) RC circuit. (b) Rise time. (c) Frequency response.

When a step function generator is applied to an RC circuit such as shown in Fig. 16.26a, the voltage $V(t)$ starts to rise with time as follows:

$$V(t) = V_0(1 - e^{-t/RC}) \tag{16.82}$$

The curve of Eq. (16.82) is drawn in Fig. 16.26b. The rise time, which is defined as the time taken to rise from 0.1 V_0 to 0.9 V_0, is calculated from Eq. (16.82) as

$$\begin{aligned} 0.9 &= 1 - e^{-t_9/RC} \\ 0.1 &= 1 - e^{-t_1/RC} \end{aligned} \tag{16.83}$$

and $t_r = t_9 - t_1$ is

$$t_r = 2.2\ CR \tag{16.84}$$

With the same circuit, this time, a swept frequency generator is applied. The response would be something like the one shown in Fig. 16.26c. The 3-dB cutoff frequency bandwidth B is given by Eq. (16.78). Thus, the frequency bandwidth and rise time are related from Eqs. (16.78) and (16.84) as

$$Bt_r = 0.35 \tag{16.85}$$

The transmitters and detectors are all characterized by the modulation frequency bandwidth, but their rise times can be obtained from Eq. (16.85) if we assume that their frequency characteristic behaves like an RC circuit.

As for the fiber, the frequency response is characterized by the spread of the group delay. In a multimode fiber, mode dispersion is dominant and the spread of the group delay of the step-index fiber is given by Eq. (11.7), which is repeated here:

$$\Delta\tau = \frac{1}{2cn_1}(\mathrm{NA})^2 \tag{11.7}$$

and that of the graded-index fiber is given by Eq. (11.242), which is repeated here:

$$\Delta\tau = \frac{1}{8}\frac{(\mathrm{NA})^4}{cn_c^3} \tag{11.242}$$

In a single-mode fiber, the dispersion constant is given by the graph in Fig. 15.14. We can assume that the group delay spread is essentially equal to the rise time of the fiber,

$$t_r = \Delta\tau \tag{16.86}$$

where $\Delta\tau$ is the group delay.

In short, the rise time is calculated by converting the frequency response or group delay spread of the system into the rise time and summing their squares to obtain the rise time of the overall system.

16.6.4 Example of an Analog System Design

As a specific example, let us say that a 1-GHz bandwidth analog signal is to be transmitted over a 50-km distance with a signal to noise ratio greater than 1000, as shown in Fig. 16.27.

The transmitter is a double-channel planar buried heterostructure (DC-PBH) laser diode whose modulation frequency characteristics are shown in Fig. 14.12. The parameters of the other elements are as follows. The fiber transmission loss is 0.3 dB/km. The fiber is sold in spools, each wound with 2 km of fiber. The spools are spliced together to make the required system length of fiber. A PIN photodiode with responsivity $R = 0.5$ is used in the detector circuit. The wavelength of operation is $\lambda = 1.55$ μm.

First, the required average light power P_s^a to the detector is determined. The designed frequency bandwidth of the receiver is estimated to be 5% wider than required to allow

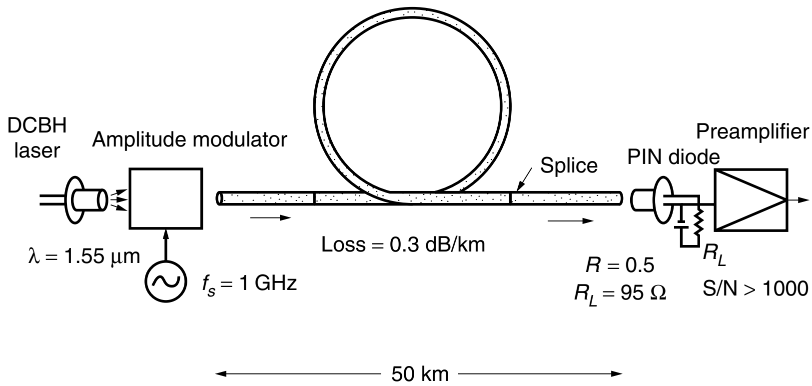


Figure 16.27 Another example of an analog system design.

later for the fiber and transmitter rise times. If 5% is insufficient, this margin can be increased and the design procedure would then be reiterated. The S/N is first assumed to be thermal noise limited. From Fig. 16.25 (or directly from Eq. (16.80)), the required P_s^a for S/N = 1000 and $B = 1.05$ GHz is $P_s^a = -14$ dBm or in terms of the peak power $P_s = -11$ dBm.

The load impedance for 1.05 GHz is, from Eq. (16.15),

$$R_L = \frac{1}{2\pi BC} = \frac{1}{2\pi(1.05 \times 10^9)(1.6 \times 10^{-12})} = 95\ \Omega$$

Next, the total attenuation loss P_A must be found. The relevant numbers are found in Table 16.4 and the result is

$$P_A = 27.4 \text{ dB}$$

Thus, the required transmitter power is, from Eq. (16.74),

$$P_T = P_S + P_A = -11 + 27.4 = 16.4 \text{ dBm}$$

An LD whose peak output is 16.4 dBm or 43.7 mW is needed. Such a power level is somewhat demanding. Later, an alternative detector circuit will be considered to see if the demand on transmitter power can be reduced.

Next, the rise time is calculated. The rise time t_S of the system is

$$t_S^2 = t_D^2 + t_F^2 + t_T^2 \tag{16.87}$$

where t_S is the rise time for the whole system, t_D is the rise time for the detector, t_F is the rise time for the fiber, and t_T is the rise time for the transmitter. The rise time t_S of Eq. (16.85) of the desired 1-GHz system is

$$t_S = \frac{0.35}{B} = \frac{0.35}{10^9} = 350 \text{ ps}$$

The rise time t_D of the detector circuit whose rise time is 5% shorter than desired to accommodate for t_F and t_T is

$$t_D = \frac{0.35}{B} = \frac{0.35}{1.05 \times 10^9} = 333 \text{ ps}$$

Next, the rise time t_F of a single-mode fiber will be calculated. The spread of the group delay $\Delta\tau$ is calculated from the wavelength bandwidth of the signal. The wavelength bandwidth $\Delta\lambda$ due to the 1-GHz modulation is

$$\Delta\lambda = 2\left(\frac{c}{f_c} - \frac{c}{f_c + 10^9}\right) = 0.016 \text{ nm}$$

where f_c is the unmodulated 1.55-μm wavelength light. From Fig. 15.14 the dispersion constant D at $\lambda = 1.55$ μm is

$$D = 17 \text{ ps/(km} \cdot \text{nm)}$$

and

$$\Delta\tau = (17)(50)(0.016) = 13.6 \text{ ps}$$

Therefore, from Eq. (16.86), $t_F = 13.6$ ps.

The modulation frequency bandwidth of the laser is shown in Fig. 14.12 as 8 GHz for $i = 30$ mA, and hence,

$$t_T = \frac{0.35}{8 \times 10^9} = 43.8 \text{ ps}$$

The overall rise time therefore is

$$\begin{aligned} \sqrt{t_D^2 + t_F^2 + t_T^2} &= \sqrt{333^2 + 13.6^2 + 43.8^2} \\ &= 336 \text{ ps} \end{aligned}$$

which is shorter than the required 350-ps system rise time and satisfies the budgeted rise-time requirement.

Lastly, let us see if the overstressed requirement of the transmitter light power can be relieved by installing the transimpedance circuit shown in Fig. 16.16 in the detector circuit. Earlier, Eq. (16.78) was used to obtain Eq. (16.80) relating B and R_L, but in this case, R in Fig. 16.16 has to be used and the results in Fig. 16.25 can no longer be applied. The frequency band B of the transimpedance circuit is given by

$$B = \frac{1}{2\pi(R_f/G)C} \tag{16.24}$$

and the two conditions of Eq. (16.22) for this to be true are $G \gg 1$ and $RG \gg R_f$. The impedance R of the PIN diode, being of the order of megohms, easily satisfies the second condition. In addition, the thermal noise contribution $4kTB/R$ becomes negligibly small compared to the shot noise, and the quantum-limited approximation has to be used. The quantum-limited S/N for analog modulation using a PIN diode is

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\frac{1}{2}\left(\frac{\eta e}{h\nu}P_s^a\right)^2}{2eB\frac{\eta e}{h\nu}P_s^a} \tag{16.88}$$

and

$$P_s^a = 4e\frac{h\nu}{\eta e}\left(\frac{\mathrm{S}}{\mathrm{N}}\right)B \tag{16.89}$$

Inserting parameters and typical values of $\eta e/h\nu = 0.5$, Eq. (16.89) becomes

$$P_s^a = 1.28 \times 10^{-18}\left(\frac{\mathrm{S}}{\mathrm{N}}\right)B \tag{16.90}$$

Using Eq. (16.90), $P_s^a = -28.9$ dBm or $P_s = -25.9$ dBm is obtained and P_T becomes

$$P_T = -25.9 + 27.4 = 1.5 \text{ dBm}$$

or 1.4-mW peak power. A laser diode with such an output power is readily available.

Example 16.4 A 6-MHz wide analog TV signal is to be sent over a 10-km optical fiber link between a TV station in the city and a TV tower on a hill for microwave broadcasting as shown in Fig. 16.28. Spools wound with 2 km of multimode graded-index fiber with NA = 0.2 and $n_c = 1.55$ are intended to be used. The fiber loss is 1 dB/km. The wavelength is 0.82 μm. A PIN diode with a terminal capacitance of 7 pF and zero dark current is used as the detector and a laser diode with 10-ns rise time is used as the transmitter. The laser diode is not pigtailed. The required signal to noise ratio is 40 dB.

(a) What is the required transmitter peak power of the laser diode?
(b) Is it possible to use a step-index multimode fiber?

Solution A 5% extra bandwidth is allotted for the rise-time budget and $B = 6.3$ MHz. If the system is assumed thermal noise limited, the average light power P_s^a needed for the detector is -32 dBm $= 0.63$ μW from Fig. 16.25, which verifies the thermal-noise-limited S/N from Fig. 16.22. However, according to the specification, this PIN diode has a terminal capacitance of 7 pF while a capacitance of 1.6 pF was assumed when Figs. 16.22 and 16.25 were prepared. Note from Eq. (16.79) that P_s^a is proportional to $\sqrt{C}$, and hence the required power P_s^a is

$$P_s^a = (0.63)\sqrt{\frac{7}{1.6}} = 1.3 \text{ μW}$$

Point $P_s^a = 1.3$ μW and $R_L = (2\pi BC)^{-1} = 3609\ \Omega$ is again located in the thermal-noise-limited region in Fig. 16.22. Alternatively, from Eq. (16.35), the shot noise power is $N_{\text{shot}} = 4.73 \times 10^{-15}$ W. From Eq. (16.36), the thermal noise power is

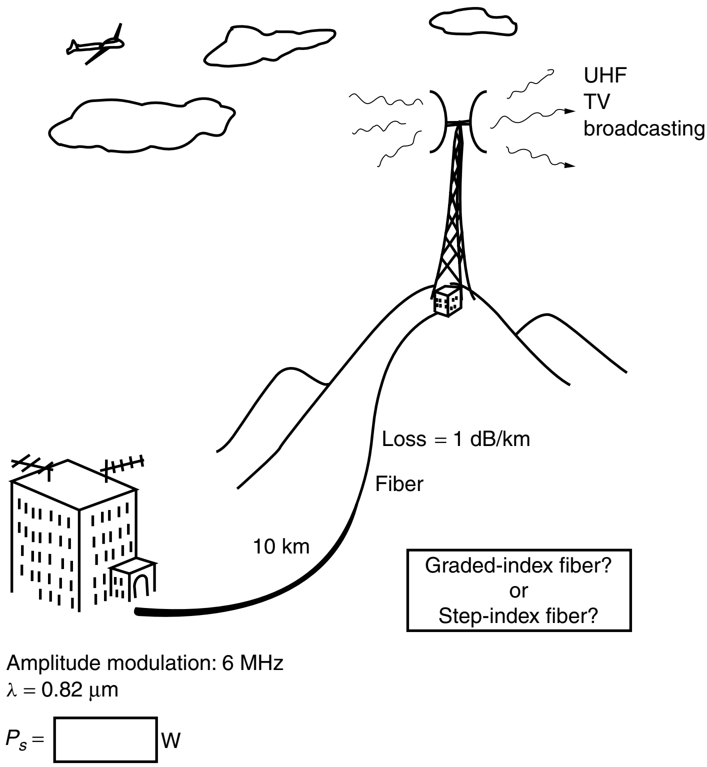


Figure 16.28 Another example of an analog system design.

$N_{\text{th}} = 1 \times 10^{-13}$ W. Thus, the assumption of the thermal-noise-limited system is verified for $C = 7$ pF. The peak power P_s is $2P_s^a$ and

$$P_s = 2.6\ \mu\text{W} = -25.9\ \text{dBm}$$

The power attenuation list is:

Coupling loss from the source to fiber	5 dB
Fiber loss (1) (10)	10 dB
Connector loss	2 dB
Splice loss (0.1) (5)	0.5 dB
Degradation	3 dB
Margin of safety	5 dB
	25.5 dB

Hence, the required transmitter power P_T is

$$P_T = P_S + P_A = -25.9 + 25.5 = -0.4\ \text{dBm}$$
$$= 0.91\ \text{mW}$$

(b) The system rise time t_S is

$$t_S = \frac{0.35}{B} = \frac{0.35}{6 \times 10^6} = 58 \text{ ns}$$

The rise time for the detector is

$$t_D = \frac{0.35}{(1.05)B} = \frac{0.35}{6.3 \times 10^6} = 56 \text{ ns}$$

The spread of the arrival time of the pulse in the graded-index fiber is calculated from Eq. (11.242):

$$\Delta\tau = \frac{(\text{NA})^4}{8cn_c^3}L = \frac{(0.2)^4}{8(3 \times 10^5)(1.55^3)}(10) = 1.8 \text{ ns}$$

Thus, the total budget is

$$\sqrt{56^2 + 1.8^2 + 10^2} = 56.9 \text{ ns}$$

Next, the possibility of using a step index multimode fiber is considered. The group delay spread is given by Eq. (11.7) as

$$\Delta\tau = \frac{(NA)^2}{2cn_1}L = \frac{(0.2)^2}{2(3 \times 10^5)1.55}(10) = 430 \text{ ns}$$

The fiber alone exceeds the system rise time of 58 ns and thus this fiber cannot be used. In other words, from Eq. (16.85), the bandwidth of the fiber is

$$B = \frac{0.35}{t_F} = \frac{0.35}{4.3 \times 10^{-7}} = 814 \text{ kHz}$$

which is too narrow for the 6-MHz system. □

16.6.5 Required Frequency Bandwidth for Amplifying Digital Signals

The design procedure for the digital system is more or less the same as that of the analog system. The differences are the units used. In digital systems, BER is used instead of S/N, and bit rate B_t is used instead of bandwidth B. Since the conversion between BER and S/N has already been done (see Fig. 16.24), the conversion between B and B_t will be explained before going on to the next design procedure.

The baseband electrical pulse, such as the one shown in Fig. 16.29a, is represented by the rectangle function from Section 1.4.1 as

$$f(t) = \Pi\left(\frac{t}{T}\right) \tag{16.91}$$

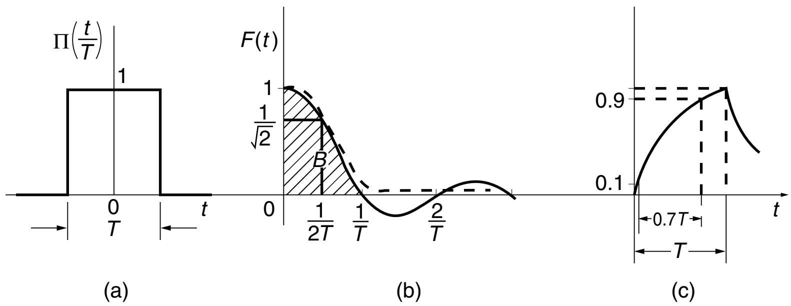


Figure 16.29 Relationship between NRZ pulse, its Fourier transform, and the output of an amplifier with $B = 1/2T$. (a) NRZ pulse in the time domain. (b) NRZ pulse in the frequency domain. (c) Output from an amplifier with $B = 1/2T$.

where T is the period of the NRZ pulse. The frequency spectrum $F(f)$ of this pulse is obtained from its Fourier transform as

$$F(f) = T \text{ sinc } Tf \tag{16.92}$$

As shown in Fig. 16.29b, the first zero crossing point appears at $f = 1/T$ and the 3-dB down (one-half of the power and $1/\sqrt{2}$ of the voltage) point appears at approximately the midpoint between the origin and the first zero crossing. Suppose this pulse is applied to an amplifier whose 3-dB bandwidth B is set at $1/2T$. The frequency characteristic of the amplifier is shown by the dotted lines in Fig. 16.29b. As seen from the figure, the spectral components beyond the first zero crossing cannot be amplified by this amplifier, even though the major frequency components are retained in the passband.

If it is approximated that this amount of retained spectrum will do, then the required frequency bandwidth of the amplifier is

$$B = \frac{1}{2T} \tag{16.93}$$

The pulse of the NRZ code in Table 16.2 takes up one entire period of coding and the bit rate B_t and period are related by

$$B_t = \frac{1}{T} \tag{16.94}$$

Thus, the required frequency bandwidth to amplify the NRZ pulses at the bit rate B_t is, from Eqs. (16.93) and (16.94),

$$B = \tfrac{1}{2}B_t \tag{16.95}$$

The conclusion is that the required 3-dB frequency bandwidth of the amplifier is one-half of the bit rate of the NRZ pulses.

Let us see how the output of the amplifier looks if the bandwidth of the amplifier is set according to Eq. (16.93). The rise time t_r when a step function is applied to an amplifier with a 3-dB cutoff at B is expressed by Eq. (16.85). Inserting Eq. (16.93) into (16.85) gives

$$t_r = 0.7\ T \tag{16.96}$$

The rise time of the output is 70% of the period as shown in Fig. 16.29c. For most designs, an amplified signal such as this is satisfactory. The rise time required for the RZ code is $0.35T$. If an amplifier is used with $B = N/2T$, which is N times wider than that specified by Eq. (16.93), then the rise time of the output becomes N times faster and the output more closely resembles the input. However, for the purpose of digital communication, a faithful reconstruction of the pulse is a luxury because all that is really needed in binary coding is whether or not the pulse is present, and the bandwidth of the amplifier specified by Eq. (16.95) is sufficient for the NRZ pulses.

In the case of the RZ coding, the pulsewidth becomes one-half of the pulse of NRZ coding, and the rise time has to be twice as fast to retain the same quality of the output pulse. For RZ coding, B has to be twice as wide as set by Eq. (16.93) and

$$B = \frac{1}{T} \tag{16.97}$$

Comparing Eq. (16.97) with (16.94),

$$B = B_t \tag{16.98}$$

Thus, for the RZ coding, the frequency bandwidth of the amplifier must be identical with that of the bit rate of the digital signal. This concludes the explanation of the conversion between B and B_t. Now let us return to the design procedure for digital modulation.

16.6.6 Digital System Design

When explaining the procedure of calculating power for digital systems, some of the calculated results for analog modulation can be modified and utilized. The required BER can be converted into the required S/N using the graph in Fig. 16.24. The bit rate B_t can be converted into the required frequency bandwidth B of the detector circuit using Eq. (16.95) for the NRZ code and Eq. (16.98) for the RZ code. Hence, the methods of calculating power are similar for both analog and digital modulations.

The case of the NRZ code will provide an example. The peak light power P_s, which is the light power for the "1" bit, is used in the expressions.

First, the peak light power P_s needed for a given BER will be calculated assuming a thermal-noise-limited S/N. If half the bits are "0"s and the other half are "1"s, then the signal to noise ratio is

$$\frac{\mathrm{S}}{\mathrm{N}} = \frac{\frac{1}{2}\left(\dfrac{\eta e}{h\nu} M P_s\right)^2}{\dfrac{4kTB}{R_L}} \tag{16.99}$$

The frequency bandwidth requirement of the digital case is only one-half of the bit rate as given by Eq. (16.95). The value of the load impedance, therefore, is double that of the analog case; explicitly,

$$R_L = \frac{1}{\pi C B_t} \tag{16.100}$$

Inserting Eqs. (16.95) and (16.100) into Eq. (16.99) gives

$$P_s = \frac{1}{M}\frac{h\nu}{\eta e}\sqrt{4kT\pi C\left(\frac{\mathrm{S}}{\mathrm{N}}\right)}B_t \tag{16.101}$$

Inserting $\eta e/h\nu = 0.5$ A/W, $T = 293$ K, and $C = 1.6$ pF,

$$P_s = 5.70 \times 10^{-16}\frac{1}{M}\sqrt{\frac{\mathrm{S}}{\mathrm{N}}}B_t \tag{16.102}$$

The S/N for a given BER is found from the conversion curve in Fig. 16.24. The required peak powers P_s with $M = 1$ and $B_t = 1$ Gb/s for the three cases BER $= 10^{-2}$, 10^{-4}, and 10^{-9} are tabulated for comparison.

PEAK POWER P_s REQUIRED FOR VARIOUS BER

BER	S/N	$\sqrt{\mathrm{S/N}}$	P_s (dBm)
10^{-2}	21.9	4.68	−25.7
10^{-4}	54.8	7.4	−23.8
10^{-9}	144	12.0	−21.7

As seen from the table, a change in BER does not profoundly influence the value of P_s. A rule of thumb is that the value of P_s for 10 times the BER can be found by subtracting 0.5 dBm from the original value of P_s. For instance, the value of P_s for BER $= 10^{-8}$ is obtained from that for BER $= 10^{-9}$ as $-21.7 - 0.5 = -22.2$ dBm.

$P_s^a(= \frac{1}{2}P_s)$ is plotted against B_t for a PIN photodiode in Fig. 16.30 for BER $= 10^{-9}$ and $C = 1.6$ pF. As in the case of analog modulation, this curve is reexamined to see whether or not the thermal-noise-limited assumption was correct. Referring to Fig. 16.22, the curves of not only the PIN photodiode but also the APD in Fig. 16.30 do indeed satisfy the thermal-noise-limited assumption. This is simply because the digital system does not need a large value of S/N so that P_s is smaller and the lower equation in Eq. (16.55) holds. Unless such elaborate circuits as mentioned in Section 16.4.3 are used, the PIN and APD detectors are usually thermal noise limited for digital signals.

The peak value P_s has been used for the NRZ code instead of the average P_s^a. The average value, however, is much easier to measure. The vender's specification of the output power from the laser diode or LED is according to the peak power. The designer has to convert between the two values as necessary. Equation (16.102) can be expressed in terms of P_s^a using

$$P_s^a = \tfrac{1}{2}P_s \tag{16.103}$$

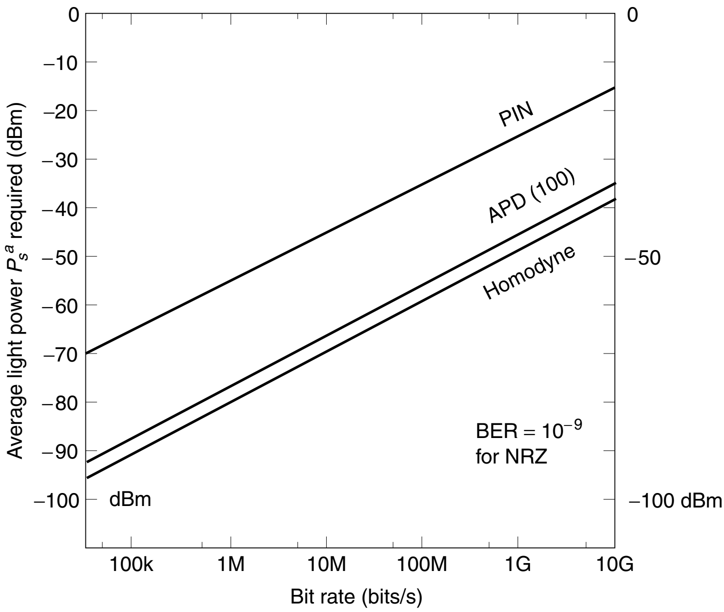


Figure 16.30 Input average light power P_s^a versus bit rate (bits/s) for NRZ code. For the value of the RZ code add 3 dB. The PIN photodiode and APD are thermal noise limited but the homodyne is quantum limited.

For the thermal-noise-limited case for the PIN photodiode, Eq. (16.102) becomes

$$P_s^a = 2.85 \times 10^{-16} \frac{1}{M} \sqrt{\frac{\mathrm{S}}{\mathrm{N}}} B_t \tag{16.104}$$

Lastly, homodyne detection is considered. With

$$2P_s^a = P_s, \quad B_{\mathrm{IF}} = 2B, \quad B = \frac{B_t}{2}, \quad M^x = 1, \text{ and } \frac{\eta e}{h\nu} = 1 \tag{16.105}$$

Eq. (16.53) for homodyne detection becomes

$$\begin{aligned} P_s^a &= \tfrac{1}{2} e \left(\frac{\mathrm{S}}{\mathrm{N}}\right) B_t \\ &= 0.8 \times 10^{-19} \left(\frac{\mathrm{S}}{\mathrm{N}}\right) B_t \end{aligned} \tag{16.106}$$

These curves for $BER = 10^{-9}$ or S/N = 144 are plotted in Fig. 16.30.

If a peak light power P_s is needed in the process of selecting a laser diode for the NRZ code, then 3 dB is added to the value obtained by the graph in Fig. 16.30. For the RZ code, $P_s^a = \frac{1}{4} P_s$ and 6 dB is added to the value obtained by the graph. This concludes the power calculations.

The rise-time calculations are quite similar to the case of analog modulation and will be explained in the following example.

16.6.7 Example of Digital System Design

Let us construct a 500-Mb/s fiber-optic link with RZ code and BER $= 10^{-9}$. The components used are a 3-mW peak power nonpigtailed laser diode with a rise time of 200 ps as the transmitter and an APD with $\eta e/h\nu = 0.5$ and $x = 0.3$, and $M = 100$ as the receiver. The wavelength used is $\lambda = 1.55$ µm.

Find the maximum distance of the transmission if a single-mode fiber with 0.3-dB/km loss is used. The fiber is available in lengths of 2 km. The problem is outlined in Fig. 16.31.

16.6.7.1 Power Requirement

The required average light power is found from Fig. 16.30 and is equal to $P_s^a =$ -48.7 dBm. The point for $P_a^s = -48.7$ dBm and $B = B_t = 500$ Mb/s is indeed in the thermal-noise-limited region in Fig. 16.22. The required peak power for the RZ code is found by adding 6 dB:

$$-48.7 + 6 = -42.7 \text{ dBm}$$

The transmitter's 3-mW peak power in dBm is 4.8 dBm.

The affordable attenuation P_A between the transmitter and receiver is

$$P_A = P_T - P_S = 4.8 - (-42.7) = 47.5 \text{ dB}$$

The essential losses for the fiber link are as follows:

Coupling loss from the source to fiber	5 dB
Connector loss	2 dB
Degradation	3 dB
Margin of safety	5 dB
	15 dB

After considering the attenuation losses listed above, the rest of the loss allowed is

$$47.5 - 15 = 32.5 \text{ dB}$$

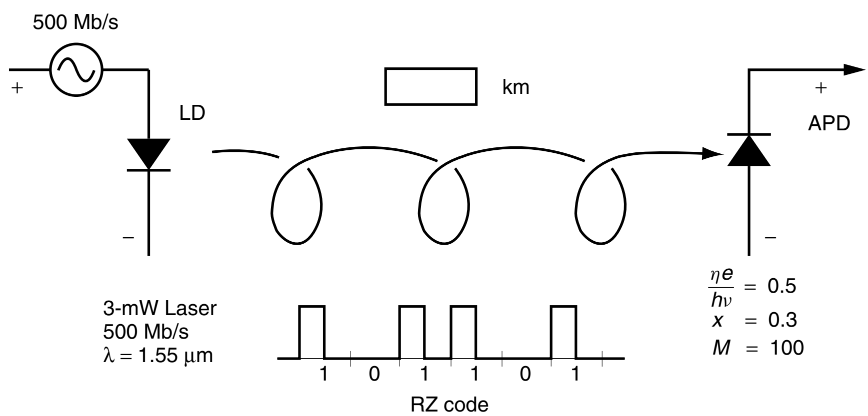


Figure 16.31 Layout of a digital system design.

The fiber has to be spliced every 2 km with 0.1-dB loss, which is equivalent to 0.05-dB/km splicing loss. The total length x of the fiber is

$$32.5 = (0.3 + 0.05)x$$

$$x = 92.9 \text{ km}$$

16.6.7.2 Rise-Time Requirement

An extra 5% is added to the bit rate to allow for the fiber and transmitter. The required rise time for the RZ code must be half as short as that of the NRZ code given by Eq. (16.96):

$$t_S = 0.35\ T = \frac{0.35}{5 \times 10^8} = 700 \text{ ps}$$

The rise time of the detector is

$$t_D = \frac{0.35}{5.25 \times 10^8} = 667 \text{ ps}$$

Next, the spread of the delay time of the single-mode fiber is calculated. For the RZ code, the necessary frequency bandwidth B from Eq. (16.98) is $B = 500$ MHz. The wavelength bandwidth $\Delta\lambda$ due to the modulation is

$$\Delta\lambda = 0.008 \text{ nm}$$

The dispersion constant D from Fig. 15.14 at $\lambda = 1.55$ μm is $D = 17$ ps/(km·nm) and

$$t_F = \Delta\tau = (17)(92.9)(0.008) = 12.6 \text{ ps}$$

Thus,

$$\sqrt{t_D^2 + t_F^2 + t_T^2} = \sqrt{667^2 + 12.6^2 + 200^2} = 696 \text{ ps}$$

which shows that the rise-time requirement is also satisfied.

Example 16.5 A fiber-optic communication system that can transmit the U.S. standard rating DS-4 PCM with BER $= 10^{-10}$ will be installed on a 100-km trunk line. Homodyne detection with the NRZ code is used. The wavelength is $\lambda = 1.55$ μm. A single-mode fiber with transmission loss of 0.4 dB/km in 2-km spools is used.

(a) What is the required transmitter peak power of the pigtailed laser diode?
(b) What is the rise-time requirement?

Solution The question is summarized in Fig. 16.32.

(a) From Table 16.3, DS-4 has a bit rate of $B_t = 274.175$ Mb/s. With 5% margin, $B_t = 288$ Mb/s. However, the curve in Fig. 16.30 was calculated for BER $= 10^{-9}$. From the rule of thumb of Section 16.6.6, 0.5 dBm should be added to this value to

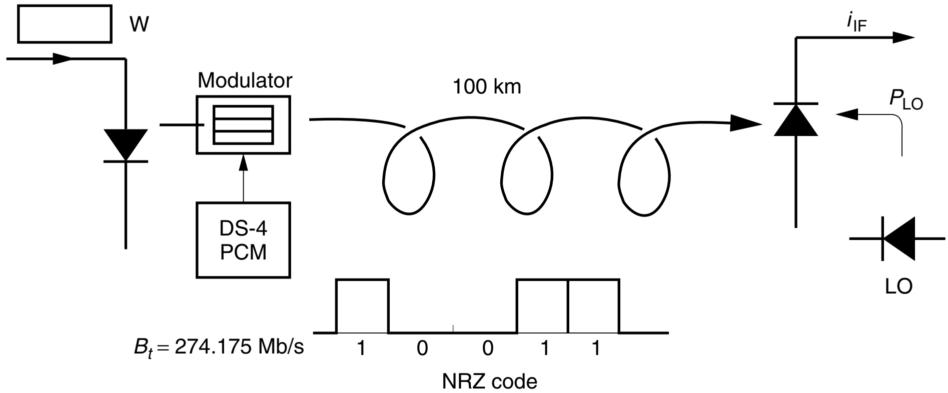


Figure 16.32 Layout of Example 16.5.

obtain $P_s^a = -55.6$ dBm, which corresponds to $-55.6 + 3 = -52.6$-dBm peak power. The tabulated attenuations of the components are as follows:

Fiber loss (0.4)(100)	40 dB
Connector loss	2 dB
Splice loss (0.1)(49)	5 dB
Degradation	3 dB
Margin of safety	5 dB
	55 dB

Finally, the required transmitter peak power is

$$P_T = P_R + P_A = -52.6 + 55 = 2.4 \text{ dBm}$$
$$= 1.7 \text{ mW}$$

(b) The system rise time is

$$t_S = \frac{0.35}{B} = \frac{0.35}{274.175 \times 10^6} = 1.28 \text{ ns}$$

The rise time for the detector is

$$t_D = \frac{0.35}{(1.05)B} == 1.22 \text{ ns}$$

In order to calculate the dispersion of the fiber, the wavelength spread $\Delta\lambda$ due to the modulation is calculated as

$$\Delta\lambda = 0.0044 \text{ nm}$$

Thus, the spread of the group delay is, from Fig. 15.14,

$$t_F = \Delta\tau = (17)(100)(0.0044) = 7.48 \text{ ps}$$

Thus, the resultant rise time is

$$\sqrt{(1.22)^2 + (7.48 \times 10^{-3})^2 + 0.18^2} = 1.23 \text{ ns}$$

and is shorter than the desired system rise time. □

PROBLEMS

16.1 The parameters of a receiver system are $\eta e/h\nu = 0.5$, $M = 1$, $C = 1.6$ pF, $I_d = 0$, and $T = 293$ K. The average input is $P_s^a = -23$ dBm and $B = 2$ kHz. No special frequency band compensation circuit is installed in the preamplifier. Is the S/N of the receiver system thermal noise limited or quantum limited?

16.2 A fiber-optic communication link will be made using a multimode step-index fiber with a core refractive index of $n_1 = 1.55$ and a numerical aperture NA $= 0.2$. The transmitter laser diode has a rise time of $t_T = 20$ ns and the detector has a rise time of $t_D = 50$ ns. The signal to be transmitted is a 4-MHz analog signal. Considering only the rise-time requirement, what is the maximum distance of transmission?

16.3 What is the required peak power to a PIN photodiode in the following system. A transimpedance circuit is used in the preamplifier and the detector can be considered quantum limited. The system operates at 100 Mb/s in RZ mode with 10^{-10} BER rate. The responsivity of the PIN diode is $R = 0.5$.

16.4 A fiber-optic communication link is to be established between two locations separated by 20 km with the following specifications.

(1) The frequency bandwidth of the signal is 300 MHz. Assume also that a 300-MHz low-pass filter has been installed.
(2) ASK modulation with NRZ code is employed.
(3) The receiver is a silicon APD, type 2382 of Hamamatsu Photonics.
(4) The transmission loss of the optical fiber is 2 dB/km.
(5) The sum of all other losses including splicing and connector loss is 2 dB. Both long-term degradation and safety margin are disregarded.
(6) The light output power from the laser diode without modulation is 20 mW, and the light wavelength is $\lambda = 830$ nm.
(7) All the characteristics are at 25°C.

(a) Find the peak current into the load.
(b) Determine the received electrical signal power.
(c) Find the shot noise power.
(d) Determine the thermal noise power.

(e) Find the signal to noise ratio.

(f) Is the system quantum limited or thermal noise limited?

(g) If $M = 1$ is chosen, is the answer to (f) still the same?

(h) What is the NEP of this APD?

(i) What is the minimum detectable light power of the APD?

(j) In this example, the span distance was 20 km. What is the theoretical limit of the span distance if the same light source, fiber, detector, and ASK modulation scheme are used? Disregard dispersion considerations.

(k) If only the power requirement is considered, what is the span distance if the same system is used for 300-Mb/s NRZ modulation with BER $= 10^{-9}$?

16.5 Answer the same questions as in Problem 16.4 but use the type S2384 APD from Fig. 16.21 with $M = 60$ and with the signal bandwidth of 100 MHz. This time, the APD is connected directly to a 50-Ω input impedance preamplifier, as in Fig. 16.14.

16.6 Find the optimum value of the multiplication factor M of an APD analytically from the viewpoint of the signal to noise ratio.

16.7 What happens to the cutoff frequency in a frequency region where the stray current $j\omega C_s V_f$ in Fig. 16.16 can no longer be neglected, where C_s is the stray capacitance in the feedback circuit?

16.8 A digital signal whose probability density functions are shown in Fig. P16.8 is applied to a discriminator. The signal level of "1" is s_1 volts and that of "0" is s_0 volts. The variance for "1" is σ_1 and that for "0" is σ_0. Find an expression that enables one to calculate the threshold voltage v_{th} that minimizes the bit error rate. Verify also that the optimum value of the threshold voltage for $\sigma_0 = \sigma_1 = \sigma$ is $v_{\text{th}} = \frac{1}{2}(s_1 + s_0)$.

16.9 Figure 16.8 shows an example of a multiplexing system designed to deliver 10 TV channels to each of 15 households. Each TV channel occupies 4-MHz bandwidth, and the TV channels are spaced by 8 GHz. The distance between the TV station and a household is 10 km on average, and a S/N of 30 dB is required.

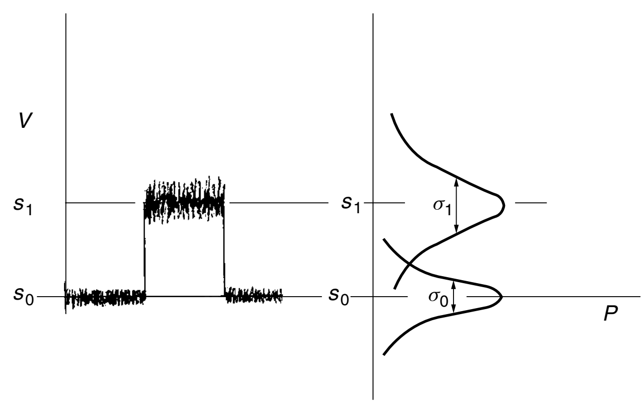


Figure P16.8 A pulse with different probability density function for "0" and "1" states.

The fiber is single-mode fiber with a loss of 1 dB/km at $\lambda = 1.55$ μm and 3 dB/km at $\lambda = 1.3$ μm and the fiber is spooled in lengths of 2 km. Laser diodes operating at $\lambda = 1.55$ μm are being proposed as the transmitters. Heterodyne detection systems with a rise time of 10^{-12} s and $\eta e/h\nu = 0.5$ are used as the receiver. The scrambler has 0.5-dB loss to each TV channel. For the rise-time budget, t_T is ignored. AM modulation is used.

(a) What is the required peak light output power?

(b) Is the rise-time requirement satisfied for the proposed transmitter?

(c) If a source wavelength of $\lambda = 1.3$ μm is used, is the rise-time requirement satisfied?

(d) What is the modification to the power requirement for $\lambda = 1.3$-μm transmitters?

REFERENCES

1. K. Iizuka (Producer), "Fiber optics labs from around the world," paneled by C. Kao, P. Cochrane, H. Kogelnik, T. Uchida, Y. Suematsu, and T. Nakahara, IEEE video, Product Number VC0022-4-POX, 1992.
2. S. Ryu, *Coherent Lightwave Communication Systems*, Artech House, Boston, 1995.
3. D. J. H. Maclean, *Optical Line Systems*, Wiley, New York, 1996.
4. J. M. Senior, *Optical Fiber Communications, Principles and Practice*, 2nd ed., Prentice Hall, New York, 1992.
5. J. Gowar, *Optical Communication Systems*, 2nd ed., Prentice Hall, New York, 1993.
6. J. C. Palais, *Fiber Optic Communications*, 4th ed., Prentice Hall, New York, 1998.
7. G. P. Agrawal, *Fiber-Optic Communication Systems*, 2nd ed., Wiley, New York, 1997.
8. L. G. Kazovsky, S. Benedetto, and A. E. Willner, *Optical Fiber Communication Systems*, Artech House, Boston, 1996.
9. J. P. Powers, *An Introduction to Fiber Optic Systems*, Aksen Associates, Irwin, Boston, 1993.
10. Special issue on "Multiple Wavelength Technologies and Networks," *J. Lightwave Technol.*, **14**(6), 932–1577 (1996).
11. A. E. Willner, "Mining the optical bandwidth for a terabit per second," *IEEE Spectrum*, 32–41 (Apr. 1997).
12. H. A. Willebrand and B. S. Ghuman, "Fiber optics without fiber," *IEEE Spectrum*, 41–45 (August 2001).
13. J. Pearson, *Basic Communication Theory*, Prentice Hall, New York, 1992.
14. R. E. Ziemer and W. H. Tranter, *Principles of Communications: Systems, Modulation, and Noise*, 3rd ed., Houghton Mifflin Co., Boston, 1990.
15. L. W. Couch II, *Digital and Analog Communication Systems*, 5th ed., Prentice Hall, New York, 1997.
16. T. Matsuda, A. Naka, and S. Saito, "Comparison between NRZ and RZ signal formats for in-line amplifier transmission in the zero-dispersion regime," *J. Lightwave Technol.*, **16**(3), 340–348 (1998).
17. S. Yamazaki, M. Shibutani, N. Shimosaka, S. Murata, T. Ono, M. Kitamura, K. Emura, and M. Shikada, "A coherent optical FDM CATV distribution system," *J. Lightwave Technol.*, **8**(3), 396–405, 1990.

18. T. Miya, "Silica-based lightwave circuits: passive and thermally active devices," *IEEE J. Selected Topics Quantum Electron.*, **6**(1), 38–45 (2000).
19. Catalogue of the Hamamatsu Corporation, Japan.
20. M. S. Park and R. A. Minasian, "Ultra-low-noise and wideband-tuned optical receiver synthesis and design," *J. Lightwave Technol.*, **12**(2), 254–259 (1994).
21. E. M. Kimber, B. L. Patel, I. Hardcastle, and A. Hadjifotiou, "High performance 10 Gb/s PIN-FET optical receiver," *Electron. Lett.* **28**(2), 120–122 (1992).
22. M. Miyashita, K. Maemura, K. Yamamoto, T. Shimura, M. Nogami, K. Motoshima, T. Kitayama, and Y. Mitsui, "An ultra broadband GaAs MESFET preamplifier IC for a 10 Gb/s optical communication system," *IEEE Trans. Microwave Tech.*, **40**(12), 2439–2444 (1992).
23. Y. Akahori, M. Ikeda, A. Kohzen, and Y. Akatsu, "11 GHz ultrawide-bandwidth monolithic photoreceiver using InGaAs PIN PD and InAlAs/InGaAs HEMTs," *Electron. Lett.*, **30**(3), 267–268 (1994).
24. A. P. Freundorfer and T. L. Nguyen, "Noise in distributed MESFET preamplifiers," *IEEE J. Solid-State Circuits*, **31**(8), 1100–1111 (1996).
25. A. P. Freundorfer and P. Lionais, "A low-noise broad-band GaAs MESFET monolithic distributed preamplifier," *IEEE Photonics Technol. Lett.*, **7**(4), 424–426 (1995).
26. K. Takahata, Y. Muramoto, H. Fukano, K. Kato, A. Kozen, S. Kimura, Y. Imai, Y. Miyamoto, O. Nakajima, and Y. Matsuoka, "Ultrafast monolithic receiver OEIC composed of multimode waveguide p-i-n photodiode and HEMT distributed amplifier," *IEEE J. Selected Topics Quantum Electron.*, **6**(1), 31–37 (2000).
27. N. Takachio, K. Iwashita, S. Hata, K. Onodera, K. Katsura, and H. Kikuchi, "A 10 Gb/s optical heterodyne detection experiment using a 23 GHz bandwidth balanced receiver," *IEEE Trans. MTT* **38**(12), 1900–1905 (1990).
28. J. Minkoff, *Signals, Noise, and Active Sensors: Radar, Sonar, Laser Radar*, Wiley, New York, 1992.
29. J. R. Barry and E. A. Lee, "Performance of coherent optical receivers," *Proc. IEEE* **78**(8), 1369–1394 (1990).
30. P. O. Börjesson and C. W. Sunberg, "Simple approximations of the error function $Q(x)$ for communications applications," *IEEE Trans. Commun.*, **27**(3), 639–643 (1979).