18

Physical Media and Channels

Ultimately the design of a digital communication system depends on the properties of the channel. The channel is typically a part of the digital communication system that we cannot change. Some channels are simply a physical medium, such as a wire pair or optical fiber. On the other hand, the radio channel is part of the electromagnetic spectrum, which is divided by government regulatory bodies into bandlimited radio channels that occupy disjoint frequency bands. In this book we do not consider the design of the *transducers*, such as antennas, lasers, and photodetectors, and hence we consider them part of the channel. Some channels, notably the telephone channel, are actually composites of multiple transmission subsystems. Such *composite channels* derive their characteristics from the properties of the underlying subsystems.

Section 18.1 discusses composite channels. Sections 18.2 through 18.4 review the characteristics of the most common channels used for digital communication, including the transmission line (wire pair or coaxial cable), optical fiber, and microwave radio (satellite, point-to-point and mobile terrestrial radio). Section 18.5 discusses the composite voiceband telephone channel, which is often used for voiceband data transmission. Finally, Section 18.6discusses magnetic recording of digital data, as used in tape and disk drives, which has characteristics similar in many ways to the other channels discussed.

The most prevalent media for new installations in the future will be optical fiber and microwave radio, and possibly lossless transmission lines based on superconducting materials. However, there is a continuing strong interest in lossy transmission lines and voiceband channels because of their prevalence in existing installations. Thus all the media discussed in this chapter are important in new applications of digital communication.

18.1. COMPOSITE CHANNELS

It is common for many users to share a common communication medium, for example by *time-division* and *frequency-division multiplexing* (Chapter 16).

Example 18-1. Voice signals are roughly bandlimited to frequencies lower than 4 kHz. A suitable baseband channel therefore needs to pass only frequencies up to 4 kHz. Such a channel is often derived from a much higher bandwidth physical medium that is shared with other users. A *voice frequency (VF)* channel derived from a coaxial cable (Section 18.2) using *single-sideband modulation* is shown in Fig. 18-1. The SSB modulator translates the VF channel to the neighborhood of a frequency ω_c for transmission on the coaxial cable. A VF channel can be used for digital communication, as long as the modulation technique conforms to the limitations of the channel.

The channel in Fig. 18-1 is an example of a *composite channel*, because it consists of multiple subsystems. If the VF channel was designed for voice transmission, it has certain characteristics which are beyond the control of the designer of the digital communication system. The VF channel characteristics in this case depend not only on the properties of the physical medium, but also on the design of the SSB modulation system.

Composite channels usually arise in the context of *multiple access*, which is defined as access to a physical medium by two or more independent users. This is again illustrated by example.

Example 18-2. Fig. 18-2 shows a *frequency-division multiplexing (FDM)* approach using SSB modulation. In this case two VF channels are derived from a single coaxial cable using SSB modulation. The two channels are separated by using two different carrier frequencies in the two modulators, f_1 and f_2, where these frequencies are chosen far enough apart that the spectra for the

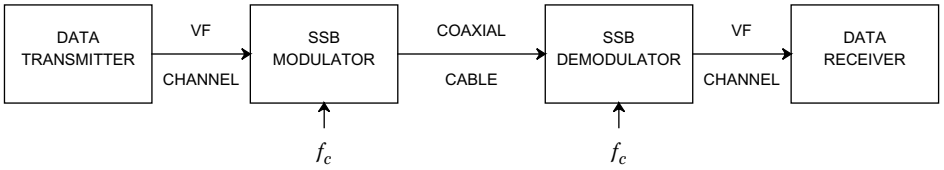

Fig. 18-1. Composite data channel derived from an SSB modulation system, where f_c is the carrier frequency.

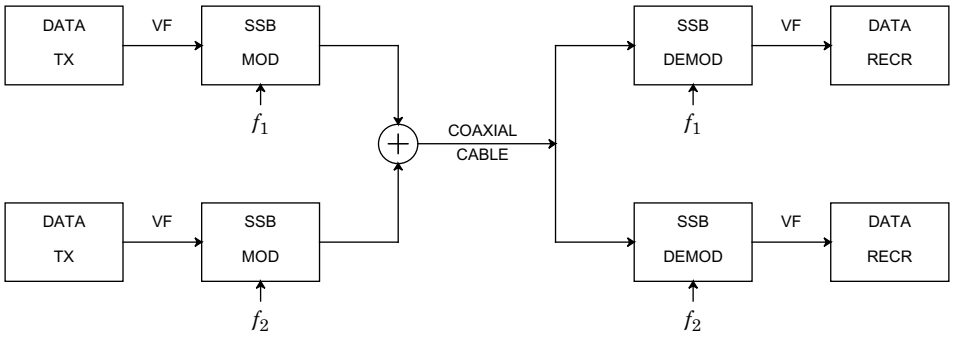

Fig. 18-2. Two data channels derived from a single coaxial cable by FDM, where f_1 and f_2 are distinct carrier frequencies. "TX" is the transmitter and "RECR" is the receiver.

two modulated VF channels do not overlap. These two VF channels can be used independently for digital communication.

In fact FDM is a very common technique in the telephone network for deriving many VF channels from a single physical medium such as coaxial cable or microwave radio (thousands rather than just two as in the example!).

Another common composite channel is illustrated in the following example.

Example 18-3. A VF channel derived from a digital transmission system using *pulse-code modulation (PCM)* is illustrated in Fig. 18-3. The PCM system samples the VF channel at 8 kHz, corresponding to a maximum bandwidth of 4 kHz, and then quantizes each sample to eight bits. The total bit rate for the PCM encoded VF channel is 64 kb/s. This derived VF channel may be used for data transmission; again, any digital modulation technique can be used subject to basic constraints imposed by the PCM system. The total bit rate that can be transmitted through this derived VF channel is less than 64 kb/s, in fact more of the order of 20–30 kb/s. The direct transmission of the bit stream over the digital transmission system would obviously be more efficient, but the situation in Fig. 18-3 is still very common due to the presence of much existing PCM equipment for voice transmission and the desire to transmit data over a channel designed primarily for voice.

Physical media as well as the composite channels derived from them impose constraints on the design of a digital communication system. Many of these constraints will be mentioned in this chapter for particular media. The nature of these constraints usually fall within some broad categories:

- A *bandwidth constraint*. Sometimes this is in the form of a channel attenuation which increases gradually at high frequencies, and sometimes (particularly in the case of composite channels) it is in the form of a very hard bandwidth limit.
- A *transmitted power constraint*. This is often imposed to limit interference of one digital communication system with another, or imposed by the inability of a composite channel to transmit a power level greater than some threshold, or by a limitation imposed by the power supply voltage of the digital communication system itself. This power constraint can be in the form of a *peak-power constraint*, which is essentially a limit on the transmitted voltage, or can be an *average power constraint*.

Example 18-4. An FDM system such as that in Fig. 18-2, where there are perhaps thousands of channels multiplexed together, is designed under assumptions on the average power of the channels. From this average power, the total power of the multiplexed signal can be deduced; this power is adjusted relative to the point at which amplifiers in the system start to become nonlinear. If a significant number of VF channels violate the average power constraint, then the multiplexed signal will overload the amplifiers, and the resulting nonlinearity will cause *intermodulation distortion* and interference between VF channels.

Example 18-5. The PCM system of Fig. 18-3 imposes a bandwidth constraint on the data signal, which must be less than half the sampling rate. A peak power constraint is also imposed by the fact that the quantizer in the PCM system has an overload point beyond which it clips the input signal.

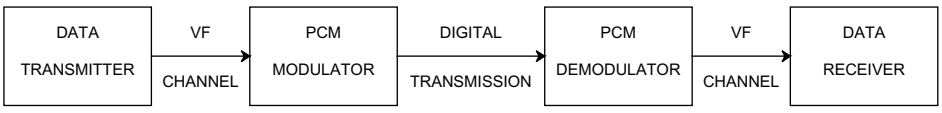

Fig. 18-3. Data transmission over a PCM-derived VF channel.

Example 18-6. The regenerative repeaters in Fig. 1-3 are usually powered by placing a high voltage on the end of the system and then stringing the repeaters in series (like Christmas tree lights!); a fraction of the total voltage appears across each repeater. The power consumption of the repeaters is limited by the applied voltage, the ohmic loss of the cable, and the number of repeaters. This places an average power constraint on the transmitted signal at each repeater. In practice there is also a peak power constraint due to the desire not to have to generate a signal voltage higher than the supply voltage drop across the repeater. An additional factor limiting the transmitted power for wire-pair systems is crosstalk into other communication systems in the same multi-pair cable, as discussed in Section 18.2.4.

Example 18-7. The optical fiber medium (Section 18.3) becomes significantly nonlinear when the input power exceeds about one milliwatt. Thus, in many applications there is a practical limit on the average transmitted power.

In addition to placing constraints on the transmitted signal, the medium or composite channel introduces *impairments* which limit the rate at which we can communicate. We will see many examples of this in this chapter.

18.2. TRANSMISSION LINES

One of the most common media for data transmission in the past has been the transmission line composed of a pair of wires or a coaxial cable. Coaxial cable is commonly used for digital communication within a building, as in a local-area network, and for high-capacity long-distance facilities in the telephone network. Wire pairs are much more extensively used, primarily for relatively short distance trunking in the telephone network between switching machines in metropolitan areas. The spacing between regenerative repeaters is typically about 1.5 km, with bit rates in the range of 1.5–6 Mb/s on wire pair to 270–400 Mb/s on coaxial cable. In addition, wire pairs are used for connection of the telephone instrument to the central office, and while this connection is primarily for analog voiceband transmission, there is work proceeding to use this same medium for digital communication at 144 kb/s or higher in the Integrated Services Digital Network (ISDN). This is called the *digital subscriber loop*, and requires a distance for transmission of about 4-5 kilometers without repeaters.

18.2.1. Review of Transmission Line Theory

A *uniform transmission line* is a two-conductor cable with a uniform cross-section. It may consist of a pair of wires twisted together (*twisted wire cable*) or a cable with a cylindrical outer conductor surrounding a wire (*coaxial cable*). While the details of the cable characteristics depend on the cross-section geometry, the basic theory does not.

A uniform transmission line can be represented by a pair of conductors as shown in Fig. 18-4. We denote the termination of the line on the right as $x = 0$, and the source of the line on the left as $x = -L$, where x is the distance along the line and L is the length of the line. Assume that the line is excited with a complex exponential with radian frequency ω. The the voltage and current along the line will be a function of both the frequency ω and the distance x. Writing the voltage and current at a point x, the dependence on time is given by the complex exponential,

$$V(x, \omega) = V(x)e^{f t}, \qquad I(x, \omega) = I(x)e^{f t}, \tag{18.1}$$

where $V(x)$ and $I(x)$ are complex numbers which summarize the amplitude and phase of the complex exponential at distance x.

The voltage and current as a function of distance along the line consists of two propagating waves, one from source to termination and the other from termination to source. The first we call the *source wave*, and the latter we call the *reflected wave*. The total voltages and currents are the sum of the two waves, given by

$$V(x) = V_+e^{-\gamma x} + V_-e^{\gamma x}, \qquad I(x) = \frac{1}{Z_0}(V_+e^{-\gamma x} - V_-e^{\gamma x}). \tag{18.2}$$

In these equations the V_+ terms correspond to the source wave, and the V_- terms correspond to the reflected wave. The complex impedance Z_0 is called the *characteristic impedance* of the transmission line, since it equals the ratio of the voltage to current at any point of the line (independent of x) for either the source or reflected wave. The other complex quantity in this equation is γ, which is called the *propagation constant*. The real and imaginary parts of γ are of importance in their own right, so in

$$\gamma = \alpha + j\beta \tag{18.3}$$

the real part α and imaginary part β are called respectively the *attenuation constant* and *phase constant*. The attenuation constant has the units of *nepers per unit distance* and the phase constant has the units of *radians per unit distance*.

There are three things that distinguish the source wave and its reflection:

- The amplitude and phase as expressed by V_+ and V_- are different.
- The current is flowing in opposite directions.
- The sign of the exponent is different.

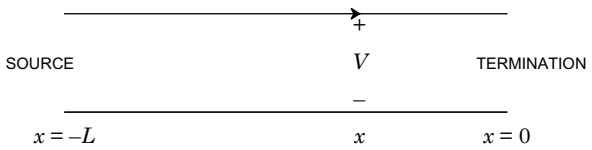

Fig. 18-4. A uniform transmission line, where x is the distance along the line.

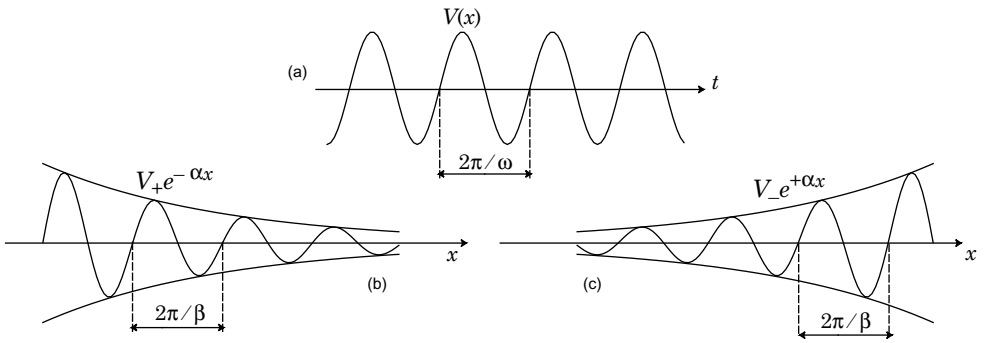

Fig. 18-5. The voltage on a uniform transmission line. a. The voltage at one point in the line as a function of time. b. The magnitude of the voltage vs. distance for the source wave at a fixed time. c. (b) repeated for the reflected wave.

The third difference is illustrated in Fig. 18-5. The dependence of both waves on time at any point along the line is shown in Fig. 18-5a. This is of course just a sinusoid of constant amplitude and phase, where both amplitude and phase depend on the frequency of the wave. For a fixed time $t = t_0$, the source wave is shown in Fig. 18-5b as a function of distance x, and is given by

$$V(x) = V_+ e^{-\alpha x} e^{-j\beta x}. \tag{18.4}$$

The amplitude of this wave decreases with distance x. The *wavelength* of the wave, or distance between nulls, is $2\pi/\beta$.

The phase shift of the complex exponential with radian frequency ω for a transmission line of length L is βL radians, which corresponds to $\beta L / 2\pi$ cycles. This phase shift represents a propagation delay of the sinusoid, and we can readily figure out the size of the delay. Since each cycle corresponds to $2\pi/\omega$ seconds from Fig. 18-5a, it follows that the total delay of the complex exponential is

$$\frac{\beta L}{2\pi} \text{ cycles} \cdot \frac{2\pi}{\omega} \frac{\text{sec}}{\text{cycles}} = \frac{\beta}{\omega} L \text{ sec}. \tag{18.5}$$

The propagation velocity of the wave on the transmission line is therefore related to the frequency and phase constant by

$$v = \frac{\omega}{\beta}. \tag{18.6}$$

Since α is always greater than zero, the magnitude of the wave is also decaying exponentially with distance in accordance with the term $e^{-\alpha x}$. This implies that at any frequency the loss of the line in dB is proportional to the length of the line. We get a power loss in dB

$$\gamma_0 \cdot L = 10 \log_{10}\left(\frac{P_T}{P_R}\right), \tag{18.7}$$

where $\gamma_0 = 20\alpha \log_{10} e$ is the loss in dB per unit distance. Since α is frequency dependent, so too is γ_0.

Similarly, shown in Fig. 18-5c is the reflected wave amplitude as a function of distance along the line. This wave is also decaying exponentially with distance in the direction of propagation, which is from termination to source.

Based on these relationships, we can determine the voltage along the line for any source and termination impedances by simply matching up boundary conditions.

Example 18-8. A transmission line terminated in impedance Z_L is shown in Fig. 18-6a. What is the relative size of the incident and reflected waves at the termination? This quantity is called the *voltage reflection coefficient* and is usually denoted by Γ. The boundary condition is that Z_L is the ratio of the voltage to current at $x = 0$, so from (18.2),

$$Z_L = Z_0 \frac{V_+ + V_-}{V_+ - V_-}, \qquad \Gamma = \frac{V_-}{V_+} = \frac{Z_L - Z_0}{Z_L + Z_0}. \tag{18.8}$$

Several special cases are of interest. When the load impedance is equal to the characteristic impedance, $Z_L = Z_0$, then the reflection coefficient is zero, $\Gamma = 0$. When the line is open circuited, $Z_L = \infty$, then $\Gamma = 1$ indicating that the reflected voltage is the same as the incident wave at the point of the open circuit. Finally, when the line is closed circuited, $Z_L = 0$, then $\Gamma = -1$ indicating that the reflected voltage is the negative of the incident voltage.

Example 18-9. For the terminated transmission line of Fig. 18-6a, what is the impedance looking into the line as a function of its length? Taking the ratio of the voltage to current from (18.2) and (18.8),

$$\frac{V(x)}{I(x)} = Z_0 \frac{e^{-\gamma x} + \Gamma e^{\gamma x}}{e^{-\gamma x} - \Gamma e^{\gamma x}}, \qquad Z_{\text{in}} = \frac{V(-L)}{I(-L)} = Z_0 \frac{1 + \Gamma e^{-2\gamma x}}{1 - \Gamma e^{-2\gamma x}}. \tag{18.9}$$

When the line is terminated in its characteristic impedance, $\Gamma = 0$ and the input impedance is equal to the characteristic impedance.

Example 18-10. For the terminated line of Fig. 18-6b with a source impedance of Z_S, what is the voltage transfer function from source V_{in} to the load? Writing the node voltage equation at the source,

$$V(-L) = V_{in} - I(-L)Z_S \tag{18.10}$$

and we have the two additional relations

$$V(-L) = V_+(e^{\gamma L} + \Gamma e^{-\gamma L}), \qquad I(-L) = \frac{V_+}{Z_0}(e^{\gamma L} - \Gamma e^{-\gamma L}), \tag{18.11}$$

which enable us to solve for the three constants V_+, $V(-L)$, and $I(-L)$. Finally, the output voltage is

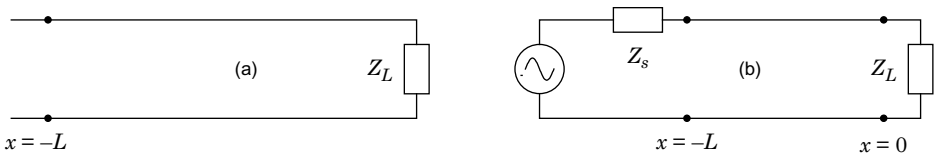

Fig. 18-6. A terminated transmission line. a. Without a source termination. b. With a source termination.

$$V(0) = V_+(1 + \Gamma) \, . \tag{18.12}$$

Putting this all together, the desired voltage transfer function is

$$\frac{V(0)}{V_{\text{in}}} = \frac{Z_0(1 + \Gamma)}{(Z_0 + Z_S)e^{\gamma L} + \Gamma(Z_0 - Z_S)e^{-\gamma L}} \, . \tag{18.13}$$

When the source impedance is equal to the characteristic impedance, $Z_S = Z_0$ this simplifies to

$$\frac{V(0)}{V_{\text{in}}} = \frac{1 + \Gamma}{2} e^{-\gamma L}. \tag{18.14}$$

When further the line is terminated in its characteristic impedance, $\Gamma = 0$ and the transfer function consists of the attenuation and phase shift of the transmission line (times another factor of 0.5 corresponding to the attenuation due to the source and termination impedances). When the line is short-circuited, then the transfer function is zero as expected since $\Gamma = -1$. What happens when the line is open circuited?

Transmission lines are often analyzed, particularly where computer programs are written, using the concept of a *chain matrix* [1]. A *twoport network* is a network as illustrated in Fig. 18-7, which has an input port and output port and for which the input and output currents are complementary as shown. The chain matrix relates the input voltage and current to the output voltage and current, *viz.*

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}, \tag{18.15}$$

where all quantities are complex functions of frequency. The chain matrix characterizes the twoport transfer function completely, and can be used to analyze connections of twoports (such as transmission lines, thereby serving to analyze nonuniform transmission lines). Its importance arises from the following fact.

Exercise 18-1. Show that if two twoports are connected in series, then the chain matrix of the combination twoport is the product of the first chain matrix times the second chain matrix.

The chain matrix of a uniform transmission line is easily calculated, giving us a ready technique for systematically analyzing combinations of transmission lines with other circuit elements.

Exercise 18-2. Show that the chain matrix of a uniform transmission line is

$$\begin{bmatrix} \cosh(\gamma L) & Z_0 \sinh(\gamma L) \\ \sinh(\gamma L)/Z_0 & \cosh(\gamma L) \end{bmatrix}. \tag{18.16}$$

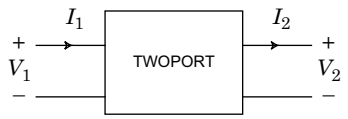

Fig. 18-7. Illustration of a twoport network with definition of voltages and currents for the chain matrix.

18.2.2. Cable Primary Constants

The characteristic impedance and propagation constant are called *secondary parameters* of the cable because they are not related directly to physical parameters. A simple model for a short section of the transmission line is shown in Fig. 18-8. This model is in terms of four parameters: the conductance G in mhos per unit length, the capacitance C in farads per unit length, the inductance L in henries per unit length, and the resistance R in ohms per unit length. All of these parameters of the transmission line are functions of frequency in general, and they differ for different cross-sections (for example, twisted pair vs. coaxial cable). In general these parameters are determined experimentally for a given cable.

This lumped-parameter model becomes an exact model for the transmission line as the length of the line $dx \to 0$, and is useful since it displays directly physically meaningful quantities. These parameters are called the *primary constants* of the transmission line. The secondary parameters can be calculated directly in terms of the primary constants as

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \qquad \gamma = \sqrt{(R + j\omega L)(G + j\omega C)}\,. \tag{18.17}$$

Example 18-11. A *lossless transmission* line is missing the two dissipative elements, resistance and conductance. New superconducting materials show promise of actually being able to realize this ideal. The secondary parameters in this case are

$$Z_0 = \sqrt{\frac{L}{C}}, \qquad \gamma = f\sqrt{LC}\,. \tag{18.18}$$

The characteristic impedance of a lossless transmission line is real-valued and hence resistive. A pure resistive termination is often used as a reasonable approximation to the characteristic impedance of a *lossy* transmission line, although the actual characteristic impedance increases at low frequencies and includes a capacitive reactive component. The propagation constant is imaginary, and since $\alpha = 0$ the lossless transmission line has no attenuation as expected. The propagation velocity on a lossless transmission line is

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}}\,. \tag{18.19}$$

The primary constants of actual cables depend on many factors such as the geometry and the material used in the insulation. For twisted wire pairs [2] the capacitance is independent of frequency for the range of frequencies of interest (0.083 μFarads per mile, or 0.0515 μFarads per kilometer is typical), the conductance is negligibly small, the inductance is a slowly varying function of frequency decreasing from about one milliHenries (mH) per mile or 0.62 mH per

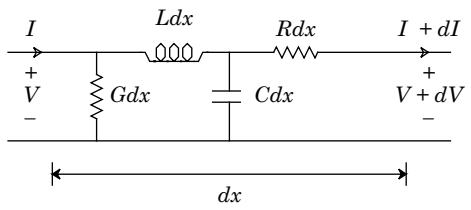

Fig. 18-8. Lumped-parameter model for a short section of transmission line.

km, at low frequencies to about 70% of that value at high frequencies, and the resistance is proportional to the square root of frequencies at high frequencies due to the *skin effect* (the tendency of the current to flow near the surface of the conductor, increasing the resistance).

Example 18-12. What is the velocity of propagation on a twisted wire cable? From (18.19), for a lossless line

$$v = \frac{1}{\sqrt{(0.083 \cdot 10^{-6})(10^{-3})}} = 1.1 \cdot 10^5 \text{ miles/sec} = 1.76 \cdot 10^5 \text{ km/sec}. \tag{18.20}$$

Example 18-13. Since the speed of light in free space is $3 \cdot 10^5$ km/sec, the velocity on the line is a little greater than half the speed of light. The delay is about 5.65 μsec per km. This approximation is valid on practical twisted wire pairs for frequencies where $R \ll \omega L$.

Coaxial cable is popular for higher frequency applications primarily because the outer conductor effectively shields against radiation to the outside world and conversely interference from outside sources. At lower frequencies near voiceband this shielding is ineffective and hence the coaxial cable does not have any advantage over the more economical twisted wire pair. In terms of primary constants, the main difference between coaxial cable and wire pair is that the coaxial inductance is essentially independent of frequency.

Example 18-14. A more accurate model than Example 18-11 of a cable, wire pair or coaxial, would be to assume that G only is zero. Then the propagation constant of (18.17) becomes

$$\alpha = \omega \sqrt{\frac{LC}{2}} \left\{ \sqrt{1 + \frac{R^2}{\omega^2 L^2}} - 1 \right\}^{1/2}, \qquad \beta = \omega \sqrt{\frac{LC}{2}} \left\{ \sqrt{1 + \frac{R^2}{\omega^2 L^2}} + 1 \right\}^{1/2}. \tag{18.21}$$

At frequencies where $R \ll \omega L$,

$$\alpha \approx \frac{R}{2} \sqrt{\frac{C}{L}} \quad \text{nepers per unit length,} \tag{18.22}$$

$$\beta \approx \omega \sqrt{LC}. \tag{18.23}$$

Hence the velocity relation of (18.19) is still valid in this range of frequencies. Since at high frequencies R increases as the square root of frequency, the attenuation constant in nepers (or dB) has the same dependency. It follows that the loss of the line in dB at high frequencies is proportional to the square root of frequency.

Example 18-15. If the loss of a cable is 40 dB at 1 Mhz, what is the approximate loss at 4 MHz? The answer is 40 dB times the square root of 4, or 80 dB.

The results of Example 18-14 suggest that the propagation constant is proportional to frequency. This *linear phase* model suggests that the line offers, in addition to attenuation, a constant delay at all frequencies. However, a more refined model of the propagation constant [3] shows that there is an additional term in the phase constant proportional to the square root of frequency. This implies that the cable will have some *group delay*, which is delay dependent on frequency. This implies that the different frequency components of a pulse launched into the line will arrive at the termination with slightly different delays. Both the frequency-dependent attenuation and the group delay cause *dispersion* on the transmission line, or spreading in time

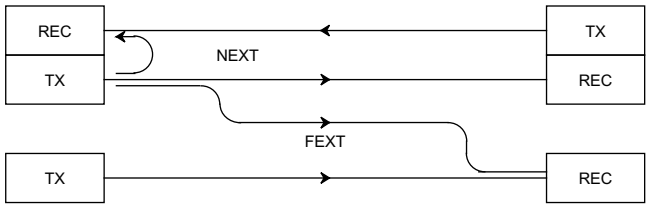

Fig. 18-9. Illustration of two types of crosstalk: far-end crosstalk (FEXT) and near-end crosstalk (NEXT).

of a transmitted pulse. The attenuation causes dispersion because the band-limiting effect broadens the pulse, and delay distortion causes dispersion because the different frequency components arrive with different delays.

18.2.3. Impedance Discontinuities

The theory presented thus far has considered a single uniform transmission line. In practice, it is common to encounter different *gauges* (diameters) of wire connected together. These gauges changes do not affect the transmission materially, except for introducing a slight discontinuity in impedance, which will result in small reflections. A more serious problem for digital transmission in the subscriber loop between central office and customer premises is the *bridged tap*, an additional open circuited wire pair bridged onto the main cable pair.

18.2.4. Crosstalk

An important consideration in the design of a digital communication system using a transmission line as a physical medium is the *range* or *distance* which can be achieved between regenerative repeaters. This range is generally limited by the high frequency gain which must be inserted into the receiver equalization to compensate for cable attenuation. This gain amplifies noise and interference signals which may be present, causing the signal to deteriorate as the range increases. The most important noise and interference signals are thermal noise (due to random motion of electrons), impulse noise (caused by switching relays and similar mechanisms), and crosstalk between cable pairs. Crosstalk, and interference from external sources such as power lines, can be minimized by using *balanced transmission*, in which the signal is transmitted and received as a difference in voltage between the two wires; this helps because external interference couples approximately equally into the two wires and hence is approximately canceled when the difference in voltage is taken at the receiver. A common way to achieve balanced transmission is to use *transformer coupling* of the transmitter and receiver to the wire pair; in addition, this affords additional protection against damage to the electronics due to foreign potentials such as lightning strikes.

There are two basic crosstalk mechanisms, *near-end crosstalk (NEXT)* and *far-end crosstalk (FEXT)*, illustrated in Fig. 18-9. NEXT [4] represents a crosstalk of a local transmitter into a local receiver, and experiences an attenuation which is accurately modeled by

$$|H_{NEXT}(f)|^2 = K_{NEXT}\ |f|^{1.5} \tag{18.24}$$

where $H_{NEXT}(f)$ is the transfer function experienced by the crosstalk. FEXT represents a crosstalk of a local transmitter into a remote receiver, with an attenuation given by

$$|H_{FEXT}(f)|^2 = K_{FEXT}|C(f)|^2|f|^2 \tag{18.25}$$

where $C(f)$ is the loss of the cable. Where present, NEXT will dominate FEXT because FEXT experiences the loss of the full length of the cable (in addition to the crosstalk coupling loss) and NEXT does not. Both forms of crosstalk experience less attenuation as frequency increases, and hence it is advantageous to minimize the bandwidth required for transmission in a crosstalk limited environment.

18.3. OPTICAL FIBER

The optical fiber cable is capable of transmitting light for long distances with high bandwidth and low attenuation. Not only this, but it offers freedom from external interference, immunity from interception by external means, and inexpensive and abundant raw materials. It is difficult to imagine a more ideal medium for digital communication!

The use of an optical dielectric waveguide for high performance communication was first suggested by Kao and Hockham in 1966 [5]. By 1986 this medium was well developed and was rapidly replacing wire pairs and coax in many new cable installations. It allows such a great bandwidth at modest cost that it will also replace many of the present uses for satellite and radio transmission. Thus, it appears that digital communication over wire-pairs and coax will be mostly limited to applications constrained to use existing transmission facilities (such as in the digital subscriber loop).

Digital transmission by satellite and radio will be limited to special applications that can make use of their special properties. For example, radio communication is indispensable for situations where one or both terminals are mobile, for example in digital mobile telephony (Section 18.4) or deep space communication. Radio is also excellent for easily bridging geographical obstacles such as rivers and mountains, and satellite is excellent for spanning long distances where the total required bandwidth is modest and the installation of an optical fiber cable would not be justified. Furthermore, satellite has unique capabilities for certain types of multiple-access situations (Chapter 18) spread over a wide geographical area.

18.3.1. Fiber Optic Waveguide

The principle of an optical fiber waveguide [6] can be understood from the concept of *total internal reflection*, shown in Fig. 18-10. A light wave, represented by a single ray, is incident on a boundary between two materials, where the angle of incidence is θ_1 and the angle of refraction is θ_2. We define a *ray* as the path that the center of a slowly diverging beam of light takes as it passes through the system; such a beam must have a diameter large with respect to the wavelength in order to be approximated as a plane wave [7]. Assuming that the index of refraction n_1 in the incident material is greater than the index of refraction of the refraction medium, n_2, or $n_1 > n_2$. Then Snell's Law predicts that

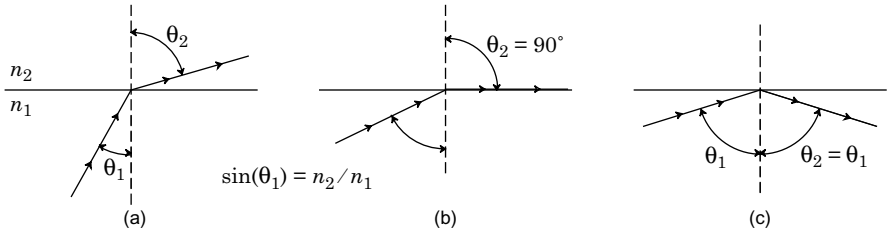

Fig. 18-10. Illustration of Snell's Law and total internal reflection. a.Definition of angles of incidence θ_1 and refraction θ_2. The angle of refraction is larger if $n_1 > n_2$. b. The critical angle of incidence at which the angle of refraction is ninety degrees. c. At angles larger than the critical angle, total internal reflection occurs.

$$\frac{\sin(\theta_1)}{\sin(\theta_2)} = \frac{n_2}{n_1} < 1 \,. \tag{18.26}$$

The angle of refraction is larger than the angle of incidence. Shown in Fig. 18-10b is the case of a critical incidence angle where the angle of refraction is ninety degrees, so that the light is refracted along the material interface. This corresponds to *critical* incident angle

$$sin(\theta_1) = \frac{n_2}{n_1} \,. \tag{18.27}$$

For angles larger than (18.27), there is total internal reflection as illustrated in Fig. 18-10c, where the angle of reflection is always equal to the angle of incidence.

This principle can be exploited in an *optical fiber waveguide* as illustrated in Fig. 18-11. The *core* and *cladding* materials are glass, which transmits light with little attenuation, while the *sheath* is an opaque plastic material that serves no purpose other than to lend strength, absorb any light that might otherwise escape, and prevent any light from entering (which would represent interference or crosstalk). The core glass has a higher index of refraction than the cladding, with the result that incident rays with a small angle of incidence are captured by total internal reflection. This is illustrated in Fig. 18-12, where a light ray incident on the end of the fiber is captured by total internal reflection as long as the angle of incidence θ_1 is below a critical angle (Problem 18-4). The ray model predicts that the light will bounce back and forth, confined to the waveguide until it emerges from the other end. Furthermore, it is obvious that the path length of a ray, and hence the transit time, is a function of the incident angle of the ray (Problem 18-5).

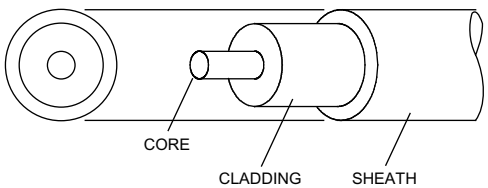

Fig. 18-11. An optical fiber waveguide. The core and cladding serve to confine the light incident at narrow incident angles, while the opaque the sheath serves to give mechanical stability and prevent crosstalk or interference.

This variation in transit time for different rays manifests itself in *pulse broadening* — the broadening of a pulse launched into the fiber as it propagates — which in turn limits the pulse rate which can be used or the distance that can be transmitted or both. The pulse broadening can be reduced by modifying the design of the fiber, and specifically by using a *graded-index fiber* in which the index of refraction varies continuously with radial dimension from the axis.

The foregoing ray model gives some insight into the behavior of light in an optical fiber waveguide; for example, it correctly predicts that there is a greater pulse broadening when the index difference between core and cladding is greater. However, this model is inadequate to give an accurate description since in practice the radial dimensions of the fiber are on the order of the wavelength of the light. For example, the ray model of light predicts that there is a continuum of angles for which the light will bounce back and forth between core-cladding boundaries indefinitely. A more refined model uses Maxwell's equations to predict the behavior of light in the waveguide, and finds that in fact there are only a discrete and finite number of angles at which light propagates in zigzag fashion indefinitely. Each of these angles corresponds to a *mode* of propagation, similar to the modes in a metallic waveguide carrying microwave radiation. When the core radius is many times larger than the wavelength of the propagating light, there are many modes; this is called a *multimode* fiber. As the radius of the core is reduced, fewer and fewer modes are accommodated, until at a radius on the order of the wavelength only one mode of propagation is supported. This is called a *single mode* fiber. For a single mode fiber the ray model is seriously deficient since it depends on physical dimensions that are large relative to the wavelength for its accuracy. In fact, in the single mode fiber the light is not confined to the core, but in fact a significant fraction of the power propagates in the cladding. As the radius of the core gets smaller and smaller, more and more of the power travels in the cladding.

For various reasons, as we will see, the transmission capacity of the single mode fiber is greater. However, it is also more difficult to splice with low attenuation, and it also fails to capture light at the larger incident angles that would be captured by a multimode fiber, making it more difficult to launch a given optical power. In view of its much larger ultimate capacity, there is a trend toward exclusive use of single mode fiber in new installations, even though multimode fiber has been used extensively in the past [6]. In the following discussion, we emphasize the properties of single mode fiber.

We will now discuss the factors which limit the bandwidth or bit rate which can be transmitted through a fiber of a given length. The important factors are:

- *Material attenuation*, the loss in signal power that inevitably results as light travels down an optical waveguide. There are four sources of this loss in a single mode fiber — scattering of the light by inherent inhomogeneities in the molecular structure of the glass

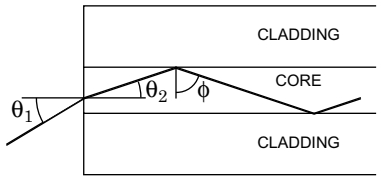

Fig. 18-12. Ray model of propagation of light in an optical waveguide by total internal reflection. Shown is a cross-section of a fiber waveguide along its axis of symmetry, with an incident light ray at angle θ_1 which passes through the axis of the fiber (a meridional ray).

crystal, absorption of the light by impurities in the crystal, losses in connectors, and losses introduced by bending of the fiber. Generally these losses are affected by the wavelength of the light, which affects the distribution of power between core and cladding as well as scattering and absorption mechanisms. The effect of these attenuation mechanisms is that the signal power loss in dB is proportional to the length of the fiber. Therefore, for a line of length L, if the loss in dB per kilometer is γ_0, the total loss of the fiber is $\gamma_0\ L$ and hence the ratio of transmitted power P_T to received power P_R obeys

$$\gamma_0 L = 10 \cdot \log_{10} \frac{P_T}{P_R}, \qquad P_R = P_T \cdot 10^{-\gamma_0 L/10}. \tag{18.28}$$

This exponential dependence of loss vs. length is the same as for the transmission lines of Section 18.2.

- *Mode dispersion*, or the difference in group velocity between different modes, results in the broadening of a pulse which is launched into the fiber. This broadening of pulses results in interference between successive pulses which are transmitted, called *intersymbol interference* (Chapter 8). Since this pulse broadening increases with the length of the fiber, this dispersion will limit the distance between regenerative repeaters. One significant advantage of single mode fibers is that mode dispersion is absent since there is only one mode.
- *Chromatic* or *material dispersion* is caused by differences in the velocity of propagation at different wavelengths. For infrared and longer wavelengths, the shorter wavelengths arrive earlier than relatively longer wavelengths, but there is a crossover point at about *1.3* μm beyond which relatively longer wavelengths arrive earlier. Since practical optical sources have a non-zero bandwidth, called the *linewidth*, and signal modulation increases the optical bandwidth further, material dispersion will also cause intersymbol interference and limit the distance between regenerative repeaters. Material dispersion is qualitatively similar to the dispersion that occurs in transmission lines (Section 18.2) due to frequency-dependent attenuation. The total dispersion is usually expressed in units of picoseconds pulse spreading per GHz source bandwidth per kilometer distance, with typical values in the range of zero to 0.15 in the 1.3–1.6 μmeter minimum attenuation region [9][10]. It is very important that since the dispersion passes from positive to negative in the region of 1.3 μmeter wavelength, the dispersion is very nearly zero at this wavelength. A typical curve of the magnitude of the chromatic dispersion vs. wavelength is shown in Fig. 18-13, where the zero is evident. The chromatic dispersion can be made negligibly small over a relatively wide range of wavelengths. Furthermore, the frequency

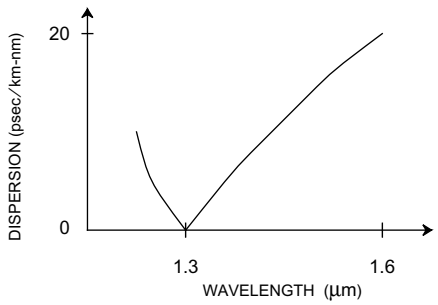

Fig. 18-13. Typical chromatic dispersion in silica fiber [10]. Shown is the magnitude of the dispersion; the direction of the dispersion actually reverses at the zero-crossing.

of this zero in chromatic dispersion can be shifted through waveguide design to correspond to the wavelength of minimum attenuation.

With these impairments in mind, we can discuss the practical and fundamental limits on information capacity for a fiber. The fundamental limit on attenuation is due to the intrinsic material scattering of the glass in the fiber — this is known as *Rayleigh scattering*, and is similar to the scattering in the atmosphere of the earth that results in our blue sky. The scattering loss decreases rapidly with wavelength (as the fourth power), and hence it is generally advantageous to choose a longer wavelength. The attenuation due to intrinsic absorption is negligible, but at certain wavelengths large attenuation due to certain impurities is observed. Particularly important are hydroxyl (OH) radicals in the glass, which absorb at 2.73 μmeters wavelength and harmonics. At long wavelengths there is infrared absorption associated fundamentally with the glass, which rises sharply starting at 1.6 μmeters.

A loss curve for a state-of-the-art fiber is shown in Fig. 18-14. Note the loss curves for two intrinsic effects which would be present in an ideal material, Rayleigh scattering and infrared absorption, and additional absorption peaks at 0.95, 1.25, and 1.39 μm due to OH impurities. The lowest losses are at approximately 1.3 and 1.5 μm, and these are the wavelengths at which the highest performance systems operate. The loss is as low as about 0.2 dB/km, implying potentially a much larger repeater spacing for optical fiber digital communication systems as compared to wire-pairs and coax. A curve of attenuation vs. frequency in Fig. 18-15 for wire cable media and for optical fiber illustrates that the latter has a much lower loss.

The loss per unit distance of the fiber is a much more important determinant of the distance between repeaters than is the bit rate at which we are transmitting. This is illustrated for a single-mode fiber in Fig. 18-16, where there is an attenuation-limited region where the curve of repeater spacing vs. bit rate is relatively flat. As we increase the bit rate, however, we eventually approach a region where the repeater spacing is limited by the dispersion (mode dispersion in a multimode fiber and chromatic dispersion in a single mode fiber). The magnitude of the latter can be quantified simply by considering the Fourier transform of a transmitted pulse, and in

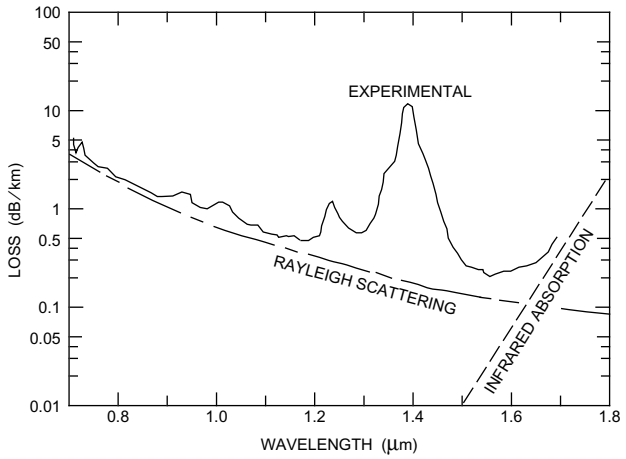

Fig. 18-14. Observed loss spectrum of an ultra-low-loss germanosilicate single mode fiber together with the loss due to intrinsic material effects [11].

particular its bandwidth W. The spreading of the pulse will be proportional to the repeater spacing L and the bandwidth W, with a constant of proportionality D. Thus, if we require that this dispersion be less than half a pulse-time at a pulse rate of R pulses per second,

$$D \cdot L \cdot W < \frac{1}{2R} . \tag{18.29}$$

The bandwidth W of the source depends on the *linewidth*, or intrinsic bandwidth in the absence of modulation, and also on the *modulation*. Since a non-zero linewidth will increase the bandwidth and hence the chromatic dispersion, we can understand fundamental limits by assuming zero linewidth. We saw in Chapter 5 that the bandwidth due to modulation is approximately equal to the pulse rate, or $W \approx R$, and hence (18.29) becomes

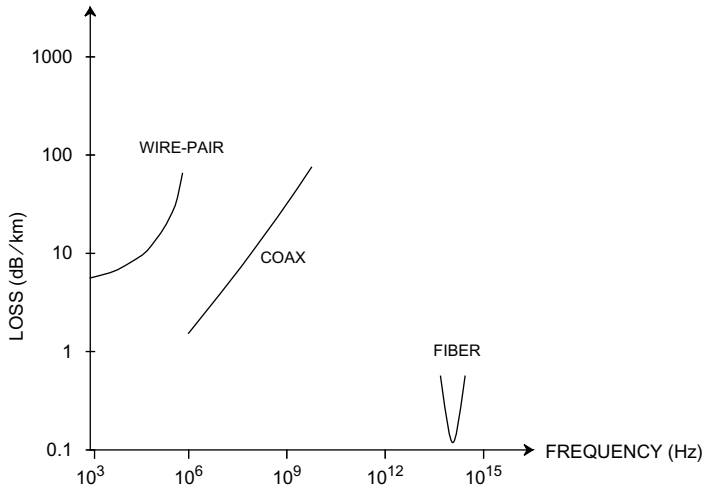

Fig. 18-15. Attenuation vs. frequency for wire cable and fiber guiding media [10].The band of frequencies over which the fiber loss is less than 1 dB/km is more than 10^{14} Hz.

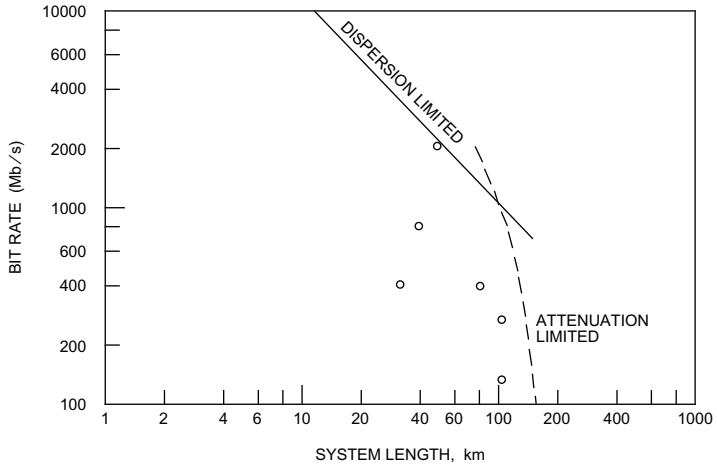

Fig. 18-16. Trade-off between distance and bit rate for a single mode fiber with a particular set of assumptions [8].The dots represent performance of actual field trial systems.

$$R^2 L < \frac{1}{2D}. \tag{18.30}$$

This equation implies that in the region where dispersion is limiting, the repeater spacing L must decrease rather rapidly as the bit rate is increased, as shown in Fig. 18-16. More quantitative estimates of the limits shown in Fig. 18-16 are derived in Problem 18-6 and Problem 18-11.

As a result of these considerations, first generation (about 1980) optical fiber transmission systems typically used multimode fiber at a wavelength of about 0.8 μmeter, and achieved bit rates up to about 150 Mb/s. The Rayleigh scattering is about 2 dB per km at this wavelength, and the distance between regenerative repeaters was in the 5 to 10 km range. Second generation systems (around 1985) moved to single mode fibers and wavelengths of about 1.3 μmeters, where Rayleigh scattering attenuation is about 0.2 dB/km (and practical attenuations are more on the order of 0.3 dB/km).

The finite length of manufactured fibers and system installation considerations dictate fiber connectors. These present difficult alignment problems, all the more difficult for single mode fibers because of the smaller core, but in practice connector losses of 0.1 to 0.2 dB can be obtained even for single mode fibers. Since it must be anticipated in any system installation that accidental breakage and subsequent splicing will be required at numerous points, in fact connector and splicing loss is the dominant loss in limiting repeater spacing.

Bending loss is due to the different propagation velocities required on the outer and inner radius of the bend. As the bending radius decreases, eventually the light on the outer radius must travel faster than the speed of light, which is of course impossible. What happens instead is that significant attenuation occurs due to a loss of confined power. Generally there is a trade-off between bending loss and splicing loss in single mode fibers, since bending loss is minimized by confining most of the power to the core, but that makes splicing alignment more critical.

In Fig. 18-16, the trade-off between maximum distance and bit rate is quantified for a single mode fiber for a particular set of assumptions (the actual numerical values are dependent on these assumptions). At bit rates below about one Gb/s (10^9 bits per second) the distance is limited by attenuation and receiver sensitivity. In this range the distance decreases as bit rate increases since the receiver sensitivity decreases (see Section 18.3.3). At higher bit rates, pulse broadening limits the distance before attenuation becomes important. The total fiber system capacity is best measured by a figure of merit equal to the product of the bit rate and the distance between repeaters (Problem 18-10), measured in Gb-km/sec. Current commercial systems achieve capacities on the order of 100 to 1000 Gb-km/sec.

18.3.2. Sources

While optical fiber transmission uses light energy to carry the information bits, at the present state of the art the signals are generated and manipulated electrically. This implies an electrical-to-optical conversion at the input to the fiber medium and an optical-to-electrical conversion at the output. There are two available light sources for fiber digital communication systems: the *light-emitting diode (LED)* and the *semiconductor injection laser*. The semiconductor laser is the more important for high-capacity systems, so we emphasize it here.

In contrast to the LED, the laser output is *coherent*, meaning that it is nearly confined to a single frequency. In fact the laser output does have non-zero linewidth, but by careful design the linewidth can be made small relative to signal bandwidths using a structure called *distributed feedback (DFB)*. Thus, coherent modulation and demodulation schemes are feasible, although commercial systems use intensity modulation. The laser output can be coupled into a single mode fiber with very high efficiency (about 3 dB power loss), and can generate powers in the 0 to 10 mW range [9] with one mW (0 dBm) typical [10]. The laser is necessary for single mode fibers, except for short distances, because it emits a narrower beam than the LED. The light output of the laser is very temperature dependent, and hence it is generally necessary to monitor the light output and control the driving current using a feedback circuit.

There is not much room for increasing the launched power into the fiber because of nonlinear effects which arise in the fiber [9], unless, of course, we can find ways to circumvent or exploit these nonlinear phenomena.

18.3.3. Photodetectors

The optical energy at the output of the fiber is converted to an electrical signal by a *photodetector*. There are two types of photodetectors available — the *PIN photodiode*, popular at about 100 Mb/s and below, and the *avalanche photodiode (APD)* (popular above 1 Gb/s) [12][10]. The cross-section of a PIN photodiode is shown in Fig. 18-17. This diode has an intrinsic (non-doped) region (not typical of diodes) between the n- and p-doped silicon. Photons of the received optical signal are absorbed and create hole-electron pairs. If the diode is reverse biased, there is an electric field across the depletion region of the diode (which includes the intrinsic portion), and this electric field separates the holes from electrons and sweeps them to the contacts, creating a current proportional to the incident optical power. The purpose of the intrinsic region is to enlarge the depletion region, thereby increasing the fraction of incident photons converted into current (carriers created outside the depletion region, or beyond diffusion distance of the depletion region, recombine with high probability before any current is generated).

The fraction of incident photons converted into carriers that reach the electrodes is called the *quantum efficiency* of the detector, denoted by η. Given the quantum efficiency, we can easily predict the current generated as a function of the total incident optical power. The energy of one photon is $h\nu$ where h is Planck's constant ($6.6 \cdot 10^{-34}$ Joule-sec) and ν is the optical frequency, related to the wavelength λ by

$$\nu\lambda = c\,, \tag{18.31}$$

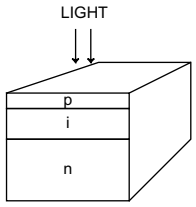

Fig. 18-17. A PIN photodiode cross-section. Electrode connection to the n-and p-regions creates a diode, which is reverse-biased.

where c is the speed of light ($3 \cdot 10^8$ m/sec). If the incident optical power is P watts, then the number of photons per second is $P/h\nu$, and if a fraction η of these photons generate an electron with charge q ($1.6 \cdot 10^{-19}$ Coulombs) then the total current is

$$i = \frac{\eta q P}{h\nu}. \tag{18.32}$$

Example 18-16. For a wavelength of 1.5 μm and quantum efficiency of unity, what is the *responsivity* (defined as the ratio of output current to input power) for a PIN photodiode? It is

$$\frac{i}{P} = \frac{q}{h\nu} = \frac{q\lambda}{hc} = \frac{(1.6 \cdot 10^{-19})(1.5 \cdot 10^{-6})}{(6.6 \cdot 10^{-34})(3.0 \cdot 10^8)} = 1.21 \text{ amps/watt}. \tag{18.33}$$

If the incident optical power is 1 nW, the maximum current from a PIN photodiode is 1.21 nA.

With PIN photodiodes (and more generally all photodetectors), there is a trade-off between quantum efficiency and speed. Quantum efficiencies near unity are achievable with a PIN photodiode, but this requires a long absorption region. But a long intrinsic absorption region results in a correspondingly smaller electric field (with resulting slower carrier velocity) and a longer drift distance, and hence slower response to an optical input. Higher speed inevitably results in reduced sensitivity.

Since very small currents are difficult to process electronically without adding significant thermal noise, it is desirable to increase the output current of the diode before amplification. This is the purpose of the APD, which has internal gain, generating more than one electron-hole pair per incident photon. Like the PIN photodiode, the APD is also a reverse-biased diode, but the difference is that the reverse voltage is large enough that when carriers are freed by a photon and separated by the electric field they have enough energy to collide with the atoms in the semiconductor crystal lattice. The collisions ionize the lattice atoms, generating a second electron-hole pair. These secondary carriers in turn collide with the lattice, and additional carriers are generated. One price paid for this gain mechanism is an inherently lower bandwidth. A second price paid in the APD is the probabilistic nature of the number of secondary carriers generated. The larger the gain in the APD, the larger the statistical fluctuation in current for a given optical power. In addition, the bandwidth of the device decreases with increasing gain, since it takes some time for the avalanche process to build up.

Both PIN photodiodes and APD's exhibit a small current which flows in the absence of incident light due to thermal excitation of carriers. This current is called *dark current* for obvious reasons, and represents a background noise signal with respect to signal detection.

18.3.4. Model for Fiber Reception

Based on the previous background material and the mathematics of Poisson processes and shot noise (Section 3.4) we can develop a statistical model for the output of an optical fiber detector. This signal has quite different characteristics from that of other media of interest, since random quantum fluctuations in the signal are important. Since the signal itself has random fluctuations, we can consider it to have a type of *multiplicative noise*.

In commercial systems, the *direct detection* mode of transmission is used, as pictured in Fig. 18-18. In this mode, the intensity or power of the light is directly modulated by the electrical source (data signal), and a photodetector turns this power into another electrical signal. If the input current to the source is $x(t)$, then the output power of the source is proportional to $x(t)$.

Two bad things happen to this launched power as it propagates down the fiber. First, it is attenuated, reducing the signal power at the detector. Second, it suffers dispersion due to chromatic dispersion (and mode dispersion for a multimode fiber), which can be modeled as a linear filtering operation. Let $g(t)$ be the impulse response of the equivalent dispersion filter, including the attenuation, so that the received power at the detector is

$$P(t) = x(t) * g(t). \tag{18.34}$$

In the final conversion to electrical current in the photodetector, the situation is a bit more complicated since quantum effects are important. The incident light consists of discrete photons which are converted to photoelectron-hole pairs in the detector. Hence, the current generated consists of discrete packets of charge generated at discrete points in time. Intuitively we might expect that the arrival times of the charge packets for a Poisson process (Section 3.4) since there is no reason to expect the interarrival times between photons to depend on one other. In fact, this is predicted by quantum theory. Let $h(t)$ be the response of the photodetector circuit to a single photoelectron, and then an outcome for the detected current $y(t)$ is a filtered Poisson process

$$Y(t) = \sum_m h(t - t_m) \, , \tag{18.35}$$

where the $\{t_m\}$ are Poisson arrival times. The Poisson arrivals are characterized by the rate of arrivals, which is naturally proportional to the incident power,

$$\lambda(t) = \frac{\eta}{h\nu} P(t) + \lambda_0 \tag{18.36}$$

where η is the quantum efficiency and λ_0 is a dark current. Note from Campbell's theorem (Section 3.4.4) that the expected detected current is

$$E[Y(t)] = \lambda(t) * h(t) = \frac{\eta}{h\nu} x(t) * g(t) * h(t) + \lambda_0 H(0) \, . \tag{18.37}$$

The equivalent input-output relationship of the channel is therefore characterized, with respect to the mean-value of the detector output current, by the convolution of the two filters — the dispersion of the fiber and the response of the detector circuitry. Of course, there will be statistical fluctuations about this average. This simple linear model for the channel that is quite

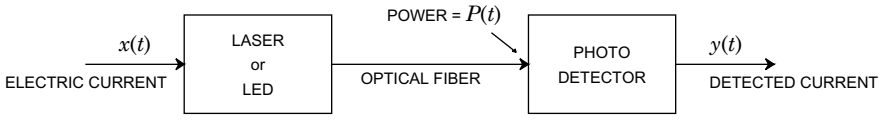

Fig. 18-18. Elements of a direct detection optical fiber system.

accurate unless the launched optical power is high enough to excite nonlinear effects in the fiber and source-detector.

Avalanche Photodetector

In the case of an APD, we have to modify this model by adding to the filtered Poisson process of (18.35) the random multiplier resulting from the avalanche process,

$$y(t) = \sum_m g_m h(t - t_m) , \tag{18.38}$$

the statistics of which has already been considered in Chapter 3. Define the mean and second moment of the avalanche gain,

$$\overline{G} = E[G_m] , \qquad \overline{G^2} = E[G_m{}^2] . \tag{18.39}$$

Then from Section 3.4.6, we know that the effect of the avalanche gain on the second order statistics of (18.35) is to multiply the mean value of the received random process by $\overline{G}$ and the variance by $\overline{G^2}$.

If the avalanche process were deterministic, that is precisely $\overline{G}$ secondary electrons were generated for each primary photoelectron, then the second moment would be the square of the mean,

$$\overline{G^2} = \overline{G}^2. \tag{18.40}$$

The effect of the randomness of the multiplication process is to make the second moment larger, by a factor F_G greater than unity,

$$\overline{G^2} = F_G \overline{G}^2 \tag{18.41}$$

where of course $F_G = \overline{G^2}/\overline{G}^2$. The factor F_G is called the *excess noise factor*. In fact, a detailed analysis of the physics of the APD [13] yields the result

$$F_G = k \cdot \overline{G} + (2 - 1/\overline{G}) \cdot (1 - k) \tag{18.42}$$

where $0 \le k \le 1$ is a parameter under the control of the device designer called the *carrier ionization ratio*. Note that as $k \to 1$, $F_G \to \overline{G}$, or the excess noise factor is approximately equal to the avalanche gain. This says that the randomness gets larger as the avalanche gain gets larger. On the other hand, as $k \to 0$, $F_G \to 2$ for large $\overline{G}$, or the excess noise factor is approximately independent of the avalanche gain. Finally, when $\overline{G} = 1$ (there is no avalanche gain), $F_G = 1$ and there is no excess noise. This is the PIN photodiode detector.

Fiber and Preamplifier Thermal Noise

Any physical system at non-zero temperature will experience noise due to the thermal motion of electrons, and optical fiber is no exception. This noise is often called *thermal noise* or *Johnson noise* in honor of J.B. Johnson, who studied this noise experimentally at Bell Laboratories in 1928. Theoretical study of this noise based on the theory of quantum mechanics was carried out by H. Nyquist at about the same time. Thermal noise is usually approximated as white Gaussian noise. The Gaussian property is a result of the central limit theorem and the fact that thermal noise is composed of the superposition of many independent actions. The

white property cannot of course extend to infinite frequencies since otherwise the total power would be infinite, but rather this noise can be considered as white up to frequencies of 300 GHz or so. Nyquist's result was that thermal noise has an available noise power per Hz of

$$N(\nu) = \frac{h\nu}{e^{h\nu/kT_n} - 1}, \tag{18.43}$$

where h is Planck's constant, ν is the frequency, k is Boltzmann's constant ($1.38 \cdot 10^{-23}$ Joules per degree Kelvin), and T_n is the temperature in degrees Kelvin. By *available noise power* we mean the power delivered into a load with a matched impedance. If we consider this as a two sided spectral density, we have to divide by two.

At frequencies up through the microwave, the exponent in (18.43) is very small, and if we approximate e^x by $1 + x$ we get that the spectrum is approximately white,

$$N(\nu) \approx kT_n . \tag{18.44}$$

This corresponds to a two-sided spectral density of size

$$\frac{N_0}{2} = kT_n/2 . \tag{18.45}$$

However, at high frequencies, this spectrum approaches zero exponentially, yielding a finite total power.

There are two possible sources of thermal noise — at the input to the detector, and in the receiver preamplifier. At the input to the detector only thermal noise at optical frequencies is relevant (the detector will not respond to lower frequencies), and at these frequencies the thermal noise will be negligible.

Example 18-17. At room temperature kT_n is $4 \cdot 10^{-21}$ Joules. At 1 GHz, or microwave frequencies, $h\nu$ is about 10^{-24} Joules, and we are well in the regime where the spectrum is flat. However, at 1 μm wavelength, or $\nu = 3 \cdot 10^{14}$ Hz, $h\nu$ is about $2 \cdot 10^{-19}$ Joules, and $h\nu/kT_n$ is about 50. Thus, the thermal noise is much smaller than kT_n at these frequencies. Generally thermal noise at optical frequencies is negligible in optical fiber systems at wavelengths shorter than about 2 μm [14].

Since the signal level is very low at the output of the detector in Fig. 18-18, we must amplify the signal using a preamplifier as the first stage of a receiver. Thermal noise introduced in the preamplifier is a significant source of noise, and in fact in many optical systems is the dominant noise source. Since the signal at this point is the baseband digital waveform, it occupies a bandwidth extending possibly up to microwave frequencies but not optical frequencies, hence the importance of thermal noise. This thermal noise is the primary reason for considering the use of an APD detector in preference to a PIN photodiode. A more detailed consideration of the design of the preamplifier circuitry is given in [14].

18.3.5. Advanced Techniques

Two exciting developments have been demonstrated in the laboratory: *soliton transmission* and *erbium-doped fiber amplifiers*. The soliton operates on the principle of the *optical Kerr effect*, a nonlinear effect in which the index of refraction of the fiber depends on the optical power. As previously mentioned, in chromatic dispersion, the index of refraction also depends

on the wavelength. Solitons are optical pulses that have a precise shape and peak power chosen so that the Kerr effect produces a chirp (phase modulation) that is just appropriate to cancel the pulse broadening induced by group-velocity dispersion. The result is that all the wavelengths can be made to travel at the same speed, essentially eliminating material dispersion effects. In soliton transmission, material attenuation is the only effect that limits repeater spacing.

An optical amplifier can be constructed out of a fiber doped with the rare-earth element erbium, together with a semiconductor laser pumping source. If the pumping source wavelength is 0.98 or 1.48 μm, then the erbium atoms are excited into a higher state, and reinforce 1.55 μm incident light by stimulated emission. With about 10 mW of pumping power, gains of 30 to 40 dB at 1.55 μm can be obtained. A receiver designed using this optical amplifier is shown in Fig. 18-19. The optical amplifier has gain G, which actually depends on the input signal power because large signals deplete the excited erbium atoms and thereby reduce the gain. The amplifier also generates a spurious noise due to spontaneous emission, and the purpose of the optical bandpass filter is to filter out spontaneous noise outside the signal bandwidth (which depends on source linewidth as well as signal modulation). There is a premium on narrow linewidth sources, because that enables the optical filter bandwidth to be minimized.

The effect of the amplifier is similar to an avalanche detector, in that it increases the signal power (rendering electronic thermal noise insignificant) while adding additional spontaneous noise. The major distinction between the amplifier and avalanche detector, however, is that much of the spontaneous noise in the amplifier can be optically filtered out, whereas in the detector it cannot. It is also possible to place optical amplifiers at intermediate points in a fiber system, increasing the repeater spacing dramatically.

18.4. MICROWAVE RADIO

The term "radio" is used to refer to all electromagnetic transmission through free space at microwave frequencies and below. There are many applications of digital transmission which use this medium, primarily at microwave frequencies, a representative set of which include *point-to-point terrestrial digital radio*, *digital mobile radio*, *digital satellite communication*, and *deep-space digital communication*.

Terrestrial digital radio systems use microwave horn antennas placed on towers to extend the horizon and increase the antenna spacing. This medium has been used in the past principally for analog transmission (using FM and more recently SSB modulation), but in recent years has gradually been converted to digital transmission due to increased demand for data services.

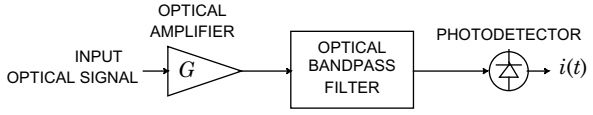

Fig. 18-19. A direct-detection optical receiver using an optical amplifier.

Example 18-18. In North America there are frequency allocations for telephony digital radios centered at frequencies of 2, 4, 6, 8, and 11 GHz [15]. In the United States there were over 10,000 digital radio links in 1986, including a cross-country network at 4 GHz.

A related application is digital mobile radio.

Example 18-19. The frequency band from 806 to 947 MHz is allocated in the United States to land mobile radio services [16]. This band is used for *cellular mobile radio* [17], in which a geographical area is divided into a lattice of cells, each with its own fixed omnidirectional base antenna for transmission and reception. As a vehicle passes through the cells, the associated base antenna is automatically switched to the closest one. An advantage of this concept is that additional mobile telephones can be accommodated by decreasing the size of the cells and adding additional base antennas.

Satellites are used for long-distance communication between two terrestrial antennas, where the satellite usually acts as a non-regenerative repeater. That is, the satellite simply receives a signal from a terrestrial transmitting antenna, amplifies it, and transmits it back toward another terrestrial receiving antenna. Satellite channels offer an excellent alternative to fiber and cable media for transmission over long distances, and particularly over sparse routes where the total communication traffic is small. Satellites also have the powerful characteristic of providing a natural multiple access medium, which is invaluable for random-access communication among a number of users. A limitation on satellites is limited power available for transmission, since the power is derived from solar energy or expendable resources. In addition, the configuration of the launch vehicles usually limit the size of the transmitting and receiving antennas (which are usually one and the same). Most communication satellites are put into synchronous orbits, so that they appear to be stationary over a point on the earth. This greatly simplifies the problems of antenna pointing and satellite availability.

In deep-space communication, the object is to transmit data to and receive data from a platform that is at a great distance from earth. This application includes the features of both satellite and mobile communication, in that the vehicle is usually in motion. As in the satellite case, the size of the antenna and the power resources at the space vehicle are limited.

With the exception of problems of multipath propagation in terrestrial links, the microwave transmission channel is relatively simple. There is an attenuation introduced in the medium due to the spreading of the energy, where this attenuation is frequency-independent, and thermal noise introduced at the antenna and in the amplifiers in the receiver. These aspects of the channel are covered in the following subsections followed by a discussion of multipath distortion.

18.4.1. Microwave Antennas and Transmission

Microwave propagation through free-space is very simple, as there is an attenuation due to the spreading of radiation. The attenuation varies so slowly with frequency that it can be considered virtually fixed within the signal bandwidth. Consider first an *isotropic antenna*; namely, one that radiates power equally in all directions. Assume the total radiated power is P_T $watts$, and assume that at distance d meters from this transmit antenna there is a receive antenna with area A_R meters^2. Then the maximum power that the receive antenna could capture is the transmit power times the ratio of A_R to the area of a sphere with radius d, which is $4\pi d^2$. There are two factors which modify this received power. First, the transmit antenna can

be designed to focus or concentrate its radiated energy in the direction of the receiving antenna. This adds a factor G_T called the *transmit antenna gain* to the received power. The second factor is the *antenna efficiency* η_R of the receive antenna, a number less than (but hopefully close to) unity; the receive antenna does not actually capture all the electromagnetic radiation incident on it. Thus, the received power is

$$P_R = P_T \frac{A_R}{4\pi d^2} G_T \eta_R \,. \tag{18.46}$$

At microwave frequencies, aperture antennas (such as horn or parabolic) are typically used, and for these antennas the achievable antenna gain is

$$G = \frac{4\pi A}{\lambda^2} \eta \,, \tag{18.47}$$

where A is the area of the antenna, λ is the wavelength of transmission, and η is an efficiency factor. Expression (18.47) applies to either a receiving or transmitting antenna, where the appropriate area and efficiency are substituted. This expression is intuitively pleasing, since it says that the antenna gain is proportional to the square of the ratio of antenna dimension to wavelength. Thus, the transmit antenna size in relation to the wavelength is all that counts in the directivity or gain of the antenna. This antenna gain increases with frequency for a given antenna area, and thus higher frequencies have the advantage that a smaller antenna is required for a given antenna gain. The efficiency of an antenna is typically in the range of 50 to 75 percent for a parabolical reflector antenna and as high as 90 percent for a horn antenna [18].

An alternative form of the received power equation can be derived by substituting for the area of the receive antenna in (18.46) in terms of its gain in (18.47),

$$\frac{P_R}{P_T} = G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2 ; \tag{18.48}$$

this is known as the *Friis transmission equation*. The term in brackets is called the *path loss*, while the terms G_T and G_R summarize the effects of the two antennas. While this loss is a function of wavelength, the actual power over the signal bandwidth does not generally vary appreciably where the bandwidth is very small in relation to the center frequency of the modulated signal. The Friis equation does not take into account other possible sources of loss such as rain attenuation and antenna mispointing.

The application of these relations to a particular configuration can determine the received power and the factors contributing to the loss of signal power. This process is known as generating the *link power budget*, as illustrated by the following examples.

Example 18-20. Determine the received power for the link from a synchronous satellite to a terrestrial antenna for the following parameters: Height 40,000 km, satellite transmitted power 2 watts, transmit antenna gain 17 dB, receiving antenna area 10 meters^2 with perfect efficiency, and frequency 11 GHz. The wavelength is $\lambda = c/\nu = 3 \cdot 10^8 / 11 \cdot 10^9 = 27.3$ mm. The receive antenna gain is

$$10\log_{10} G_R = 10\log_{10} \frac{4\pi \cdot 10}{(27.3 \cdot 10^{-3})^2} = 52.2 \,. \tag{18.49}$$

Next, the path loss is

$$10\log_{10}\left(\frac{\lambda}{4\pi d}\right)^2 = 20\log_{10}\left(\frac{27.3\cdot 10^{-3}}{4\pi\cdot 4\cdot 10^{7}}\right) = -205.3 \text{ dB}. \tag{18.50}$$

Finally, we are in a position to calculate the received power, which we express in dBW (decibels relative to one watt) and recognizing that the transmit power is 3 dBW,

$$10\log_{10}\left(\frac{\lambda}{4\pi d}\right)^2 = 3 \text{ dBW} + 17 + 52.3 - 205.3 = -133 \text{ dBW}. \tag{18.51}$$

Example 18-21. The Mariner-10 deep-space mission to Mercury in 1974 used a transmitter power of 16.8 watts and frequency 2.3 GHz. The transmit antenna diameter on the spacecraft was 1.35 meters with efficiency 0.54, which results in an antenna gain of 27.6 dB. The terrestrial receive antenna diameter was 64 meters with efficiency 0.575, for an antenna gain of 61.4 dB. The distance from the spacecraft to ground was $1.6\cdot 10^{11}$ meters, for a path loss of 263.8 dB. Finally, the received power was

$$10\log_{10}P_R = 10\log_{10}16.8 + 27.6 + 61.4 - 263.8 = -162.6 \text{ dBW}. \tag{18.52}$$

The two dominant effects of microwave propagation are attenuation and delay. It is of interest to determine the effect on a passband signal, represented by its complex envelope of Section 2.4.2. Assume the attenuation is A, the distance of propagation is d, and the speed of propagation is c. The delay the signal experiences is $\tau = d/c$, and given a passband signal of the form of Fig. 2-5, the output of the channel is

$$\text{Re}\{Au(t-\tau)e^{j2\pi f_c(t-\tau)}\} = \text{Re}\{Au(t-\tau)e^{-jkd}e^{j2\text{p}f_c t}\}, \tag{18.53}$$

where

$$k = \frac{2\pi f_c \tau}{d} = \frac{2\pi f_c}{c} = \frac{2\pi}{\lambda} \tag{18.54}$$

is called the *propagation constant*. The equivalent complex-baseband channel, shown in Fig. 18-20, characterizes the effect of propagation on the equivalent complex-baseband signal. Not surprisingly, the baseband signal is delayed by τ, the same as the passband signal. In addition, there is a phase shift by $kd = 2\pi d/\lambda$ radians, or 2π radians for each wavelength of propagation distance. The equivalent complex-valued impulse response of the propagation is an impulse with delay τ and area Ae^{-jkd}, and the equivalent transfer function is $Ae^{-jkd}e^{-j2\pi f\tau}$. The

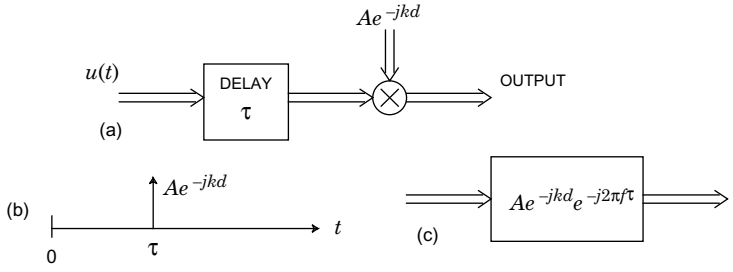

Fig. 18-20. The equivalent complex baseband channel for free-space propagation with attenuation A and distance d. a) Equivalent system, b) the equivalent impulse response, and c) the equivalent baseband transfer function.

only frequency dependence is linear in frequency, due to the delay. For mobile receivers, the effect of small changes in d on the baseband channel response is particularly significant. The effect is dramatically more pronounced for the phase shift than for the delay.

Example 18-22. For a carrier frequency of 1 GHz (typical for mobile radio), the propagation constant is $k = 2\pi f_c / c = 21$ radians/meter. Thus, a $\pi/2$ radian phase shift, which will be very significant to demodulators, occurs with every 7.4 centimeters change in propagation distance. In contrast, the propagation delay changes by 3.3 nanoseconds for each meter change in propagation distance. In relation to typical baseband bandwidths, this is totally insignificant. For example, at 1 MHz, the change in phase shift due to this delay change is only $2\pi f\tau = 2\pi \cdot 0.0033$, or roughly one degree.

18.4.2. Noise in Microwave Amplifiers

On a radio link noise enters the receiver both through the antenna and as internal noise sources in the receiver. We saw in (18.45) that both sources of noise are Gaussian and can be considered as white up through the microwave frequencies. White noise is completely specified by the spectral density $N_0/2$, given by (18.45). However, in radio transmission it is common to express this spectral density in terms of an equivalent parameter, the *noise temperature* expressed in degrees Kelvin. This custom derives from the functional form of (18.45), where N_0 is strictly a function of the temperature. The custom of specifying noise temperature derives from the fact that T_n is reasonable in size, on the order of tens or hundreds of degrees, whereas N_0 is a very small number. Note however that the total thermal noise at some point in a system may be the superposition of many thermal noise sources at different temperatures. Hence, the noise temperature is merely a convenient specification of the noise power, and is not necessarily equal to the physical temperature of any part of the system! For example, if we amplify the noise we increase the noise temperature without affecting the physical temperature of the source that generated that noise.

There are two sources of noise — the noise incident on the antenna and the noise introduced internally in the receiver. The noise incident on the antenna depends on the effective noise temperature in the direction the antenna is pointed. For example, the sun has a much higher effective temperature than the atmosphere. The noise introduced internal to the receiver depends on the design and sophistication of the receiver. It is customary to refer all noise sources to the input to the receiver (the antenna), and define an equivalent noise temperature at that point. Since the stages of a receiver typically have large gains, the noise introduced internal to the receiver usually has a much smaller noise temperature when referred to the receiver input. These receiver noise temperatures range from about four degrees Kelvin for supercooled maser amplifiers to the range of 70 to 200 degrees Kelvin for receivers without physical cooling.

Example 18-23. Continuing Example 18-21 for the Mariner-10 mission the effective noise temperature of the antenna plus receiver was 13.5 degrees Kelvin. A bit rate of 117.6 kb/s was used. What is the signal-to-noise ratio in the receiver assuming the bandwidth of the system is half the bit rate, 58.8 kHz? (We saw in Fig. 5 that this is the minimum possible bandwidth for binary transmission.) The total noise power within the receiver bandwidth would be

$$P_n = kT_nB = 1.38 \cdot 10^{-23} \cdot 13.5 \cdot 58.8 \cdot 10^3 = -169.6 \text{ dBW}. \tag{18.55}$$

The signal-to-noise ratio is therefore

$$SNR = -162.6 + 169.6 = 7.0 \text{ dB}. \tag{18.56}$$

In practice the noise bandwidth will be larger than this, and the SNR will be lower, perhaps by a couple of dB. This SNR will support data transmission, albeit at a rather poor error rate. Coding techniques (Chapters 12 and 13) can compensate for the poor SNR.

The SNR as calculated in this example is a useful quantity since it will not be changed by the large gain introduced in the receiver (both signal and noise will be affected the same way).

18.4.3. Emission Masks

A radio channel does not in itself provide any significant restriction on the bandwidth that we can use for digital communication. Moreover, the free-space channel introduces only a slowly varying dependence of attenuation on frequency (due to antenna gain). Thus, there is nothing inherent about the channel to introduce significant motivation to be spectrally efficient. Enter the regulatory agencies! Since the radio spectrum is a scarce commodity, it is carefully allocated to individual users. Unlike optical-fiber, where different users can install their own fiber, we must all share a single radio environment. To prevent significant interference between users, *spectral emission masks* are specified by regulation. An example of such a mask is shown in Fig. 18-21. In this case, the regulatory agency has assigned a nominal 30 MHz bandwidth centered at f_c to a particular user, but for practical reasons has allowed that user to transmit a small amount of power (down more than 50 dB) outside that band.

This mask is usually adhered to by placing a very sharp cutoff filter in the radio transmitter. Since this filter is imposed by external constraints, it is natural to think of this filter as being part of the channel (this logic is oversimplified of course since the filter requirements depend on the spectrum of the signal feeding the filter). From this perspective, the microwave radio channel has a very sharp cutoff at the band edges, in contrast to the media we have seen earlier which have at most a gradual increase of attenuation with frequency. This characteristic is shared by the voiceband data channel in the next section, and for this reason similar modulation techniques are often employed on the two media.

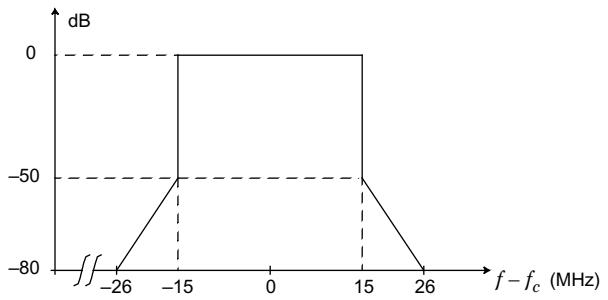

Fig. 18-21. A spectral emission mask referenced to a nominal 30 MHz channel bandwidth. The vertical axis is transmitted power spectrum referenced to the power of an unmodulated carrier. The user signal must stay under the mask. (This mask applies to the United States.)

18.4.4. Multipath Fading

We have seen how the link budget can be determined for a radio link. The calculation we made assumed for the most part idealized circumstances, whereas in practice additional *system margin* must be included in the link budget to account for foreseen or unforeseen circumstances. For example, at higher frequencies there will be an additional *rain attenuation* during rainstorms at the earth receiving antenna. In terrestrial microwave systems, there is an additional important source of attenuation that must be accounted for — *multipath fading* [19]. Both rain attenuation and multipath fading result in an attenuation on the signal path that varies with frequency. A significant difference is that unlike rain attenuation, multipath fading can result in a large frequency-dependent attenuation within the narrow signal bandwidth. This phenomenon is known as *selective fading*.

The mechanism for multipath fading, shown in Fig. 18-22, is very similar to mode distortion in multimode optical fibers and to the distortion introduced by bridged taps in wire-pairs, except that it is time varying. The atmosphere is inhomogeneous to electromagnetic radiation due to spatial variations in temperature, pressure, humidity, and turbulence. This inhomogeneity results in variations in the index of refraction, resulting in possibly two or more ray paths for electromagnetic waves to travel from transmitter to receiver. Another source of multipath is the reflection of radio waves off of obstacles, such as buildings. The effective path lengths may be different for the different rays, and in general will interfere with one another since the receiver will perceive only the sum of the signals.

We can determine the effect of multipath fading on a passband signal using the equivalent complex-baseband response for a single path and applying superposition. If we assume two paths have attenuations A_1 and A_2 and propagation distances d_1 and d_2, corresponding to propagation delays $\tau_1 = d_1/c$ and $\tau_2 = d_2/c$, we can define two parameters $\Delta d = d_1 - d_2$ and $\Delta\tau = \tau_1 - \tau_2$. Then by superposition the equivalent complex-baseband channel transfer function is

$$A_1 e^{-j2\pi f\tau_1} e^{-jkd_1} + A_2 e^{-j2\pi f\tau_2} e^{-jkd_2} = A_1 e^{-j2\pi f\tau_1} e^{-jkd_1}\left(1 + \frac{A_2}{A_1} e^{j2\pi f\Delta\tau} e^{jk\Delta d}\right). \quad (18.57)$$

The first terms have a constant and linear phase shift due to the delay τ_1, identical to the first path. The term in parentheses is important, because it can display a complicated dependence on frequency due to constructive and destructive interference of the two signals at the receiver.

The critically important parameter is $\Delta\tau$, which is called the *delay spread*. Two distinct cases can be distinguished. The first occurs when, for baseband frequencies of interest, $|f\Delta\tau| << 1$, so that the frequency-dependence of the second term is insignificant. This is called the *narrowband model*. For this case, the two path propagation is similar to a single path, in that

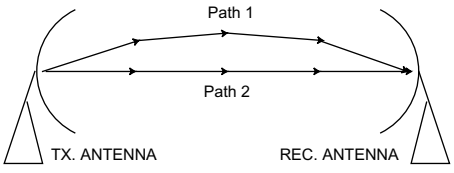

Fig. 18-22. Illustration of two ray paths between a transmit and receive radio antenna. Fading attenuation results when the two paths have different propagation delays.

it results in a delay (linear phase shift with frequency) plus a constant phase shift. The contrary case is called the *broadband model*, and results in a more complicated frequency dependence due to constructive and destructive interference.

Example 18-24. Assume that we define the transition between the narrowband and broadband model as a delay spread such that $|f\Delta\tau| = 0.005$ at the highest baseband frequency of interest. Equivalently, we expect that $f = 1/200\Delta\tau$ for the highest frequency. Then if the delay spread is 1 ns, baseband channels with a bandwidth less than 5 MHz are considered narrowband, and bandwidths greater than 5 MHz (especially those significantly greater) are considered broadband. If the delay spread increases to 100 ns, then the narrowband channel has bandwidth less than 50 kHz according to this criterion. Note that all that counts is the delay spread, and not the absolute delay nor the carrier frequency. Also note that the actual passband signal has a bandwidth double the equivalent complex baseband signal.

Example 18-25. For the two-path case, the magnitude-squared of the frequency response for the frequency-dependent term of interest is

$$|1 + \rho e^{j2\pi f\Delta\tau}|^2 = 1 + |\rho|^2 + 2 \cdot \mathrm{Re}\{\rho e^{j2\pi f\Delta\tau}\} \tag{18.58}$$

for some complex constant ρ. We will chose a delay spread of 10 nanoseconds (a typical worst-case number in an urban environment) and a fairly large $|\rho| = 0.99$. This is plotted in dB in Fig. 18-23 over a ± 50 MHz frequency range, a broadband model, and a narrower frequency range, a narrowband model. Note the large notches due to destructive interference at some frequencies, accentuated by the fact that the two paths are nearly the same amplitude. Also note the close to 6 dB gain at some frequencies due to constructive interference. The narrowband model is plotted over a ± 500 kHz frequency range, which by the criterion of Example 18-24 is a narrowband model. Note that the channel response varies only a couple of dB over this range.

The two-path model, which is usually adequate for fixed terrestrial microwave systems, suggests that fading may result in either a monotonic gain change (or slope) across the channel or as a dip (or notch) in the channel response within the bandwidth. A typical faded channel response is shown in Fig. 18-24, and the typical parameters that characterize the fade are identified [15].

In Section 18.4.1, we showed that the power loss in free-space radio propagation obeys a square-law relationship; that is, the receive power decreases as d^{-2}, or the path loss in dB increases as $20 \cdot \log_{10} d$. For terrestrial microwave transmission, the path loss increases more rapidly than in free space, typically more like d^{-4} or $40 \cdot \log_{10} d$ in dB. This can be explained

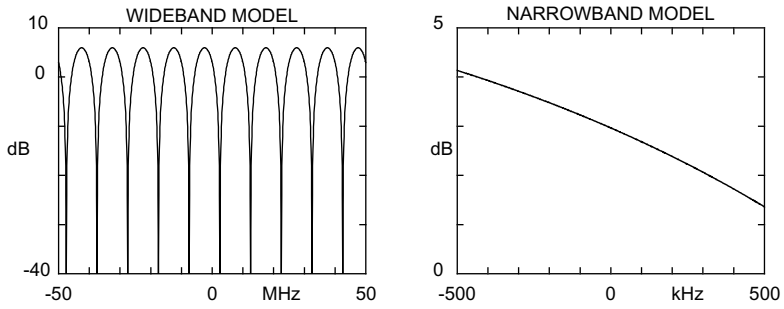

Fig. 18-23. Complex baseband channel amplitude response over a wide frequency range and a narrow frequency range for a two-path model with $\rho = 0.99j$.

using the simple model of Fig. 18-25. Even for highly directional antennas, for a large d there will be a substantial reflection off the ground interfering at the receive antenna. Typically the ground is close to a short circuit for oblique angles of incidence at microwave frequencies, implying a reflection coefficient near –1 (so that the net incident and reflected electric fields sum to zero).

Exercise 18-3. For the geometry of Fig. 18-25, consider only the reflection resulting if the ground acts like a perfect mirror, and that both the direct and indirect paths suffer a free space loss. Assuming the distance between antennas is much higher than the antenna heights, show that the resulting net power loss is approximately

$$\frac{P_R}{P_T} = \left(\frac{P_R}{P_T}\right)_{\text{freespace}} \left(\frac{4\pi h_t h_r}{\lambda d}\right)^2 . \tag{18.59}$$

Hence, the effect of the reflection is a destructive interference that increases the path loss by another factor of d^{-2} over and above the free space loss.

Note that, not unexpectedly, it is advantageous to have high antennas (the loss decreases as the square of the antenna heights).

Even when the transmitter and receiver are at fixed locations relative to one another, fading is a time-varying phenomenon for large distances (30 km or greater) due to atmospheric phenomena. Of considerable importance to designers of radio systems is not only the depth but also the duration of fades. Fortunately, it has been observed that the deeper the fade, the less

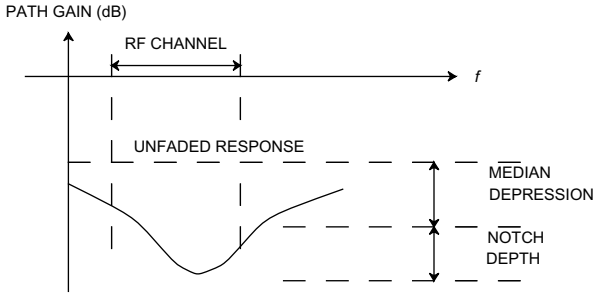

Fig. 18-24. A typical frequency-selective notch due to fading with some terminology. Note that the impact on the channel depends strongly on the location of the notch relative to the channel bandwidth.

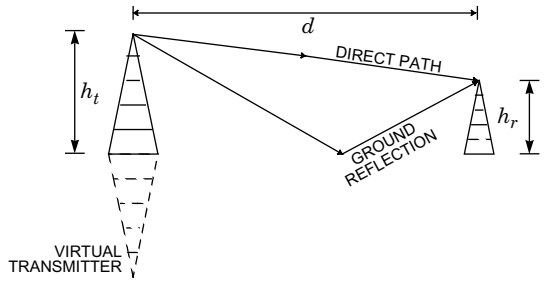

Fig. 18-25. The attenuation of a terrestrial microwave system is increased by the ground reflection. There will be ground reflections from all points between the two antennas, but the single reflection resulting if the ground acts like a perfect mirror is shown. The transmit antenna height is h_t, the receive antenna height is h_r, and the distance between antennas is d.

frequently it occurs and the shorter its duration when it does occur. Also, the severity of fades increases as the distance between antennas increases or as the carrier frequency increases. Fading can also be mitigated by using *diversity techniques*, in which two or more independent channels are somehow combined [20]. The philosophy here is that only one of these channels at a time is likely to be affected by fading.

18.4.5. Mobile Radio

One of the most appealing uses for radio transmission is for communication with people or vehicles on the move. For this type of communication there is really no alternative to radio transmission, except for infrared, which does not work well outdoors. Mobile radio exhibits some characteristics that are different from point-to-point transmission. First, antennas must generally be omnidirectional, and thus they exhibit much less antenna gain. Second, there can be obstacles to direct propagation, causing a *shadowing effect* that results in large variations in received signal power with location. Third, the most common application is in urban areas, where there are many opportunities for multiple reflections, and the two-path model is usually not accurate. Fourth, the user is often moving, resulting in extreme time-variations in transmission conditions over even short distances, as well as Doppler shift in the carrier frequency.

The two-path model is easily extended to an M-path model, again using superposition. In this case, the complex-baseband output of the channel is

$$\sum_{i=1}^{M} A_i u(t-\tau_i) e^{-jkd_i}, \tag{18.60}$$

where the A_i are real-valued attenuation coefficients, d_i is the length of the i-th path, and τ_i is the propagation delay of the i-th path. There may be a dominant path whose attenuation coefficient obeys the fourth-power law with distance, but the other coefficients depend on the reflection coefficients of indirect paths and hence bear a complicated relationship to position. Furthermore, due to shadow effects, there may even be no dominant path. for example if the mobile receiver is located behind a building; the radio waves will suffer a diffraction loss. This shadowing loss typically varies markedly over a distance of tens to hundreds of meters. If we average the received power over an area on the order of 1 km^2, we will see the fourth-power loss with distance, but if we average over an area on the order of 1 meter^2 we will see an additional fluctuation with position due to shadowing. Shadowing is often assumed to result in a log-normal distribution in local-average received power; that is, the power expressed in dB has a Gaussian distribution. The standard deviation of the power expressed in dB is roughly 4 dB for typical urban areas.

When we examine local received power, not averaged over an area, we begin to see wild fluctuations due to multipath fading. For a moving vehicle, fades of 40 dB and more below the local-average level are frequent, with successive minima occurring every half wavelength or so (a fraction of a meter at microwave frequencies). Thus, the motion of the vehicle introduces a whole new dimension to the fading experienced on a point-to-point system, where the fluctuations are much slower. This rapid fluctuation is known as *Rayleigh fading* because the distribution of the envelope of the received carrier often obeys a Rayleigh distribution [21].

To understand Rayleigh fading, we must examine the effect of vehicle motion, which results in a time variation in received carrier phase. As before, this can be understood by considering a single path, and then applying superposition to multiple paths. The geometry of a single path is shown in Fig. 18-26, including a reflection between the transmitter and receiver. As shown, a virtual transmitter can be defined behind the reflector with a linear propagation to the receiver. Let $\boldsymbol{d}$ be a vector from virtual transmitter to receiver at time $t = 0$, let $\boldsymbol{v}$ be the velocity vector for the vehicle at time $t = 0$, and let θ be the angle between $\boldsymbol{d}$ and $\boldsymbol{v}$, or the angle of incidence of the propagation path relative to the vehicle velocity. Let the scalar initial distance and velocity be $d = \| \boldsymbol{d} \|$ and $v = \| \boldsymbol{v} \|$. The vector from transmitter to receiver is $\boldsymbol{d} + \boldsymbol{v}t$, and the propagation distance as a function of time is

$$\| \boldsymbol{d} + \boldsymbol{v}t \| = (d^2 + v^2t^2 + 2\langle \boldsymbol{d}, \boldsymbol{v} \rangle t)^{1/2}, \tag{18.61}$$

where the inner product is $\langle \boldsymbol{d}, \boldsymbol{v} \rangle = dv \cdot \cos\theta$. This distance is not changing linearly with time, but can be approximated by a linear function of time.

Exercise 18-4. Show that if $t << d / v$, then (18.61) can be approximated accurately by $d + vt\cos\theta$. For example, if $d = 1$ km and $v = 30$ m/sec (approximately 100 km/hr) then the approximation holds for $t << 66$ sec.

The time scale over which the linear approximation to distance is valid is quite large relative to the significant carrier phase fluctuations, and hence it is safe to assume that the distance to the receiver is changing as $v \cdot \cos\theta \cdot t$. This change in distance has slope $+v$ when the receiver is moving directly away from the transmitter, $-v$ when it is moving directly toward the transmitter, and zero when the receiver is moving orthogonally to the transmitter.

With this basic geometric result in hand, the received complex-baseband signal is

$$A \cdot \mathrm{Re}\left\{ u\left(t - \frac{d}{c} - \frac{vt}{c}\cos\theta \right) e^{-jkd} e^{-jkvt\cos\theta} e^{j2\pi f_c t} \right\}. \tag{18.62}$$

We see here several propagation effects. First, the baseband signal $u(t)$ is delayed by a time-varying amount, due to the changing propagation distance. This effect is generally insignificant at the baseband frequencies of interest. Second, there is a static phase shift e^{-jkd} due to the propagation distance at $t = 0$. Third, and most interesting, is a phase shift that is linear with time. In effect, this is a frequency offset, known as the *Doppler shift*. The carrier frequency is shifted from f_c to $f_c - f_d$, where the Doppler frequency is

$$f_d = \frac{k}{2\pi} v\cos\theta = \frac{v}{\lambda}\cos\theta \,. \tag{18.63}$$

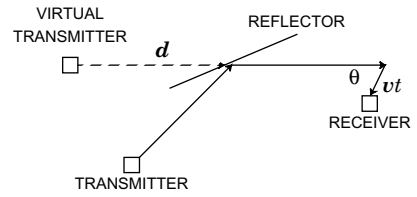

Fig. 18-26. Trajectory of motion for a vehicle moving at constant velocity, relative to a propagation path including a reflection.

When the receiver is moving away from the transmitter, the Doppler shift is negative; it is positive when the receiver is moving toward the transmitter.

Example 18-26. If the vehicle velocity is $v = 30$ m/sec (100 km/hr), and the carrier frequency is 1 GHz ($\lambda = 0.3$ meters), then the maximum Doppler shift is $f_d = v/\lambda = 100$ Hz. This illustrates that relative to the carrier frequency, the Doppler shift is typically quite small, but relative to baseband frequencies it can be relatively large. Also observe that for a constant vehicle velocity, the Doppler shift becomes larger as the carrier frequency increases.

In addition to affecting the propagation distance and angle of incidence, the reflection in Fig. 18-26 will also affect the attenuation constant and add an unknown phase shift due to the reflection coefficient.

The Doppler shift by itself might not be a big issue, since it results in an offset in a carrier frequency that might not be too precisely known in the first place. The more substantive effect occurs when there are two or more paths, each with different Doppler shifts because their incident angles at the receiver are different. If the delay spread of the different paths is small, we can assume a narrowband model; that is, the different delays of the arriving replicas of the baseband signal $u(t)$ are insignificant for the baseband frequencies of interest. The resulting superposition of different Doppler shifts can result in a rapidly fluctuating phase and amplitude. For example, for a set of paths with amplitude A_i, delays $\tau_i = \tau$ assumed to be the same on all paths (which is the narrowband model), phase shifts ϕ_i at time zero, maximum Doppler shift f_d, and angles of incidence θ_i, the receive complex baseband signal is

$$\mathrm{Re}\left\{\sum_i A_i u(t-\tau_i) e^{j(\phi_i - v_d t \cos\theta_i)}\right\} = \mathrm{Re}\{u(t-\tau) r(t)\}, \tag{18.64}$$

where

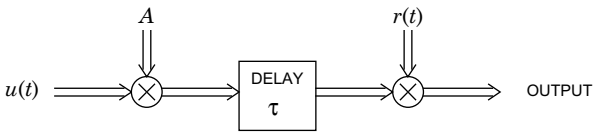

Fig. 18-27. A model for the baseband channel with a receiver in motion at uniform velocity.

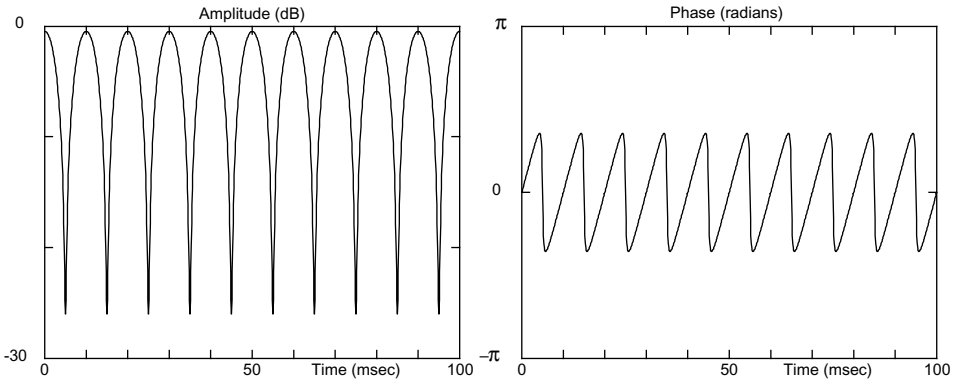

Fig. 18-28. Amplitude and phase of $r(t)$ resulting from the superposition of two paths with Doppler shift of 0 and 100 Hz, with $A_1 = 1$ and $A_2 = 0.9$.

$$r(t) = \sum_i A_i e^{j(\phi_i - v_d t \cos\theta_i)} . \tag{18.65}$$

The basic equivalent baseband channel model is shown in Fig. 18-27. The effect is multiplication by a complex-valued waveform $r(t)$. It is instructive to plot this waveform for a couple of cases. For example, we show in Fig. 18-28 the effect of adding two carriers with a relative 100 Hz Doppler shift. The result is fades at 10 msec intervals, the 10 msec being the reciprocal of the relative Doppler shift. There are periodic very rapid phase jumps, corresponding precisely to the times at which there are large amplitude fades. This effect is explained by Fig. 18-29, which shows a polar plot. The waveform $r(t)$ follows the circular trajectory shown, where the angular velocity is constant. The amplitude fades occur when the trajectory comes near the origin, which coincides with time that the phase changes most rapidly.

These plots are repeated for 40 signals arriving from uniformly-spaced directions in Fig. 18-30 and Fig. 18-31. While the result is qualitatively similar, the trajectory is much more complicated and random-looking. Again there are occasional deep amplitude fades, which coincide with rapid phase variations. The time scale of these deep fades is again on the order of 10 msec, which is the reciprocal of the maximum Doppler frequency. This also corresponds to the time the receiver traverses a half of a wavelength at the carrier frequency.

The very chaotic change in amplitude and phase with time shown in Fig. 18-30 can be characterized statistically employing the central limit theorem. Returning to the model of Fig. 18-27, we can model the multiplicative signal $r(t)$ as a random process $R(t)$. Examining this process at some point in time t_0, the phase of the i-th incident path is given by $\xi_i = \phi_i - vd\cos(\theta_i)t_0$. Since the phases ϕ_i are very sensitive functions of the initial position, it is reasonable to assume that the ξ_i are i.i.d. uniform random variables on the interval $[0, 2\pi)$. Writing the real and imaginary parts independently,

$$\mathrm{Re}\{R(t_0)\} = \sum_i A_i \cos\xi_i , \qquad \mathrm{Im}\{R(t_0)\} = \sum_i A_i \sin\xi_i , \tag{18.66}$$

where each term is the sum of independent random variables. By the central limit theorem, as the number of terms increases both $\mathrm{Re}\{R(t_0)\}$ and $\mathrm{Im}\{R(t_0)\}$ will be Gaussian distributed, and hence $R(t)$ will be a complex-valued Gaussian random variable.

Exercise 18-5. Show that, with the assumption that the ξ_i are i.i.d. uniform random variables,

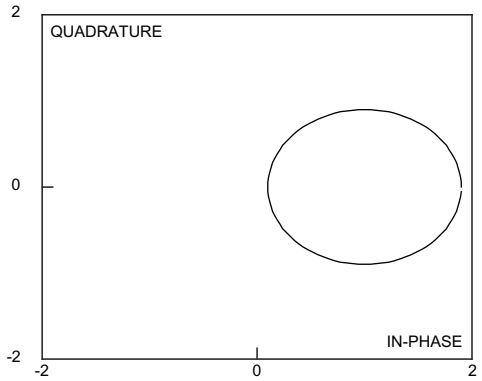

Fig. 18-29. Polar plot of the trajectory of $r(t)$ for the same case as Fig. 18-28.

$$\mathrm{E}[(\mathrm{Re}\{R(t_0)\})^2] = \mathrm{E}[(\mathrm{Im}\{R(t_0)\})^2] = \sigma^2 = \tfrac{1}{2}\sum_i A_i^2 \ , \tag{18.67}$$

$$\mathrm{E}[\mathrm{Re}\{R(t_0)\}\mathrm{Im}\{R(t_0)\}] = 0 \ . \tag{18.68}$$

When

$$R(t_0) = Re^{j\Theta} \tag{18.69}$$

is a complex-valued Gaussian random variable with identically-distributed and independent real and imaginary parts, then R is a Rayleigh-distributed random variable,

$$f_R(r) = \begin{cases} \dfrac{r}{\sigma^2} e^{-r^2/(2\sigma^2)}, & r \geq 0 \\ 0, & r < 0 \end{cases} , \tag{18.70}$$

and the phase Θ will be uniformly distributed on $[0, 2\pi)$. The amplitude R is the envelope of the received carrier, and Θ is the phase, so we can say that the envelope has a Rayleigh distribution and the phase is uniform.

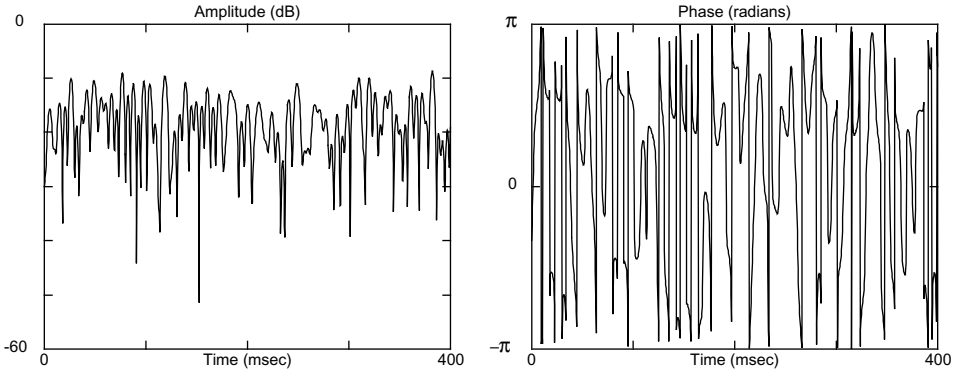

Fig. 18-30. a) The amplitude and b) phase for the superposition of 40 signals arriving at uniform angles, each with the same amplitude and random phases.

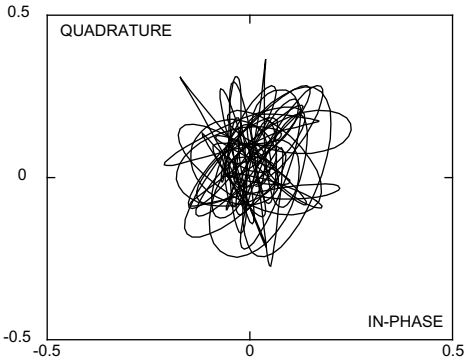

Fig. 18-31. Polar plot for the same case as Fig. 18-30.

The conclusion is that when a CW carrier is transmitted, the received signal $R(t)$ is well approximated as a complex Gaussian process. The power spectrum of that process can be calculated, if we make assumptions about the distribution of arriving power vs. angle. This is because the frequency of an arriving component depends directly on the cosine of the angle of arrival. Let $R(t)$ have power spectrum $S_R(f)$. The contribution to $R(t)$ arriving at angle θ is at frequency $f_c + f_d \cos\theta$. This implies that $S_R(f)$ is confined to frequency band $[f_c - f_d\,, f_c + f_d]$. In particular, the total power arriving in band $[f_0, f_c + f_d]$ corresponds to angle of arrivals in the range

$$f_c + f_d \cos\theta \geq f_0, \text{ or } |\theta| \leq \theta_0 = \cos^{-1}\left(\frac{f_0 - f_c}{f_d}\right). \tag{18.71}$$

If we assume, for example, that a total received power P is arriving uniformly spread over all angles $|\theta| \leq \pi$, then the portion of the power arriving in band $[f_0, f_c + f_d]$ must be $P \cdot \theta_0 / \pi$. Thus,

$$\int_{f_0}^{f_c + f_d} S_R(f)df = \frac{P}{\pi}\cos^{-1}\left(\frac{f - f_c}{f_d}\right), \tag{18.72}$$

and differentiating both sides with respect to ω_0, the power spectrum is

$$S_R(f) = \frac{P/\pi}{\sqrt{f_d^2 - (f - f_c)^2}}, \qquad |f - f_c| \leq f_d, \tag{18.73}$$

and zero elsewhere (of course the spectrum is symmetric about $f = 0$). This power spectrum is plotted for positive frequencies in Fig. 18-32, where we see that the power is concentrated in the region of frequencies $f_c \pm f_d$. A sample function of a random process with this power spectrum will look like a random version of the deterministic signal $\cos(2\pi f_c t)\cdot\cos(2\pi f_d t)$, since the latter has a Fourier transform that consists of delta functions at $f_c \pm f_d$. This AM-DSB signal is the carrier multiplied by an envelope with periodic zero crossings (fades) spaced at $\lambda/(2v)$ sec intervals. This is just the time it takes for the vehicle to travel a half wavelength. Thus, temporally, Rayleigh fading exhibits a strong tendency toward fades every half wavelength when the power is uniformly spread over all incoming angles.

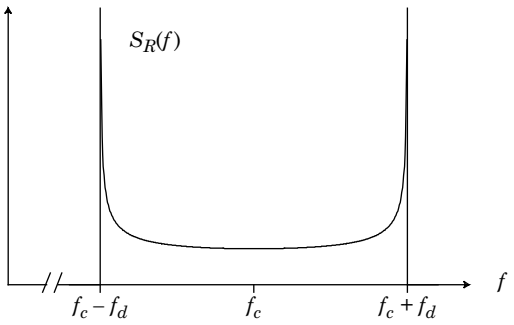

Fig. 18-32. The Doppler power spectrum of the received carrier for a vehicle traveling at velocity v, assuming the received signal power is spread uniformly over all angles of arrival.

Example 18-27. If the vehicle velocity is 100 km/hr and the carrier frequency is 1 GHz, the maximum Doppler frequency is approximately 100 Hz. This means that the individual paths coming into the receiver can have Doppler shifts on the order of ±100 Hz, or the bandwidth of the passband signal is increased by approximately 200 Hz due to the motion of the vehicle. The wavelength is about 0.3 meters, so that the time it takes the vehicle to travel a half wavelength is

$$t = \frac{0.15 \text{ meters}}{30 \text{ meters/sec}} = 5 \text{ msec} . \tag{18.74}$$

We can expect significant fades approximately every 5 msec, which happens to be the reciprocal of the 200 Hz range of Doppler shifts.

The model of Fig. 18-27 and the Rayleigh fading derivation assumed a narrowband model; that is, the delay spread is small with respect to the reciprocal of the bandwidth, or equivalently that delays τ_i in (18.64) are identical over all paths. Thus, the model must be modified to accommodate a wideband model, when the signal bandwidth is too large. Usually this is handled as follows. First, any given reflection, like off a high-rise building, is actually a complicated superposition of multiple reflections, where the delay spread across these reflections is small enough to obey the narrowband model. Thus, this single reflection can actually be represented by a narrowband Rayleigh fading model with an associated delay τ_1. Now if there is a second reflection with a significantly different delay, it can be represented by another narrowband Rayleigh fading model with delay $\tau_2 \neq \tau_1$. The broadband model follows from superposition of these narrowband models.

A two-path broadband model is illustrated in Fig. 18-33. The complex-baseband signal $u(t)$ experiences the two path delays τ_1 and τ_2, and the two delay outputs are multiplied by independent complex-Gaussian processes $r_1(t)$ and $r_2(t)$. Each path also has an associated attenuation A_i, and a static phase shift which can be subsumed in $r_1(t)$ and $r_2(t)$. This broadband model is easily generalized to an arbitrary number of paths.

18.5. TELEPHONE CHANNELS

Most locations in the world can be reached over the public telephone network, so the voiceband channel is an almost universal vehicle for data communication. The design of digital modems for telephone channels is challenging, because the channel was designed primarily for voice, and impairments that are not serious for voice can be debilitating for data signals. The telephone channel is a prime example of a composite channel, consisting of many media such

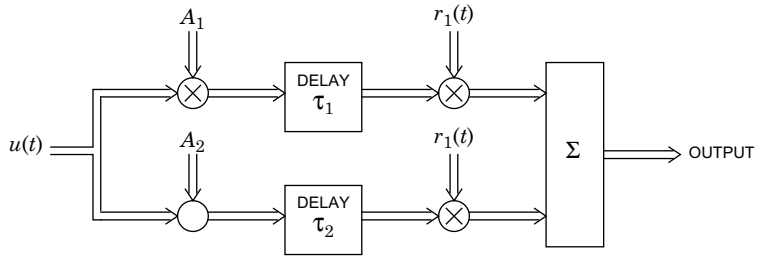

Fig. 18-33. A broadband two-path model, where each path is assumed to be independently Rayleigh fading.

as wire pairs, satellite channels, coaxial cables, terrestrial microwave links, and optical fiber. Even more important than the media are the many modulation systems built on top of these media, such as pulse-code modulation and single-sideband modulation. The characteristics of the channel vary widely depending on the particular connection. It is useful to discuss these characteristics, not only because of the importance of this particular channel, but also because we encounter many impairments that occur in other situations as well.

18.5.1. Measured Performance of Telephone Channels

Because of the wide variety of possible connections, there is no simple analytical characterization of the telephone channel. Modem designers rely rather on statistical surveys of telephone circuits. In the U.S., a comprehensive survey was conducted in 1969-70 [22] and again in 1982-83 [23]. The data in this section comes primarily from interpretation of the second survey. A modem designer needs to determine the acceptable percentage of telephone connections over which the modem will perform, and then find the parameter thresholds that are met or exceeded by that percentage of channels. The resulting thresholds can be quite sensitive to the percentage.

Example 18-28. According to the 1982-83 connection survey, 99% of end-to-end channels attenuate a 1004 Hz tone 27 dB or less. But 99.9% of channels attenuate the same tone 40 dB or less. To get the extra 0.9% coverage, an additional 13 dB of loss must be tolerated.

In Table 18-1 we give typical worst-case figures assumed for some of the impairments on the channel. The percentage of telephone channels that exceed this performance is roughly 99%. Linear distortion is a major impairment that is missing from the table because it is difficult to summarize concisely. It is discussed below, followed by discussions of the remaining impairments.

Linear Distortion

The frequency response of a telephone channel can be approximated by a linear transfer function $B(f)$, roughly a bandpass filter from 300 to 3300 Hz. This bandwidth is chosen to give acceptable voice quality in the network, and is enforced by bandpass filters in analog and digital modulation systems used in the network. A typical transfer function of a telephone channel is illustrated in Fig. 18-34, using traditional terminology that we now will explain.

Table 18-1. Typical worst-case impairments for telephone channels. Roughly 99% of the telephone circuits measured in the 1982-83 connection survey [23] meet or exceed this performance.

Impairment	Impairment
Attenuation of a 1004 Hz tone	27 dB
Signal to C-notched noise ratio	20 dB
Signal to second harmonic distortion ratio	34 dB
Signal to third harmonic distortion ratio	33 dB
Frequency offset	3 Hz
Peak to peak phase jitter (2–300 Hz)	20˚
Peak to peak phase jitter (20–300 Hz)	13˚
Impulse noise (–4 dB threshold)	4 per minute
Phase hits (20˚ threshold)	1 per minute
Round trip delay (no satellites)	50 ms

Amplitude distortion, the magnitude of the frequency response, is plotted as attenuation (or loss) vs. frequency. Amplitude distortion is often summarized as a set of *slope distortion* numbers, which attempt to capture images such as Fig. 18-34a. A typical slope distortion measure is the worst of two differences, (1) the loss at 404 Hz minus the loss at 1004 Hz and (2) the loss at 2804 Hz minus the loss at 1004 Hz. For 99% of telephone connections, that number is less than 9 dB. Several other slope distortion characterizations are found in the literature, but they are difficult to use in practice. We refer interested readers to the connection survey [23].

Interestingly, the attenuation in Fig. 18-34a is almost precisely the typical attenuation of the local loop from the 1980 survey [24] combined with the typical frequency response of the filters in the PCM modulators in Fig. 18-3, suggesting that these are the dominant sources of frequency-dependent attenuation.

Phase distortion, the deviation from linear of the phase response of $B(f)$, is traditionally described as *envelope delay distortion*. *Envelope delay* is defined as the negative of the derivative of the phase of the received signal with respect to frequency, and hence measures the deviation from a linear phase response. Envelope delay distortion is often summarized by a set of numbers, much as the magnitude response is summarized by slope distortion. For details see [24].

The overall attenuation of the channel is typically measured at 1004 Hz, where the attenuation is usually near its minimum. The attenuation is usually about 6 dB between one local switch and another, to which is added the loss of the local loops at each end.

Noise Sources

In addition to attenuation, there is noise present on the voiceband channel, primarily from four sources: *quantization noise*, *thermal noise*, *crosstalk*, and *impulse noise*. We discuss them in order of increasing importance.

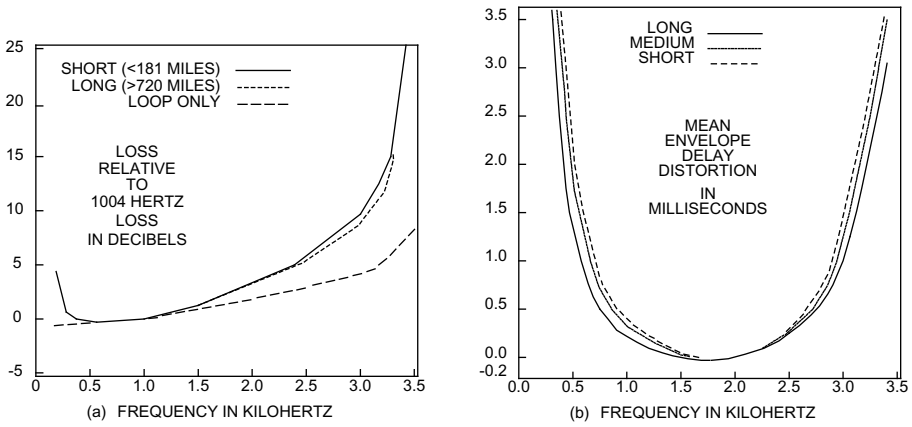

Fig. 18-34. The attenuation (a) and envelope delay distortion (b) of a typical telephone channel as a function of frequency. The attenuation is given relative to the attenuation of a 1004 Hz tone, and the envelope delay distortion relative to 1704 Hz, where it is near its minimum value [23].

Crosstalk of the type discussed in Section 18.2 is one impairment that is more severe for voice than for data, so it has largely been eliminated from the network. Induction of interfering tones at 60 Hz and its harmonics (50 Hz in Europe) from power lines is more significant. As on other communication channels, thermal noise is an important impairment. *Impulse noise* consists of sudden, large spikes of short duration and is measured by counting the number of times the noise exceeds a given threshold. Impulse noise is due to electromechanical switches in the network, such as in telephone switches and dial telephones. Impulse noise is not well characterized, and modem designs are not heavily influenced by its presence.

The dominant source of noise is *quantization error* introduced by PCM systems, as in Fig. 18-3. Quantization error is a consequence of using a limited number of bits to represent each sample in the PCM system. While the quantization error is deterministically dependent on the signal, the randomness of the signal usually gives quantization error a "noise-like" characteristic. It has an approximately white power spectrum, and the level of noise is usually measured by the *signal-to-quantization-noise ratio (SQNR)*. The SQNR for a single quantizer as encountered in the U.S telephone network is illustrated in Fig. 18-35. Note that over an input range of about 30 dB (-40 to -10 dBm0) the SQNR varies by only about 6 dB (33 to 39 dB). This relatively constant SQNR implies that the quantization error power varies almost in direct proportion to the signal power; that is, it is not constant independent of the signal as for thermal noise. A thorough study of this noise is given in [25]. For the fastest voiceband data modems, it can be the dominant impairment. For lower speed modems, it is adequately approximated by white Gaussian noise.

In Table 18-1, the noise is labeled *C-notched noise*, which refers to a particular filter applied to the noise prior to measurement of power. This filter is chosen on the basis of subjective effects for voice, and if the noise is white has no effect beyond a fixed offset in the measured power. Quantization error power is measured by applying a *holding tone*, usually at 1004 Hz, and filtering out this tone with a deep (–50 dB) notch filter prior to the measurement of the remaining quantization error with a C-message weighted filter.

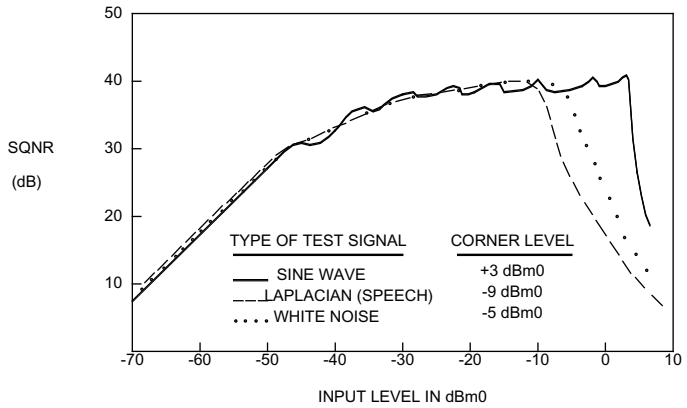

Fig. 18-35. SQNR as a function of the absolute input signal power for three different types of inputs. The Gaussian and Laplacian inputs are random, with the latter approximating the p.d.f. of speech samples [23].

Nonlinear Distortion

Nonlinear distortion is due to imperfections in amplifiers and also to tracking errors between A/D and D/A converters. Because of its relatively low level, non-linear distortion is a significant impairment only for the most elaborate, highest data-rate modems.

Frequency Offset

Frequency offset is peculiar to telephone channels and channels with Doppler shift. If the input to the channel is $x(t)$, with Fourier transform $X(f)$, and the channel has a frequency offset of ω_0 radians, and no other impairments, then the output of the channel has Fourier transform

$$Y(f) = \begin{cases} X(f - f_0), & \text{for } f > 0 \\ X(f + f_0), & \text{for } f < 0 \end{cases} . \tag{18.75}$$

This small shift in the spectrum of signal has important implications for carrier recovery (Chapter 15) and echo cancellation (Chapter 20).

Exercise 18-6. We saw in Chapter 2 that passband data signals can be expressed in the form

$$x(t) = \sqrt{2}\,\mathrm{Re}\{s(t)e^{j2\pi f_c t}\} , \tag{18.76}$$

where the complex envelope $s(t)$ is a complex-valued baseband data signal and f_c is the carrier frequency. Show that the effect of a frequency offset on the channel is a received signal

$$y(t) = \sqrt{2}\,\mathrm{Re}\{s(t)e^{j2\pi(f_c - f_0)t}\} . \tag{18.77}$$

In effect, the carrier frequency has been shifted by f_0. Assume that $f_0 > 0$ and $s(t)$ is bandlimited so that $S(f) = 0$ for $|f| > f_c$.

Frequency offset is a consequence of using slightly different frequencies to modulate and demodulate single-sideband (SSB) signals in analog transmission facilities (Fig. 18-1). It is allowed because it has no perceptible effect on speech quality, and can be compensated by proper design in voiceband data modems.

Phase Jitter

Phase jitter on telephone channels is primarily a consequence of the sensitivity of oscillators used for carrier generation in SSB systems (Fig. 18-1) to fluctuations in power supply voltages. Since power supply fluctuations are often at 60 Hz or harmonics thereof, the largest components of phase jitter are often at these frequencies. Phase jitter is measured by observing the deviation of the zero crossings of a 1004 Hz tone from their nominal position in time.

Phase jitter can be viewed as a generalization of frequency offset. If the phase jitter on a channel is $\theta(t)$, the effect on the transmitted signal of (18.76) is a received signal of the form

$$y(t) = \mathrm{Re}\{s(t)e^{j(2\pi f_c t + \theta(t))}\} . \tag{18.78}$$

A phase jitter of $\theta(t) = 2\pi f_0 t$ amounts to frequency offset. It is common for $\theta(t)$ to have oscillatory components at the power line frequency (50 or 60 Hz) and harmonics. If we simply demodulate this signal using the carrier $e^{j2\pi f_c t}$, we recover a distorted baseband signal $s(t)e^{j\theta(t)}$

rather than the desired $s(t)$. To mitigate this distortion, it is common in carrier recovery (Chapter 15) to include algorithms designed to track and remove this undesired phase jitter.

A phase hit is an abrupt change in the nominal phase of a received sinusoidal signal lasting at least 4 ms. There is little that can be done to defend a modem against this degradation, but it must be taken into account in the design of the carrier recovery (Chapter 15).

Delay and Echo

Delay and *echo* are the final impairments in telephone channels that we will consider. A simplified telephone channel is shown in Fig. 18-36. The *local loop*, which is the twisted wire pair connecting the central office with customer premise, is used for transmission in both directions. Both signals share the same wire pair. At the central office, a circuit called a *hybrid* separates the two directions of transmission. Longer distance facilities are *four-wire*, meaning that the two directions of transmission are physically separated.

One possible implementation of the hybrid circuit is shown in Fig. 18-37. The signal from the other end of the two-wire facility is fed through to the receive port. The transmit signal appears at the transformer as a voltage divider with impedances R and Z_0, where the latter is the input impedance of the two-wire facility. We cancel this undesired feedthrough by constructing another voltage divider with a *balance* impedance Z_B. When $Z_B = Z_0$, the loss from transmit to receive port is infinite. In practice, a fixed compromise impedance Z_B is used, and a component of the receive signal (A) can leak through to (B) with an attenuation as small as 6 to 10 dB due to the variation in impedance of the two-wire facility.

The signal and two types of echo paths for the configuration of Fig. 18-36 are shown in Fig. 18-38. An *echo* is defined as a signal component that has taken any path other than the *talker speech path*. The *talker echo* is the signal that leaks through the far-end hybrid and

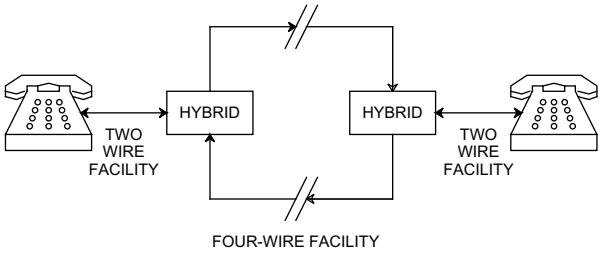

Fig. 18-36. A simplified telephone channel, showing the two-wire local loop and the four-wire transmission facility.

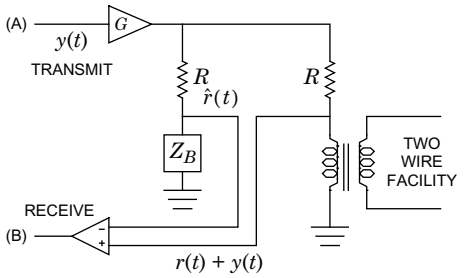

Fig. 18-37. An electronic hybrid. To avoid leakage of the receive signal (A) into the transmit path (B) the impedance Z_B should exactly match the impedance of the transformer and two-wire facility.

returns to the sender (talker). The *listener echo* is the component of the talker echo that leaks through the near-end hybrid and returns again to the listener. This echo is similar to multipath propagation on radio channels (Section 18.4). The length of the telephone channel determines the round-trip echo delay. Echoes from the near end of the connection typically undergo zero to 2 msec of delay, whereas far-end echoes can have round-trip delays of 10-60 msec for terrestrial facilities, or up to 600 msec on satellite connections.

To mitigate the effects of echo on speech quality, several strategies co-exist on the network. The effect of each strategy on data signals is different. For short delays, loss is added in the talker speech path, which is advantageous because the echoes experience this loss more than once. This loss, plus the loss of the subscriber loops at each end, is the source of the attenuation that must be accommodated by data transmission; it can be as high as 40 dB (at 1004 Hz). For longer delays, devices known as *echo suppressors* and *echo cancelers* are added to the connection.

A *full-duplex* (FDX) modem is one that transmits and receives on the same telephone channel. Such a modem requires an internal two-to-four-wire conversion, as shown in Fig. 18-39. Because of imperfect balance impedances of the hybrids, some of the transmitted signal echoes into the receiver and interferes with the weaker data signal from the far end. The hybrid echo loss may be as low as about 6 dB, and the received signal may have experienced as much as 40 dB loss, so the desired far-end signal may be as much as about 34 dB *below* the echo. Ways of dealing with this problem are discussed in Chapters 17 and 20.

18.5.2. Channel Capacity Compared to Practical Modems

Rough estimates of the capacity of a voiceband telephone channel indicate it is over 30,000 b/s. In Table 18-2 we summarize the bit rates achieved by existing standardized voiceband data modems. Bit rates as high as 28,800 b/s are envisioned, although the higher rates may be achievable on a smaller fraction of possible connections. Indeed, several of the higher speed modems are used exclusively with *leased lines*, which can be conditioned for guaranteed quality.

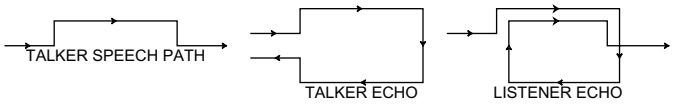

Fig. 18-38. Three of many possible signal paths in a simplified telephone channel with a single two-to-four-wire conversion at each end.

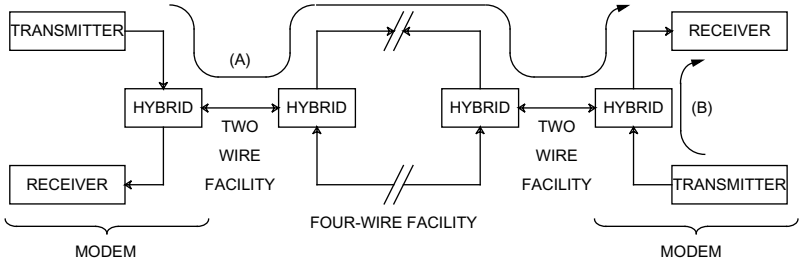

Fig. 18-39. Two modems connected over a single simplified telephone channel. The receiver on the right must be able to distinguish the desired signal (A) from the signal leaked by its own transmitter (B).

18.6. MAGNETIC RECORDING CHANNELS

Digital communication is used not only for communication over a distance (from here to there), but also for communication over time (from now to then). The latter application is called *digital storage* or *recording*, and is usually accomplished using a magnetic medium in the form of a tape or disk. More recently, particularly in the context of read-only applications, optical storage media have been used as well.

Example 18-29. The compact disk ROM, an offshoot of a similar consumer audio technology, allows 600 megabytes of data to be stored on a single plastic disk 12 cm in diameter [26]. The bits are stored as small pits in the surface, and are read by spinning the disk, shining a laser diode on the surface, and detecting the reflected light with an optical pickup.

Digital recording is of course used extensively in computing systems, but is increasingly used in addition for the storage of music [27][28] or voice.

Example 18-30. The compact disk digital audio system, which is a great commercial success, records music digitally using a similar technology to the compact disk ROM. The music is converted to digital using 16 bits per sample at a sampling rate of 44.1 kHz for each of two channels, for a total bit rate of about 1.4 Mb/s. Up to 70 minutes of material can be recorded on a single disk.

Example 18-31. Digital storage on disk drives is used in speech store-and-forward systems, which are essentially the functional replacement for telephone answering systems, except that they serve a number of customers.

Digital recording offers some of the same advantages over analog recording as we discussed earlier for transmission. The principle advantage again is the *regenerative effect*, in which the recording does not deteriorate with time (except for the introduction of random errors which can be eliminated by coding techniques) or with multiple recordings and re-recordings. An additional advantage is the compatibility of digital recording with digital signal processing, which offers very powerful capabilities.

Table 18-2. Important standardized voiceband data modems are summarized here. The "duplex" column indicates whether a single channel is shared for both directions of transmission (full) or separate channels must be used for each direction (half). For full duplex modems, it also indicates whether frequency division multiplexing (FDM) or echo cancellation (EC) is used for multiple access (Chapter 17). The "CCITT std" column identifies the international standard that applies to this type of transmission. Finally, the "modulation" column identifies the type of modulation, which are discussed in Chapters 5 and 13. The numbers indicate the number of symbols in the alphabet. The "TC" in the V.32 and V.33 refers to trellis coding (Chapter 13).

speed (b/s)	symbol rate (Hz)	duplex — method	CCITT standard	modulation
≤ 300	≤ 300	full — FDM	V.21	2-FSK
1200	1200	half	V.23	2-FSK
1200	600	full — FDM	V-22	4-PSK
2400	1200	half	V.26	4-PSK
2400	600	full — FDM	V.22bis	16-QAM
2400	1200	full — EC	V.26ter	4-PSK
4800	1600	half	V.27	8-PSK
4800	2400	full — EC	V.32	4-QPSK
9600	2400	half	V.29	16-AM/PM
9600	2400	full — EC	V.32	32-QAM + TC
14,400	2400	full — EC	V.32bis	128-QAM + TC
≤ 28,800	≤ 3429	full — EC	V.fast (V.34)	1024-QAM + TC

Magnetic tape or disk can be considered as a transmission medium in much the same manner as other media such as wires and fibers [29][30]. We will now briefly discuss the properties of that medium.

18.6.1. Writing and Reading

In the writing process, a magnetic field is generated in an electromagnet called a *head* as it passes at high speed over a ferric oxide magnetic medium, thereby orienting the direction of magnetization along a track in a nearby magnetic medium on the disk or tape [31]. On reading, when the oriented magnetic pattern passes under that same head, it produces a voltage that can be sensed and amplified.

There are two basic types of recording. *Saturation recording* is almost always used for digital recording, in which the magnetization is saturated in one direction or the other. Thus, in saturation recording, the medium is constrained to be used for binary transmission; that is, only two levels are allowed. This is in contrast to wire and coaxial media in which multi-level transmission can be considered. The other form of magnetic recording is *a.c. bias recording*, in which the signal is accompanied by a much larger and higher frequency bias sinusoid for the purpose of linearizing the channel. A.c. bias recording is necessarily used in analog recording, where linearity is important, but has not been applied to digital recording because of the deterioration in signal-to-noise ratio and the fact that saturation recording is appropriate for binary modulation and demodulation techniques.

The magnetic recording process is qualitatively illustrated in Fig. 18-40. For saturation recording, the voltage applied to the write head assumes one positive and one negative value corresponding to the two directions of desired magnetization. In Fig. 18-40a it is assumed that a square wave corresponding to the binary sequence "1101" is applied to the write head. This waveform correspondence to a bit sequence is called *non-return to zero*, or *NRZ*. The bit stream is recorded on linear (tape) or circular (disk) *tracks* on the magnetic medium, and one track is shown in Fig. 18-40b. Note the two directions of magnetization, schematically indicated by the arrows. The voltage on the read head (which is physically the same as the write head) during a read operation is shown in Fig. 18-40c. As long as the magnetization is constant, no voltage is induced in the read head coil, but upon a *change* in magnetization there is a voltage induced (recall that the voltage induced in a coil is proportional to the *derivative* of the magnetic field). The polarity of that voltage is determined by the *direction of change* in magnetization.

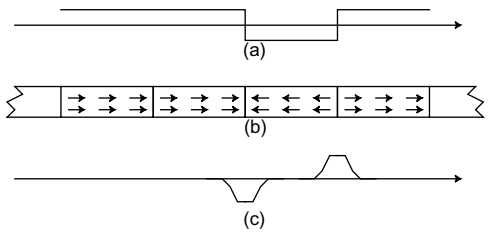

Fig. 18-40. Illustration of magnetic recording. a. The NRZ waveform applied to the record head corresponding to bit sequence "1101." The abscissa is time, but this is proportional to distance on the medium for constant velocity of the head. b. The magnetization of one track after saturation recording. b. The voltage on the read head coil corresponding to position of the read head, which at constant velocity is the same as time.

This write-read magnetic recording process can be viewed as a communication channel if we observe only the input and output voltage waveforms in Fig. 18-40a andFig. 18-40c and ignore the physical medium of Fig. 18-40b. Both of these waveforms represent signals in time, just like in communications, although there is a conceptually insignificant and indeterminate time delay between the write and read operations. Viewed as a communication channel, we see that the magnetic recording channel of Fig. 18-40 inherently includes a differentiation operation. Another way of looking at this is that the channel is sensitive to only the *transitions* in the input waveform rather than its polarity. Therefore, from a digital communication point of view, we want to map the input bits into transitions in the input waveform rather than absolute polarity.

18.6.2. Linearity of the Magnetic Channel

The magnetic channel can be made linear in a special sense to be specified now. This linearity is a very desirable feature, in that it will greatly simplify system design.

The view of Fig. 18-40 is oversimplified in that it assumes that the magnetization is in either one direction or the other. In fact, the tape medium contains a multiplicity of tiny magnetic particles, and each particle must indeed be magnetized in one direction or the other. The total net magnetization can assume almost a continuum of values, depending on the number of particles magnetized in each direction. Unfortunately this continuum of magnetization depends nonlinearly on the applied magnetic field, and displays hysteresis, and therefore the write process is highly nonlinear. On the other hand, the read process is very linear, in that the voltage induced on the read head is a linear function of the magnetization.

If the applied field to the recording head is strong enough and held long enough so that the medium is fully saturated, then the output of the read head displays a form of superposition. This is because this saturation destroys the memory of the hysteresis. This form of superposition is illustrated in Fig. 18-41. If the response to a positive transition at time t is $h(t)$, and the response to a negative transition at time $t + \Delta$ is $-h(t + \Delta)$, then the response to the positive followed by negative transition obeys superposition, and is

$$h(t) - h(t + \Delta) \; . \tag{18.79}$$

This is true with great accuracy *as long as* the time between transitions Δ is larger than some threshold Δ_0. This threshold is determined by the time to achieve full saturation of the medium in one direction or the other since the last transition, and depends on the design of the write head as well as the medium.

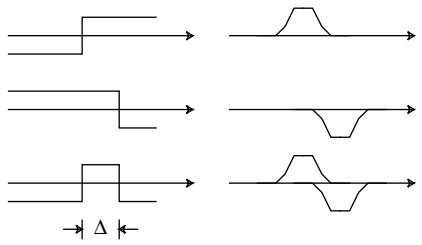

Fig. 18-41. Superposition in the reading process of magnetic recording.

18.6.3. Noise on the Magnetic Channel

The noise impairments are very complicated on the magnetic channel, consisting of additive and multiplicative noise components. A major source of noise is due to the granularity of the medium. The total response of the head is the superposition of the responses to a multiplicity of magnetic particles. This discrete nature of the signal is similar to the quantum nature of the optical detection process (Section 18.3) with one important distinction. In optical detection we have only photons (or photoelectrons) or the absence of same, whereas in magnetics there are both positively and negatively magnetized particles. Thus, in optical detection, when there is no signal incoming there is also no quantum noise (neglecting dark current). In the magnetic reading process the particles are present whether or not there is a signal, or putting it another way the absence of a signal is represented by an equal number of positively and negatively magnetized particles. Hence, the *granular noise* in magnetic recording is present independent of the signal, and is therefore truly an additive noise phenomenon. Its spectrum is not white because it is filtered by the read head response, and in fact its power spectrum tends to be similar to the spectrum of the read signal.

Zero crossing jitter results from variations in the composition of the medium and the distance between the write head and the medium. The effect is a small jitter in the position of the read pulses. Another phenomenon is *amplitude modulation* of the received signal, a multiplicative noise phenomenon due to medium density fluctuations. An extreme form of amplitude modulation is the tape dropout, in which the signal level gets too small for reliable detection. Since dropouts are a reproducible effect, depending on position on the medium, they are often flagged in disk files so that these areas are not used for recording. Another phenomenon is interference from adjacent tracks, which is similar to the crosstalk experienced in multi-wire-pair cables and the interference between radio systems.

18.7. FURTHER READING

The literature on communication media is vast and scattered. We offer here some suggestions that may help the reader get started. The basic theory of transmission lines is covered in [1]. There are a number of books devoted to optical fiber devices [32][6][33][34][35][36][37][38] and a smaller number that concentrate on the noise and system issues of primary interest here [39][13][14]. A special issue of *IEEE Journal on Selected Areas in Communications* (November, 1984) is devoted to undersea lightwave communication. It contains numerous useful articles describing the special problems that arise in this hostile environment. Another issue (November, 1985) covers fiber optics for local communication, and concentrates on networking issues. Yet another issue (December, 1986) is devoted to terrestrial fiber optics, and includes papers on reliability, economics, and networking issues. Finally, the Optical Society of America's *Journal of Lightwave Technology* is a good source of information.

There are available books on satellite [18] and mobile radio [21] design. Special issues of the *IEEE Journal on Selected Areas in Communications* in July 1984, June 1987, and January 1989 are is devoted to mobile radio communication. Many of the papers propose modulation schemes that are robust in the presence of multipath fading. More specifically directed to

multipath fading channels is another special issue (February 1987). Another issue (April 1987) is devoted to point-to-point digital radio, and yet another (January 1985) to broadcasting satellites.

Further information about characteristics of the telephone channel is best obtained by going directly to the published line surveys [22][23][24]. Special issues of *IEEE Journal on Selected Areas in Communications* (September 1984, and August and December 1989) are devoted to modulation and coding for the telephone channel.

A special issue of *IEEE Journal on Selected Areas in Communications* in January 1992 is devoted to recent results on magnetic recording channels.

Problems

Problem 18-1. Show that for a terminated transmission line with real-valued characteristic impedance, the maximum power to the load is obtained in Fig. 18-6b when $Z_S = Z_L = Z_0$.

Problem 18-2. For a transmission line, derive the relation $\lambda f = v$, where f is the frequency of a propagating wave in Hz, λ is the wavelength in meters, and v is the velocity of the wave in meters/sec.

Problem 18-3. In subscriber loop wire-pair cables, it is common in some countries to have *bridged taps*, which are *open circuited* wire-pairs bridged in parallel on the main pair. Assume that a source has impedance equal to the wire-pair characteristic impedance, the wire-pair is terminated at the other end by its characteristic impedance, and that the wire-pair has a single bridged tap. Let the distance from source to tap be L_1, from tap to termination L_2, and let the length of the bridged tap be L_3.

(a) Find an expression for the transfer function of the wire-pair including bridged tap. Be sure to take advantage of the simplifications due to the terminations with the characteristic impedance.

(b) Show that when the bridged tap is very long, it causes a fixed attenuation at all frequencies. What is that attenuation?

(c) State intuitively what you would expect the response at the termination to be to a single transmitted pulse as a function of the length of the bridged tap.

(d) Discuss what happens intuitively when the bridged tap approaches zero length.

Problem 18-4. Use Snell's law to show that in Fig. 18-12 a ray will be captured by the fiber as long as the incident angle obeys

$$sin(\theta_1) < \sqrt{n_1^2 - n_2^2}\,. \tag{18.80}$$

This confirms that rays incident at small angles are captured, and those at larger angles are not.

Problem 18-5. Let the length of the fiber be L.

(a) Show that the path length for a ray is equal to $L \cdot \sec(\theta_2)$.

(b) Show that the path length varies from L to $(n_1^2/n_2^2)\,L$. Thus, the larger the difference in index of refraction of core to cladding, the larger the range of captured angles, but also the larger the variation in the transit time of rays through the length of the fiber.

Problem 18-6. Assuming that the chromatic dispersion in a single mode fiber is 0.15 psec/km-GHz, evaluate numerically (18.30). Sketch the curve of repeater spacing vs. bit rate in the range of repeater spacings between 1 and 1000 km as limited by dispersion.

Problem 18-7. In an optical fiber receiver, assume the received optical power is P, the bit rate is R bits/sec.

(a) Find the number of received photons per bit.

(b) Show that for a constant number of photons per bit, the required received optical power is proportional to the bit rate.

(c) Find the received optical power necessary to receive 100 photons per bit at a wavelength of 1.5 μmeter and a bit rate of 1 Gb/s.

(d) For the same conditions as part (c), assume you can launch one mwatt power into the fiber, and that the fiber loss at that wavelength is 0.2 dB per km. What is the distance that we can transmit?

Problem 18-8. A typical direct detection optical receiver requires about $N = 2000$ photons per bit in the notation of Problem 18-7.

(a) Derive the following formula [9] for the required received power at an optical detector at a wavelength of 1.5 μmeter for this value of N,

$$P_{\text{dBm}} = -65.8 + \log_{10} R_{\text{Mb}} \,, \tag{18.81}$$

where P_{dBm} is the received power in dBm required and R_{Mb} is the bit rate in Mb/s. Note how the required power increases as the bit rate increases. In particular, each order of magnitude increase in bit rate increases the required power by only one dB.

(b) Assuming 0 dBm launched power into the fiber, and 0.2 dB per km loss in the fiber, what is the allowable distance between repeaters at bit rates of 100 and 1000 Mb/s? You can assume that loss is the dominant impairment limiting repeater spacing.

Problem 18-9. Change the assumptions in Problem 18-8 to those that might better reflect fundamental limits [9]: A launched signal power of 20 dBm and 20 photons per bit required at the receiver.

Problem 18-10. Suppose we have a system requirement that a total bit rate of R_T must be transmitted over a distance of L_T using a set of parallel repeatered transmission lines (wire cable or fiber). In each repeater span we have as design parameters the bit rate B and repeater spacing L. Show that the total number of repeaters is minimized when the quantity $B \cdot L$ is maximized for the given transmission technology. Thus, if the repeaters are the dominant transmission cost, we want to maximize the product of the bit rate and the distance for each technology.

Problem 18-11.

(a) Derive the following relationship between repeater spacing and bit rate, using the assumptions of Problem 18-8, and assuming a fiber loss of γ_0 dB/km:

$$L = \frac{65.8 - \log_{10} R_{\text{Mb}}}{\gamma_0} \,. \tag{18.82}$$

You can assume that the number of received photons per bit is held constant and the transmit power is held constant at 0 dBm.

(b) Sketch this relation for the range of bit rates between 1 Mb/s and 10,000 Mb/s and a fiber loss of 0.2 dB/km and verify that Fig. 18-16 is qualitatively correct is predicting this loss-limited region.

(c) Using the results of Problem 18-10, argue that it will be best to increase the bit rate until dispersion becomes the controlling impairment, if the criterion is to minimize the number of repeaters.

Problem 18-12. The available thermal noise power in a bandwidth B Hz is kT_nB. For a resistor generating thermal noise, the noise source can be modeled as either a series voltage source or a parallel current source. Show that the voltage source has mean-squared voltage $4kT_nRB$ and the current source has mean-squared current $4kT_nB/R$ within bandwidth B for a resistance of R ohms.

Problem 18-13. At 6 GHz, what is the diameter of a circular aperture antenna that has an antenna gain of 40 dB with a 70% efficiency?

Problem 18-14. A radio channel with bandwidth 30 MHz is centered at 6 GHz. What is the difference in decibels in the path loss between the high and low end of this channel? At which end is the loss the minimum? (Caution: The antenna gains are a function of frequency.)

Problem 18-15. Compare the trade-off between repeater spacing d and transmitted power P_T, assuming that the received power P_R is held constant for the following two media:

(a) Metallic cable or fiber optics with loss γ_0 dB per km.

(b) A microwave radio system.

(c) For which medium does the transmitted power have the most impact?

Problem 18-16. Develop the following formula which relates the free-space loss between isotropic radiators in dB, the distance d_{km} between radiators in km, and the frequency f_{GHz} in GHz,

$$\text{Loss (dB)} = 92.4 + 20\log_{10} d_{\text{km}} + 20\log_{10} f_{\text{GHz}}. \tag{18.83}$$

Note the dependence of this loss on distance and frequency.

Problem 18-17. In this problem we will determine how to combine noise sources in a radio receiver with different noise temperatures to yield a single equivalent noise source. For the configuration of Fig. 18-42 where the noise temperature of the three noise sources $n_i(t)$ are T_i, find the relationship between these noise temperatures such that the two systems will have the same SNR. The parameter G is the power gain of an amplifier, i.e., the ratio of input to output power.

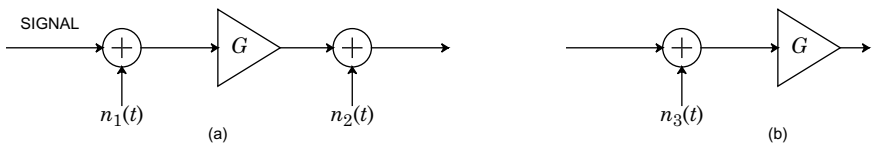

Fig. 18-42. Illustration of the combination of two noise sources into a single equivalent noise source referenced to the input of the system.

Problem 18-18. Use the results of Problem 18-17 to find the equivalent noise temperature at the input to the receiver of Fig. 18-43, where each of the circuit elements is assumed to be noiseless with an associated noise source with associated noise temperature.

Problem 18-19. Estimate the delay spread for the two-path ground-reflection model of Exercise 18-3 for a spacing of antennas by 3 km and antenna height of 50 meters. What is the approximate baseband bandwidth over which the narrowband model is applicable?

Problem 18-20. Suppose the incoming power in a Rayleigh fading scenario does not arrive at a moving vehicle uniformly spread over all angles. Describe qualitatively how you would expect the power spectrum of Fig. 18-32 to be affected under the following conditions:

(a) The vehicle is driving toward the transmitter, and more power is arriving from the direction of the transmitter than other directions.

(b) A lot of power is reflecting off a nearby mountain, so that more power is arriving at the vehicle from the side (relative to the direction of motion) than any other direction.

Problem 18-21. Consider a SSB analog voice transmission system embedded in the telephone network. Suppose that the carrier frequency f_c is nominally 1 MHz. In practice, the transmitter and receiver will be designed with components that yield modulating and demodulating frequencies that are slightly off. Component manufacturers often express the accuracy of precision parts in *parts per million*, instead of percent (which is *parts per hundred*). How accurate (in parts per million) do the modulating and demodulating oscillator frequencies have to be to guarantee less than 3 Hz frequency offset?

Problem 18-22. Suppose that a data signal

$$x(t) = \mathrm{Re}\{s(t)e^{j2\pi f_c t}\} \tag{18.84}$$

is transmitted over a telephone channel with frequency offset f_0 and sinusoidal phase jitter with frequency f_p and amplitude a. Assume there are no other impairments. Give an expression for the received signal.

References

1. Temes, G. C. and J. W. LaPatra, Introduction to Circuit Synthesis and Design, McGraw-Hill, New York, 1967.

2. Staff, Bell Laboratories Members of Technical, Transmission Systems for Communications, Western Electric Co., Winston- Salem N.C., 1970.

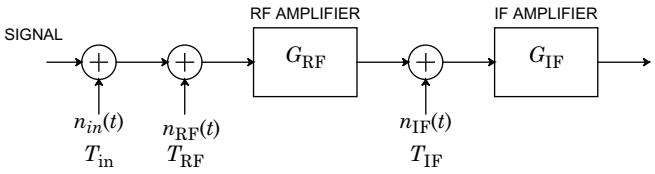

Fig. 18-43. Several noise sources introduced at the input and internally to a receiver.

3. Balloonist, P. and D. G. W. Ingram, Digital Transmission Systems, Peter Peregrinus Ltd., Stevenage England, 1976.

4. Ahamed, S. V., P. P. Bohn, and N. L. Gottfried, “A Tutorial on Two-Wire Digital Transmission in the Loop Plant,” IEEE Trans. on Communications, vol. COM-29, Nov. 1981.

5. Kao, K. C. and G.A. Hockham, “Dielectric-Fiber Surface Waveguides for Optical Frequencies,” Proc. IEE, vol. 113, p. 1151, July 1966.

6. Keck, D. B., “Fundamentals of Optical Waveguide Fibers,” IEEE Communications, vol. 23, no. 5, May 1985.

7. Verdeyen, J. T., Laser Electronics, Prentice Hall, Englewood Cliffs N.J., 1981.

8. Keck, D. B., “Single-Mode Fibers Outperform Multimode Cables,” IEEE Spectrum, vol. 20, no. 3, p. 30, March 1983.

9. Midwinter, J. E., “Performance Boundaries for Optical Fibre Systems,” NATO Advanced Study Institute, Castelvecchio, Pascoli, Italy, July 1986.

10. Henry, P. S., “Introduction to Lightwave Transmission,” IEEE Communications, vol. 23, no. 5, May 1985.

11. T. Li, “Structures, Parameters, and Transmission Properties of Optical Fibers,” Proc. IEEE, vol. 68, no. 10, p. 1175, Oct. 1980.

12. Forrest, S. R., “Optical Detectors: Three Contenders,” IEEE Spectrum, vol. 23, no. 5, p. 76, May 1986.

13. Personick, S. D., Optical Fiber Transmission Systems, Plenum Press, New York, 1981.

14. Personick, S. D., Fiber Optics Technology and Applications, Plenum Press, New York, 1985.

15. Taylor, D. and P. Hartmann, “Telecommunications by Microwave Digital Radio,” IEEE Communications Magazine, vol. 24, no. 8, p. 11, Aug. 1986.

16. Mikulski, J., “DynaT*A*C Cellular Portable Radiotelephone System Experience in the U.S. and the U.K.,” IEEE Communications Mag., vol. 24, no. 2, p. 40, Feb. 1986.

17. MacDonald, V., “The Cellular Concept,” *Bell System Technical Journal*, vol. 58, no. 1, Jan. 1979.

18. Pratt, T. and C. W. Bostian, Satellite Communications, John Wiley, New York, 1986.

19. W, Rummler, R. Coutts, and M. Liniger, “Multipath Fading Channel Models for Microwave Digital Radio,” *IEEE Communication Magazine*, vol. 11, p. 30, Nov. 1986.

20. Chamberlain, J., F. Clayton, H. Sari, and P. Vandamme, “Receiver Techniques for Microwave Digital Radio,” IEEE Communications Mag., vol. 24, no. 11, p. 43, Nov. 1986.

21. Jakes, W.C. Jr., Microwave Mobile Communications, Wiley- Interscience, New York, 1974.

22. Duffy, F. P. and T. W. Thatcher, Jr., “1969-70 Connection Survey: Analog Transmission Performance on the Switched Telecommunications Network,” *Bell System Technical Journal*, vol. 50, no. 4, pp. 1311-47, April 1971.

23. Carey, M. B., H.-T. Chen, A. Descloux, J. F. Ingle, and K. I. Park, “1982/83 End Office Connection Study: Analog Voice and Voiceband Data Transmission Performance Characterization of the Public Switched Network,” AT&T Bell Lab. Tech. J., vol. 63, no. 9, Nov. 1984.

24. Batorsky, D. V. and M. E. Burke, “1980 Bell System Noise Survey of the Loop Plant,” AT&T Bell Lab. Tech. J., vol. 63, no. 5, pp. 775-818, May-June 1984.

25. Saltzberg, B. R. and J.-D. Wang, “Second-order statistics of logarithmic quantization noise in QAM data communication,” IEEE Transactions on Communications, vol. 39, no. 10, pp. 1465-72, Oct. 1991.

26. Chen, P., “The Compact Disk ROM: How It Works,” IEEE Spectrum, vol. 23, no. 4, p. 44, April 1986.

27. Miyaoka, S., “Digital Audio is Compact and Rugged,” IEEE Spectrum, vol. 21, no. 3, p. 35, March 1984.

28. Bloom, P. J., “High-Quality Digital Audio in the Entertainment Industry,” IEEE ASSP Magazine, vol. 2, no. 4, p. 2, Oct. 1985.

29. Mallinson, J. C., “A Unified View of High Density Digital Recording Theory,” IEEE Trans. on Magnetics, vol. MAG-11, p. 1166, Sep. 1975.

30. Kobayashi, H., “A Survey of Coding Schemes for Transmission or Recording of Digital Data,” *IEEE Trans. on Communications*, vol. COM-19, p. 1087, Dec. 1971.

31. Bertram, H., “Fundamentals of the Magnetic Recording Process,” IEEE Proceedings, vol. 74, no. 11, p. 1494, Nov. 1986.

32. Marcuse, D., Light Transmission Optics, Van Nostrant Reinhold, Princeton, N.J., 1972.

33. Marcuse, D., Theory of Dielectric Optical Waveguides, Academic Press, New York, 1974.

34. Gloge, D., Optical Fiber Technology, IEEE Press, New York, 1976.

35. Unger, H. H., Planar Optical Fibers for Transmission, Clarendon Press, Oxford, 1977.

36. Midwinter, J. E., Optical Fibers for Transmission, Wiley, New York, 1979.

37. Miller, S. E. and A.G. Chynoweth, Optical Fiber Telecommunications, Academic Press, New York, 1979.

38. Taylor, H. F., Fiber Optics Communications, Artech House, Dedham, Mass., 1983.

39. Barnoski, M. K., Fundamentals of Optical Fiber Communications, Academic Press, New York, 1976.