'ies

liams

ry

ularikas

ld J. Iafrate

urmuzlu

ularikas

THE MEASUREMENT, INSTRUMENTATION, AND SENSORS HANDBOOK

VOLUME II

Ch. 87 & 88
Telemetry & Sensor Networks & Comms

Editor-in-Chief

John G. Webster

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an informa business

EX, 1985.
mann, 1993.
ddison-Wesley, 1995.
MA: Addison-Wesley,
MA: Addison-Wesley,
umentation.
Commands.

87

Telemetry

Albert Lozano-Nieto
Penn State University

87.1 Introduction

Telemetry is the science of gathering information at some remote location and transmitting the data to a convenient location to be examined and recorded. Telemetry can by done by different methods: optical, mechanical, hydraulic, electric, etc. The mechanical methods, either pneumatic or hydraulic have acceptable results for short distances and are used in environments that have a high level of electromagnetic interference and in those situations where, for security reasons, it is not possible to use electrical signals, for example, in explosive environments. More recently, use of optical fiber systems allows the measurement of broad bandwidth and high immunity to noise and interference. Other proposed telemetry systems are based on ultrasound, capacitive or magnetic coupling, and infrared radiation, although these methods are not routinely used. The discussion in this chapter will be limited to the most-used systems: telemetry based on electric signals. The main advantage of electric over mechanical methods is that electrically based telemetry does not have practical limits regarding the distance between the measurement and the analysis areas, and can be easily adapted and upgraded in already existing infrastructures. Electric telemetry methods are further divided depending on the transmission channel that they use as wire telemetry and wireless (or radio) telemetry. Wire telemetry is technologically the simplest solution. The limitations of wire telemetry are the low bandwidth and low transmission speed that it can support. However, it is used when the transmission wires can use the already existing infrastructure, as, for example, in most electric power lines that are also used as wire telemetry carriers. Wireless telemetry is more complex than wire telemetry, as it requires a final radio frequency (RF) stage. Despite its complexity, it is widely used because it can transmit information over longer distances; thus, it is used in those applications in which the measurement area is not normally accessible. It can also transmit at higher speeds and have enough capacity to transmit several channels of information if necessary.

Figure 87.1 displays a generic telemetry system. It consists of (not all the blocks will be always present) (1) transducers to convert physical variables to be measured into electric signals that can be easily processed; (2) conditioning circuits to amplify the low-level signal from the transducer, limit its bandwidth, and adapt impedance levels; (3) a signal-processing circuit that sometimes can be integrated in the previous circuits; (4) a subcarrier oscillator whose signal will be modulated by the output of the different transducers once processed and adapted; (5) a codifier circuit, which can be a digital encoder, an analog modulator, or a digital modulator, that adapts the signal to the characteristics of the transmission channel, which is a wire or an antenna; (6) a radio transmitter, in wireless telemetry, modulated by

0-8493-8347-1/99/$0.00+$.50

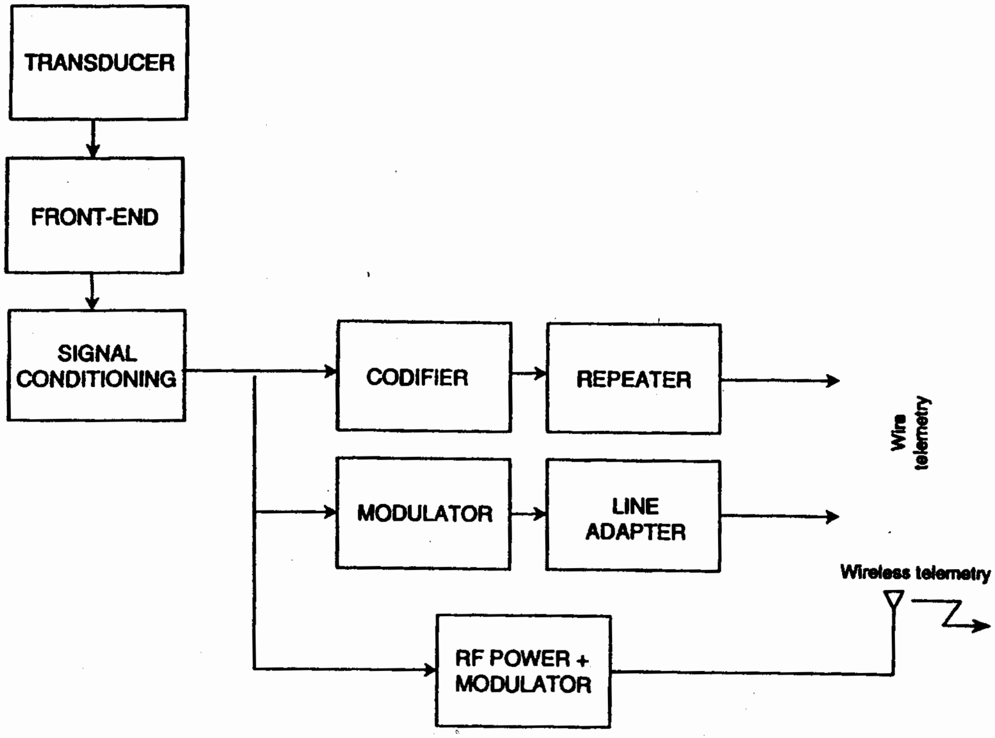

FIGURE 87.1 Block diagram for a telemetry system. Telemetry using wires can be performed in either base-band or by sending a modulated signal, while wireless telemetry uses an RF carrier and an antenna.

the composite signal; (7) an impedance line adapter, in case of wire transmission, to adapt the characteristic impedance of the line to the output impedance of the circuits connected to the adapter; and (8) for wireless communication, a transmitting antenna. The receiver end consists of similar modules. For wireless telemetry, these modules are (1) a receiving antenna designed for maximum efficiency in the RF band used; (2) a radio receiver with a demodulation scheme compatible with the modulation scheme; and (3) demodulation circuits for each of the transmitted channels. For wire telemetry, the antenna and the radio receiver are replaced by a generic front end to amplify the signal and adapt the line impedance to the input impedance of the circuits that follow. The transmission in telemetry systems, in particular wireless ones, is done by sending a signal whose analog variations in amplitude or frequency are a known function of the variations of the signals from the transducers. More recently, digital telemetry systems send data digitally as a finite set of symbols, each one representing one of the possible finite values of the composite signals at the time that it was sampled. The effective communication distance in a wireless system is limited by the power radiated by the transmitting antenna, the sensitivity fo the receiver and the bandwidth of the RF signal. As the bandwidth increases, the contribution of noise to the total signal also increases, and consequently more transmitted power is needed to maintain the same signal-to-noise ratio (SNR). This is one of the principal limitations of wireless telemetry systems. In some applications, the transmission to the receiver is done on base band, after the conditioning circuits. The advantage of base-band telemetry systems is their simplicity, although because of the base-band transmission, they are normally limited to only one channel at low speeds.

Not uncommonly, measurement system needs to acquire either different types of signals or the same type of data at different locations in the process that is being monitored. These different information signals can be transmitted using the same common carrier by multiplexing the data signals. Multiplexing allows different signals to share the same channel. Multiplexing techniques are usually considered either frequency division multiplexing (FDM) or time division multiplexing (TDM). In FDM, different subcarrier

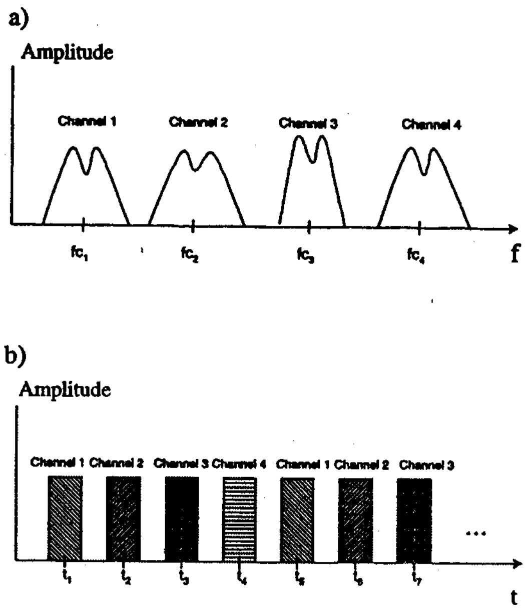

FIGURE 87.2 Basic characteristics of (a) FDM and (b) TDM signals. In FDM different channels are allocated at different subcarrier frequencies (f_{c1}, f_{c2},...) while in TDM only one channel is transmitted at a given time. The remaining channels are transmitted sequentially.

frequencies are modulated by the different measurement channel signals, which causes the information spectrum to shift from base band to the subcarrier frequency. Then, the subcarrier frequencies modulate the RF carrier signal, which allows the transmission of all desired measurement channels simultaneously. In TDM, the whole channel is assigned entirely to each measurement channel, although only during a fraction of the time. TDM techniques use digital modulation to sample the different measurement channels at different times. Then, these samples are applied sequentially to modulate the RF carrier. Figure 87.2 illustrates these concepts by showing frequency and time graphs for FDM and TDM, respectively.

Almost all instrumentation and measurement situations are candidates for use of a telemetry link. Telemetry is widely used in space applications for either telemeasurement of a distant variable or telecommandment of actuators. In most of these types of applications, for example, in space telemetry, it is very important to design the telemetry systems to minimize the consumption of power [1]. Some land-mobile vehicles, such as trains, also use telemetry systems, either wireless or by using some of the existing power wires to transmit data to the central station and receive its commands [2]. In clinical practice, the telemetry of patients increases their quality of life and their mobility, as patients do not need to be connected to a measurement system to be monitored. Several medical applications are based on implanting a sensor in a patient and transmitting the data to be further analyzed and processed either by radio [3] or by adapted telephone lines [4] from the receiving station. Optical sensors and fiber-optic communications are used in industry to measure in environments where it is not desirable to have electric signals such as explosive atmospheres [5]. The designer of a telemetry system needs also to keep in mind the conditions in which the system will have to operate. In most of the applications, the telemetry systems must operate repeatedly without adjustment and calibration in a wide range of temperatures. Finally, as different telemetry systems are developed, the need to permit tests to be made interchangeable at all ranges increases, which require compatibility of transmitting, receiving, and signal-processing equipment at all ranges. For this reason, the Department of Defense Research and Development Squad created the Guided Missiles Committee, which formed the Working Group on Telemetry. This later became the Inter-Range Instrumentation Group (IRIG) that developed Telemetry Standards. Today, the IRIG Standard 106-96 is the primary Telemetry Standard used worldwide by both government and industry.

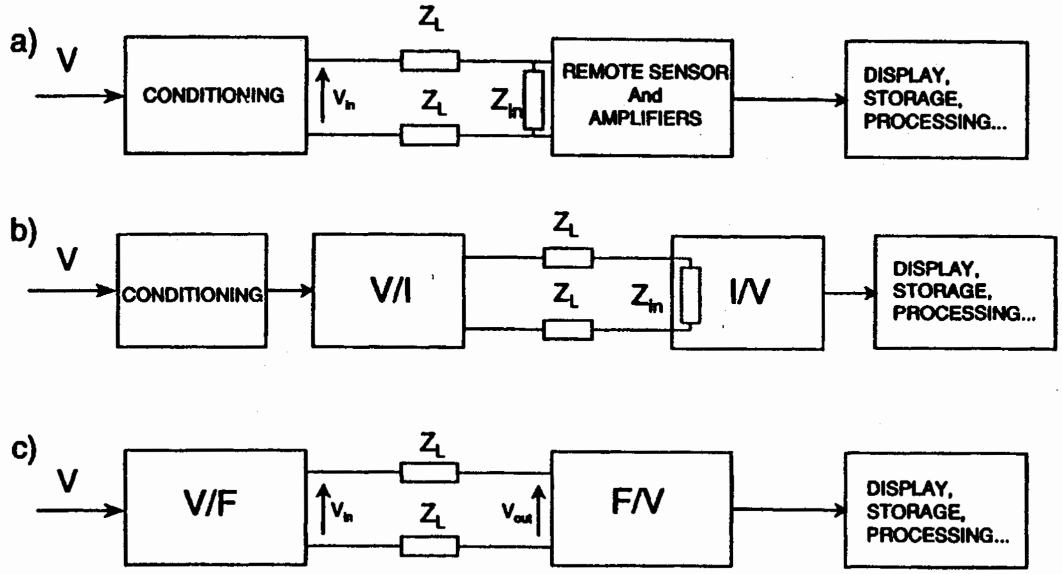

FIGURE 87.3 Different configurations for base-band telemetry. In voltage-based-base band telemetry (a) the information is transmitted as variations of a voltage signal. Current-based-base band telemetry (b) is based on sending a current signal instead of a voltage signal to neutralize the signal degradation due to the voltage divider made up by the input impedance of the receiver (Z_{in}) and the impedance of the lines (Z_L). In frequency-based base-band telemetry (c), the information is transmitted as variations of frequency which makes this system immune to noise and interference that affect the amplitude of the transmitted signal.

87.2 Base-Band Telemetry

Base-band telemetry uses a wire line to communicate the signal from the transducer after being processed and conditioned with the receiver. We will briefly describe telemetry systems based either on amplitude or frequency. More in-depth study of these base-band telemetry systems can be found in Reference 6.

Base-Band Telemetry Based on Amplitude

Voltage-Based Base-Band Telemetry

Figure 87.3a shows a simple voltage-based telemetry system. The signal from the transducer is amplified, normally to a voltage level between 1 and 15 V, and sent through a line consisting of two wires to the receiver. By making the low end of the scale 1 V, this system can detect short circuits [6]. The main problem of this configuration is the limitation on the transmission distance, which depends on the resistance of the line and the input resistance for the receiver. Also, the connecting wires form a loop that is very susceptible to interference from parasitic signals.

Current-Based Base-Band Telemetry

The limitation on transmission distance of the voltage-based system due to the impedance of the line are solved by using a current signal instead of a voltage, as is shown in Figure 87.3b. This requires an additional conversion module after the signal-processing circuits from voltage to current. At the receiver end, the signal is detected by measuring the voltage across a resistor. The most-used system in industry is the 4 to 20 mA loop. This means that 0 V is transmitted as 4 mA, while the highest voltage value is transmitted as a 20-mA current. The advantage of transmitting 4 mA for 0 V is the easy detection of an open circuit in the loop (0 mA). Other standard current values are 0 to 5, 0 to 20, 10 to 50, 1 to 5, and 2 to 10 mA. Also, voltage drops due to resistance of the wires do not affect the transmitted signal, which allows the use of thinner wires. Because this is a current mode, the parasitic voltages induced in the line

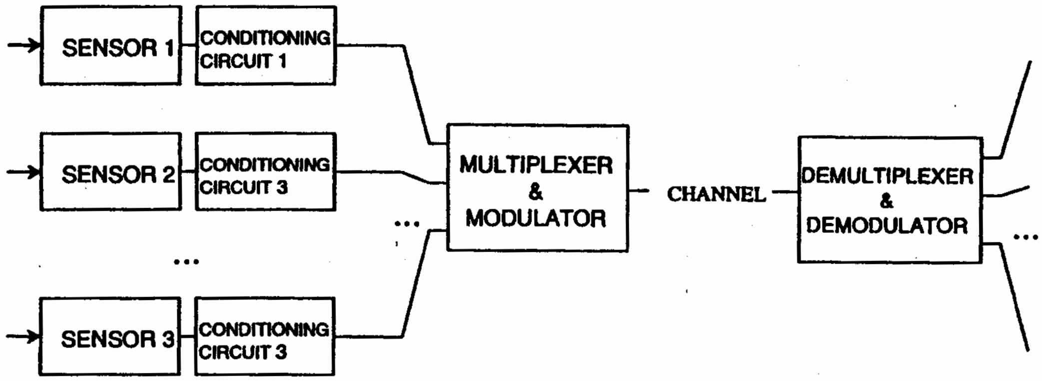

FIGURE 87.4 In multiple-channel telemetry a common transmission channel is used to transmit the measured signals from different channels using different sharing schemes.

do not affect the signal either. Current-based telemetry allows the use of grounded or floating transmitters with few modifications [6].

Base-Band Telemetry Based on Frequency

Frequency-based transmission is known to have higher immunity to noise than amplitude-based transmission. Frequency-based telemetry, shown in Figure 87.3c, is used in the presence of inductive or capacitive interference due to its immunity to noise. It also offers the possibility of isolating the receiver from the transmitter. The signal at the output of the conditioning circuit modifies the frequency of the telemetry signal, normally using a voltage-to-frequency converter. In the receiver, a frequency-to-voltage converter performs the opposite function. A special case of frequency-based telemetry is pulse telemetry, in which the modulating signal changes some characteristics of a train of pulses. Because of its importance and widespread use, pulse telemetry will be analyzed in-depth in the following sections.

87.3 Multiple-Channel Telemetry

Most of the industrial processes in which telemetry is used require the measurement of different physical variables to control the process, the measurement of only one physical variable at different locations, or normally a combination of both. In these multiple-channel measurements, base-band telemetry is not an option, as it would require building a different system for each channel. Multiple channel telemetry is achieved by sharing a common resource (transmission channel), as is shown in Figure 87.4. The sharing of the transmission channel by all the measurement channels is designated by *multiplexing*. There are two basic multiplexing techniques: FDM and TDM. In FDM, different channels are assigned to different spectral bands and the composite signal is transmitted through the communication channel. In TDM, the information for different channels is transmitted sequentially through the communication channel.

Frequency Division Multiplexing

In FDM, shown in Figure 87.5a, each measurement channel modulates a sinusoidal signal of different frequency. These sinusoidal signals are called subcarriers. Each of the modulated signals is then low-pass-filtered to ensure that the bandwidth limits are observed. After the filtering stage, all the modulated signals are fed into a summing block, producing what is known as a base-band signal. A base-band signal indicates here that the final carrier has not yet been modulated. The spectrum of the base-band signal is shown in Figure 87.5b, where it is possible to see how each measurement channel spectrum signal is allocated its own frequency. This composite signal finally modulates a carrier signal whose frequency depends on the transmission medium that is used. The signal is then fed into a transmission wire (similar

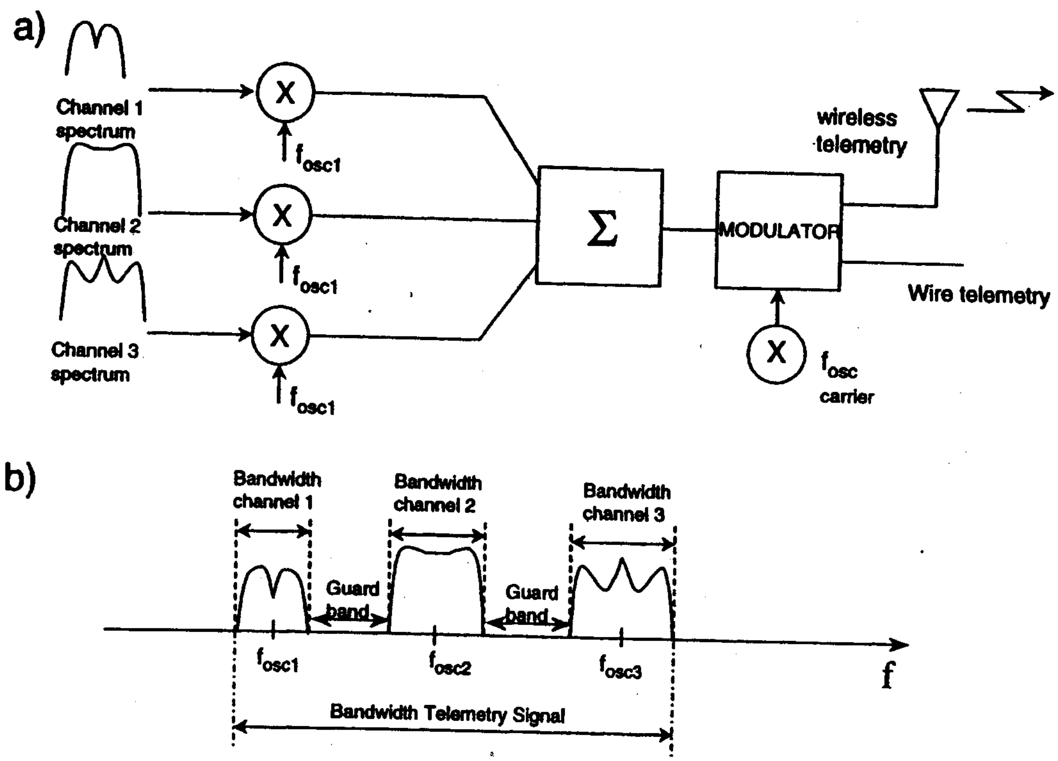

FIGURE 87.5 The different channels in an FDM system (a) are allocated at different subcarrier frequencies producing a composite signal shown in (b) that is later modulated by an RF frequency according to the transmission channel used. The guard bands limit the closeness of contiguous channels to avoid intermodulation and cross talk.

TABLE 87.1 Frequency Bands Allocated for Telemetry

Frequency band, MHz	Uses	Notes
72–76	Biotelemetry	Low power devices; restricted by Part 15 of FCC rules
88–108	Educational	Four frequencies in this band; part 90 of FCC rules
154	Industry	Band in TV channels 7–13
174–216	Biotelemetry	Low-power operations restricted to hospitals
216–222	Multiple	BW < 200 kHz
450–470	General	Telemetry as secondary basis; limited to 2 W of RF
467	Industry	Business band; limited to 2 W of RF
458–468	Biotelemetry	Band in TV channels 21–29
512–566	Biotelemetry	Low-power operations restricted to hospitals
1427–1435	Fixed	Uses in land mobile services (telemetering and telecommand)
1435–1535	Aeronautical	
2200–2290	Mobile	

to TV-broadcasting systems by cable) or, more commonly, into an antenna in the case of wireless telemetry systems. In wireless telemetry, the frequency of the carrier cannot be chosen arbitrarily, but is chosen in accordance with international agreements on the use of the electromagnetic spectrum. In the U.S., the Federal Communications Commission (FCC) is the body that regulates the allocation of frequencies for different communication services. Table 87.1 shows the most common telemetry frequency bands and their intended use. Table 87.1 is for informational purposes only, and it is not a comprehensive guide to telemetry frequencies. To find the allowed telemetry frequencies for a specific application, the maximum power allowed, and other limitations, the reader should consult the applicable FCC documents [7,8].

The allocation of bands is a process subject to change. For example, in October 1997 the FCC assigned some of the TV channel bands for patient telemetry inside hospitals, with restricted power. The FCC publishes all changes that affect frequency bands or other technical characteristics for telemetry.

At the receiver end, the carrier demodulator detects and recovers the composite base-band signal. The next step is to separate each of the subcarriers, by feeding the signal into a bank of parallel passband filters. Each channel is further demodulated, recovering the information from the transducer. The main practical problem of FDM systems is the cross talk between channels. Cross talk appears due to the nonlinearities of the electronic devices, which originates when the signal for one channel partially modulates another subcarrier in addition to the one assigned to that channel. Cross talk also originates when the spectra for two adjacent channels overlap. To avoid this effect, the subcarriers have to be chosen so that there is a separation (guard band) between the spectra of two contiguous channels. By increasing the guard band, the possibility of cross talk decreases, but the effective bandwidth also increases. The effective bandwidth equals the sum of the bandwidth of all channels, plus the sum of all the guard bands.

There are three alternative methods for each of the two modulation processes: the modulation of the measurement channel signals and the modulation of the composite signal. These methods are amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). The usual combinations are FM/FM, FM/PM, or AM/FM [6]. Here, we will analyze only on the subcarrier modulation schemes, while the modulation for the RF signal is analyzed in Chapter 81.

Subcarrier Modulation Schemes for Frequency Division Multiplexing

Subcarrier Modulation of Amplitude.

In an AM subcarrier modulation scheme, the amplitude of a particular subcarrier signal is changed according to the value of the measured channel assigned to that frequency. The resulting AM signal is given by

$$v(t) = A_c\left[1 + m(t)\right]\cos(\omega_c t)$$

where A_c is the amplitude of the carrier, $m(t)$ the modulating signal, and ω_c the frequency of the carrier.

The advantage of this type of modulation is the simplicity of the circuits that perform the modulation and the circuits required for the demodulation, in order to recover the modulating signal that carries the desired information. The percentage of modulation denotes the extent to which a carrier has been amplitude modulated. Assuming for simplicity that the modulating signal is sinusoidal of frequency ω_m, such as

$$m(t) = m \times \cos(\omega_m t)$$

the percentage of modulation (P) can be found as

$$P = m \times 100\,(\%)$$

In a more general way, the percentage of modulation (P) is expressed as

$$\frac{P}{100\%} = \frac{A_{c(\max)} - A_{c(\min)}}{2A_c}$$

where $A_{c(\max)}$ and $A_{c(\min)}$ are the maximum and minimum values that the carrier signal achieves.

Figure 87.6 shows the spectrum of an amplitude-modulated signal, assuming that the modulating signal is a band-limited, nonperiodic signal of finite energy. Figure 87.6 shows that it consists of two

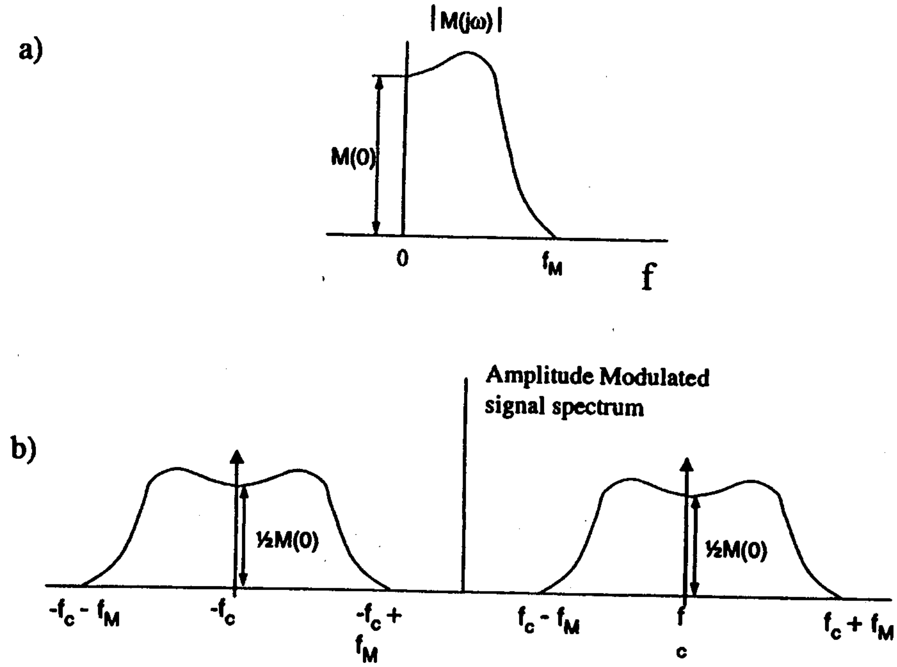

FIGURE 87.6 Resulting spectrum after amplitude modulation of a signal shown in (a). The resulting spectrum has doubled the required bandwidth, while only 0.25 of the total power is used in transmitting the desired information.

sidebands that are symmetrical in reference to the subcarrier. Figure 87.6 shows the main disadvantages of AM schemes. First, the bandwidth of the modulated channel is two times the bandwidth of the modulating signal, due to the two similar sidebands that appear. This results in an inefficient use of the spectrum. Second, the analysis of power for each of the components in Figure 87.6 shows that at least 50% of the transmitted power is used in transmitting the subcarrier, which is independent of the measured signal, as it does not contain any information. The remaining power is split between the two sidebands, which results in a maximum efficiency that it is theoretically possible to achieve of below 25%. The third main problem of AM is the possibility of overmodulation, which occurs when $m > 1$. Once a signal is overmodulated, it is not possible to recover the modulating signal with the simple circuits that are widely used for AM telemetry transmission.

The limitations of AM subcarrier modulation can be overcome using more efficient modulation techniques, such as double sideband (DSB), single sideband (SSB), and compatible single sideband (CSBB), which are also considered AM techniques. However, the complexity of these modulation systems and the cost associated with systems capable of recovering subcarrier signals modulated this way cause these not to be used in most commercial telemetry systems. Most of the available systems that use AM subcarrier techniques, use the traditional AM that has been described here, because its simplicity overcomes the possible problems of its use.

Subcarrier Modulation of Frequency.

FM (or PM) is by far the most-used subcarrier modulation scheme in FDM telemetry systems. These angle modulations are inherently nonlinear, in contrast to AM. Angle modulation can be expressed as

$$v(t) = A\cos\left[\omega_c t + \phi(t)\right]$$

where $\phi(t)$ is the modulating signal, that is, the signal from the transducers after conditioning.

It is then possible to calculate the value of the instantaneous frequency as

$$f = \frac{1}{2\pi}\frac{d}{dt}\left[\omega_c t + \phi(t)\right] = \frac{\omega_c}{2\pi} + \frac{d}{dt}\phi(t)$$

This equation shows how the signal $v(t)$ is modulated in frequency. We can analyze two parameters that can be derived from the previous equations: frequency deviation and modulation index. Frequency deviation (f_m) is the maximum departure of the instantaneous frequency from the carrier frequency. The modulation index (β) is the maximum phase deviation. The following equations show how these parameters are related. The value of the instantaneous frequency (f) is [9]

$$f = \frac{\omega_c}{2\pi} + \frac{\beta\,\omega_m}{2\pi}\cos\left(\omega_m t\right) = f_c + \beta\, f_m \cos\left(\omega_m t\right)$$

The maximum frequency deviation is Δf and is given by

$$\Delta f = \beta f_m$$

Therefore, we can write the equation for the frequency modulated signal as

$$v(t) = A\cos\left[\omega_c t + \frac{\Delta f}{f_m}\sin\left(\omega_m t\right)\right]$$

The previous equation shows that the instantaneous frequency, f, lies in the range $f_c \pm \Delta f$. However, it does not mean that all the spectral components lie in this range. The spectrum of an angle-modulated waveform cannot be written as a simple equation. In the most simple case, when the modulating signal is a sinusoidal signal, a practical rule states that the bandwidth of an FM signal is twice the sum of the maximum frequency deviation and the modulating frequency. For modulating signals commonly found in measuring systems, the bandwidth is dependent upon the modulation index; that is, as the bandwidth allocated for each channel is limited, the modulation index will also be limited.

Frequency Division Multiplexing Telemetry Standards

IRIG Standard 106-96 is the most used for military and commercial telemetry, data acquisition, and recording systems by government and industry worldwide [10]. It recognizes two types of formats for FM in FDM systems: proportional-bandwidth modulation (PBW) and constant-bandwidth modulation (CBW). It also allows the combination of PBW and CBW channels. In PBW, the bandwidth for a channel is proportional to the subcarrier frequency. The standard recognizes three classes of subcarrier deviations: 7.5, 15, and 30%. There are 25 PBW channels with a deviation frequency of 7.5%, numbered 1 to 25. The lowest channel has a central frequency of 400 Hz, which means that the lower deviation frequency is 370 Hz and the upper deviation frequency is 430 Hz. The highest channel (channel 25) has a center frequency of 560,000 Hz (deviation from 518,000 to 602,000 Hz). The center frequencies have been chosen so that the ratio between the upper deviation limit for a given channel and the lower deviation limit for the next channel is around 1.2. There are 12 PBW channels with a deviation frequency of 15%, identified as A, B, ... L. The center frequency for the lowest channel is 22,000 Hz (deviation from 18,700 Hz to 25,300 Hz), while the center frequency for the highest channel is 560,000 Hz (476,000 to 644,000 Hz), with a ratio for the center frequencies of adjacent channels being about 1.3. There are also 12 PBW channels for a deviation frequency of 30%, labeled from AA, BB, ... to LL. The center frequency for these channels is the same as that for the 15% channels.

TABLE 87.2 Characteristics of Constant Bandwidth (CBW) Channels for FDM

Channel Denomination	Frequency Deviation, kHz	Lowest Channel Center Frequency, kHz	Highest Channel Center Frequency, kHz	No. of Channels	Separation between Channels, kHz
A	±2	8	176	22	8
B	±4	16	352	22	16
C	±8	32	704	22	32
D	±16	64	1408	22	64
E	±32	128	2816	22	128
F	±64	256	3840	15	256
G	±128	512	3584	7	512
H	±256	1024	3072	4	1024

CBW channels keep the bandwidth constant and independent of its carrier frequency. There are eight possible maximum subcarrier frequency deviations labeled A (for 2 kHz deviation) to H (for 256 kHz deviation). The deviation frequency doubles from one group to the next. There are 22 A-channels, whose center frequency range from 8 to 176 kHz. The separation between adjacent channels is a constant of 8 kHz. Table 87.2 shows a summary of the characteristics of CBW channels.

IRIG Standard 106-96 gives in its appendix criteria for the use of the FDM Standards. It focuses on the limits, most of the time dependent on the hardware used, and performance trade-offs such as data accuracy for data bandwidth that may be required in the implementation of the system. The subcarrier deviation ratio determines the SNR for a channel. As a rule of thumb, the SNR varies as the three-halves power of the subcarrier deviation ratio. On the other hand, the number of subcarrier channels that can be used simultaneously to modulate an RF carrier is limited by the channel bandwidth of the RF carrier as well as considerations of SNR. Given a limited RF bandwidth, as more channels are added to the FDM system, it is necessary to reduce the deviation ratio for each channel, which reduces the SNR for each channel. It is then very important to evaluate the acceptable trade-off between the number of subcarrier channels and the acceptable SNR values. A general equation that might be used to estimate the thermal noise performance of an FM/FM channel is the following [11]:

$$\left(\frac{S}{N}\right)_d = \left(\frac{S}{N}\right)_c \left(\frac{3}{4}\right)^{1/2} \left(\frac{B_c}{F_{ud}}\right)\left(\frac{f_{dc}}{f_s}\right)\left(\frac{f_{ds}}{F_{ud}}\right)$$

where $(S/N)_d$ represents the SNR at the discriminator output, $(S/N)_c$ represents the SNR of the receiver, B_c is the intermediate-frequency bandwidth of the receiver, F_{ud} is the subcarrier discriminator output filter (at –3 dB), f_s is the subcarrier center frequency, f_{dc} is the carrier peak deviation for the subcarrier considered, and f_{ds} is the subcarrier peak deviation.

According to the Standard, the FM/FM composite FDM signal that is used to modulate an RF carrier can be of PBW format, CBW format, or a combination of both, with the only limitation that the guard bands between the channels used in the mixed format are equal or greater than the guard bands for the same channels in an unmixed format.

Time Division Multiplexing

TDM is a transmission technique that divides the time into different slots, and assigns one slot to each measurement channel. In TDM, all the transmission bandwidth is assigned entirely to each measurement channel during a fraction of the time. After the signals from the measurement channels have been low-pass filtered, they are sequentially sampled by a digital switch that samples all the measurement channels in a period of time (T) that complies with the Nyquist criteria. Figure 87.7a shows a basic block diagram for an FDM system. The output of the sampler is a train of AM pulses that contains the individual samples for the channels framed periodically, as is shown in Figure 87.7b. Finally, the composite signal modulates an RF carrier. The set of samples from each one of the input channels is called a frame. For

f els	Separation between Channels, kHz
	8
	16
	32
	64
	128
	256
	512
	1024

iency. There are eight
ı) to H (for 256 kHz
22 A-channels, whose
ınels is a constant of

ıdards. It focuses on
ıde-offs such as data
tem. The subcarrier
s as the three-halves
er channels that can
th of the RF carrier
: added to the FDM
s the SNR for each
ımber of subcarrier
timate the thermal

JR of the receiver,
riminator output
for the subcarrier

late an RF carrier
on that the guard
ırd bands for the

one slot to each
ch measurement
have been low-
ement channels
block diagram
the individual
mposite signal
d a frame. For

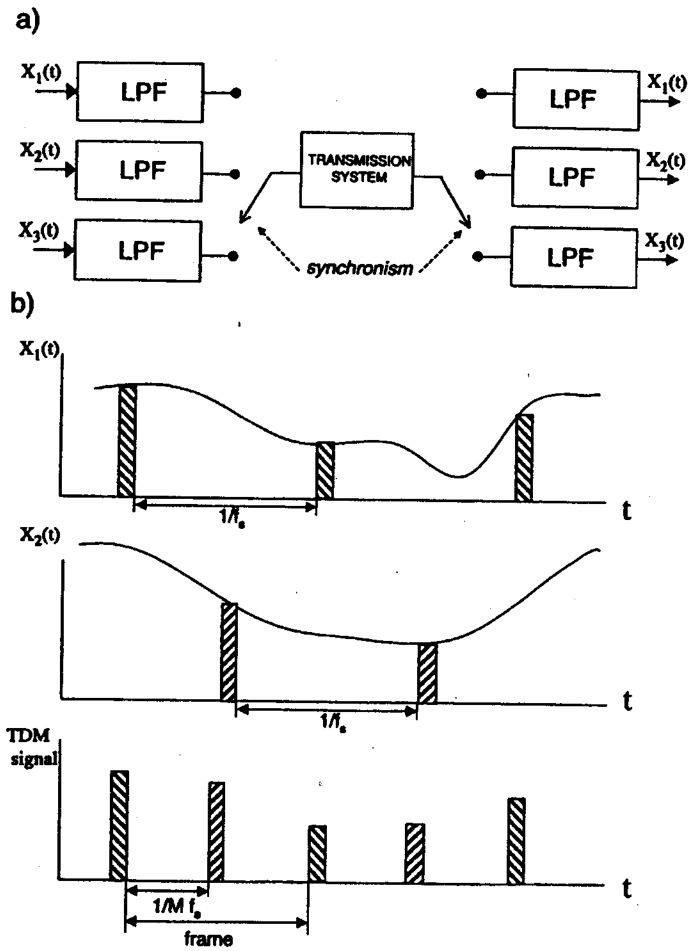

FIGURE 87.7 TDM systems (a) are based on sequentially sampling M different channels at a sampling frequency f_s, and sending the information for each channel sequentially (b). In TDM, the synchronism between transmitter and receiver is critical to recovery of the sampled signal. In this figure, the TDM signal is made of only two channels to increase readability. The blocks labeled LPF represent low-pass filters.

M measurement channels, the period between two consecutive pulses is $T_s/M = 1/Mf_s$, where T_s is the sampling period. The period between samples from the same channel is T_s. At the receiver end, by separating the digital signals into different channels by a synchronized demultipler and by low-pass filtering, it is possible to recover the original signal for each measurement channel.

TDM systems have advantages over FDM systems. First, FDM requires subcarrier modulators and demodulators for each channel, whereas in TDM only one multiplexer and demultiplexer are required. Second, TDM signals are resistant to the error sources that originate cross talk in FDM: nonideal filtering and cross modulation due to nonlinearities. In TDM, the separation between channels depends on the sampling system. However, because it is impossible in practice to produce perfectly square pulses, their rise and fall times are different from zero. It is then necessary to provide guard time between pulses, similar to the band guards in FDM systems. Cross talk in TDM can be easily estimated assuming that the pulse decay is exponential with a time constant (τ) approximately equal to

$$\tau = \frac{1}{2\pi B}$$

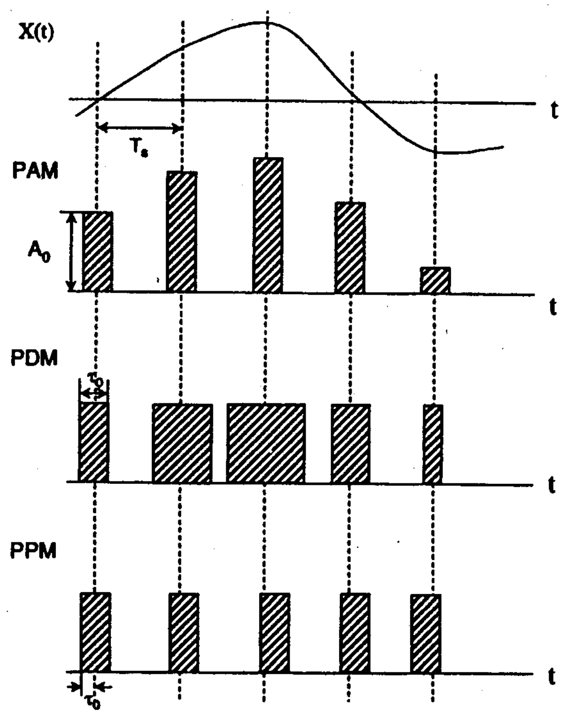

FIGURE 87.8 Different analog modulation schemes used in TDM. The variations in amplitude of the signal $x(t)$ are transmitted as amplitude variations of pulses (PAM), duration changes of pulses (PDM), or changes in the relative position of the pulses (PPM). In all the cases, the level 0 is transmitted by a pulse whose amplitude (A_0), duration (τ_0), or relative position (τ_0) is different from 0.

where B is the –3 dB channel bandwidth. The cross talk (k) between channels can be approximated as

$$k = -54.5T_g \left(\text{dB}\right)$$

where T_g is the minimum time separation between channels, called guard time.

A common situation in measurement systems occurs when the M signals that need to be measured have very different speeds. The channel sampling rate is determined by the fastest signal, thus needing an M-input multiplexer capable of handling signals at that sampling frequency. A convenient solution is to feed several slow signals into one multiplexer, then combine its output with the fast signal in a second multiplexer [6].

Analog Subcarrier Modulation Schemes for Time Division Multiplexing

In analog modulation for subcarriers the signal that results after the multiplexing and sampling process modulates a train of pulses. The most common methods for analog subcarrier modulation are pulse amplitude modulation (PAM), pulse duration modulation (PDM), and pulse position modulation (PPM). Figure 87.8 illustrates these three modulation schemes, where the pulses are shown square for simplicity. In analog modulation, the parameter that is modulated (amplitude, duration, or relative position) changes proportionally to the amplitude of the sampled signal. However, in PAM and PDM the values have an offset, so that when the value of the sample is zero, the pulse amplitude or the pulse width is different from zero. The reason for these offsets is to maintain the rate of the train of pulses constant, which is very important for synchronization purposes. The common characteristics of the different analog modulation schemes for pulses in TDM are (1) a modulated signal spectrum with a large low-frequency content, especially close to the sampling frequency; (2) the need to avoid overlaying

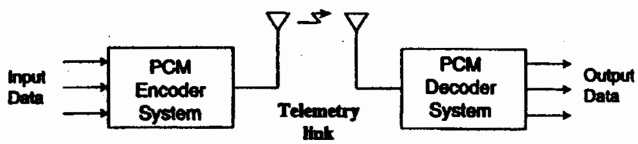

FIGURE 87.9 Block diagram showing a basic PCM link for telemetry.

between consecutive pulses in order to conserve the modulation parameters; and (3) the possibility of reconstructing the original samples from the modulated signal through low-pass filtering after demultiplexing. The reduction of noise depends on the bandwidth of the modulated signal, with this being the principal design criterion.

Pulse Amplitude Modulation.
PAM waveforms are made of unipolar, nonrectangular pulses whose amplitudes are proportional to the values of the samples. It is possible to define the modulation index using similar criteria as in analog AM. Similarly, in PAM the modulation index is limited to values less than 1.

Pulse Duration Modulation.
PDM is made of unipolar, rectangular pulses whose durations or widths depend on the values of the samples. The period between the center of two consecutive pulses is constant. The analysis of the resulting spectrum shows that it is possible to reconstruct the samples by low-pass filtering [9].

Pulse Position Modulation.
PPM is closely related to PDM, as PPM can be generated through PDM. In PPM the information resides on the time location of the pulses rather than in the pulses by themselves. It is then possible to transmit very narrow pulses to reduce the energy needed; this energy reduction is the most important advantage of PPM.

Pulse Code Modulation for Time Division Multiplexing.
All the previously analyzed subcarrier modulation schemes in telemetry systems are based on an analog signal that modulates either an analog carrier or a train of pulses. Pulse code modulation (PCM) is different: it is a digital modulation in which the measured signal is represented by a group of codified digital pulses. Two variations of PCM that are also often used are delta modulation (DM) and differential pulse code modulation (DPCM). In analog modulation schemes, the modulating signal from the transducer can take any value between the limits. If noise alters the modulating signal, it is impossible to decide its real value. Instead, if not all the values in the modulating signal are allowed, and the separation between the allowed levels is higher than the expected noise values, it is then possible to decide which were the values sent by the transmitter. This immunity against noise makes PCM systems one of the preferred alternatives for telemetry. Figure 87.9 shows the basic elements of a PCM telemetry system. A PCM encoder (or PCM commutator) converts the input data into a serial data format suitable for transmission through lines by wireless techniques. At the receiving end, a PCM decoder (or PCM decommutator) converts the serial data back into individual output data signals. PCM systems transmit data as a serial stream of digital words. The PCM encoder samples the input data and inserts the data words into a PCM frame. Words are assigned specific locations in the PCM frame, so the decoder can recover the data samples corresponding to each input signal. The simplest PCM frame consists of a frame synchronization word followed by a string of data words. The frame repeats continually to provide new data samples as the input data change. Frame synchronization enables the PCM decoder to locate the start of each frame easily.

Pulse Code Modulation Telemetry Standards.
IRIG Standard 106-96 also defines the characteristics of PCM transmission for telemetry purposes, in particular, the pulse train structure and system design characteristics. The PCM formats are divided into two classes for Standards purposes: class I and class II. The simpler types are class I, whereas the more

TABLE 87.3 Summary of the Most Relevant PCM Specifications According to IRIG 106-96

Specification	Class I	Class II
Class format support	Class I (simple formats) supported on all ranges	Class II (complex formats) requires concurrence of range involved
Primary bit representation (PCM codes)	NRZ-L, NRZ-M, NRZ-S, RNRZ-L, BiØ-L, BiØ-M, BiØ-S	Same as class II
Bit rate	10 bps to 5 Mbps	10 bps to > 5 Mbps
Bit rate accuracy and stability	0.1%	Same as class I
Bit jitter	0.1 bit	Same as class I
Bit numbering	MSB = bit number 1	Same as class I
Word length	4 to 16 bits	4 to 64 bits
Fragmented words	Not allowed	Up to 8 segments each; all segments in the same minor frame
Minor frame length	<8192 bits or <1024 words (includes synchro)	<16,384 bits (includes synchro
Major frame length	<256 minor frames	Same as class I
Minor frame numbering	First minor frame in each major frames in number 1	Same as class I
Format change	Not allowed	Frame structure is specified by frame format identification (FFI) word in every minor frame

complex types are class II. Some of the characteristics of class II systems are bit rates greater than 5 Mbit/s, word lengths in excess of 16 bits, fragmented words, unevenly spaced subcommutation, format changes, tagged data formats, asynchronous data transmission, and merger of multiple format types, among others. Table 87.3 provides a brief summary of relevant PCM specifications. Readers interested in the detailed specifications and descriptions should refer to Chapter 4 of the IRIG 106-96 Standard [10].

The following PCM codes, shown in Figure 87.10, are recognized by the IRG Standards: NRZ-L (nonreturn to zero — level), NRZ-M (nonreturn to zero — mark), NRZ-S (nonreturn to zero — space), BiØ-L (Biphase — level), BiØ-M (Bi-Phase — mark) and BiØ-S (Biphase — space). The Standard also recommends that the transmitted bit stream be continuous and contain sufficient transitions to ensure bit acquisition and continued bit synchronization. Bit rates should be at least 10 bits/s. If the bit rate is above 5 Mbit/s, the PCM system is classified as class II. In reference to the word formats, the Standard defines a fixed format as one that does not change during transmissions with regard to the frame structure, word length or location, commutation sequence, sample interval, or measurement list. Individual words may vary in length from 4 bits to not more than 16 bits in class I and not more than 64 bits in class II. Fragmented words, defined as a word divided into not more than eight segments and placed in various locations within a minor frame, are only allowed in class II. All word segments used to form a data word are constrained to the boundaries of a single minor frame. The Frame Structure allowed by the Standards for PCM telemetry specifies that data are formatted into fixed frame lengths, that contain a fixed number of equal-duration bit intervals. A minor frame is defined as the data structure in time sequence from the beginning of a minor frame synchronization pattern to the beginning of the next minor frame synchronization pattern. The minor frame length is the number of bit intervals from the beginning of the frame synchronization pattern to the beginning of the next synchronization pattern. The maximum length of a minor frame will not exceed 8192 bits nor 1024 words in class I and will not exceed 16,384 bits in class II. Minor frames consist of the synchronization pattern, data words, and subframe synchronization words if they are used. The Standard allows the use of words of different length if they are multiplexed in a single minor frame. Figure 87.11 shows a graphical representation of a PCM frame structure. Major frames contain the number of minor frames required to include one sample of every parameter in the format. Their length is defined as minor frame length multiplied by the number of minor frames contained in the major frame. The maximum number of minor frames per major frame is limited to 256.

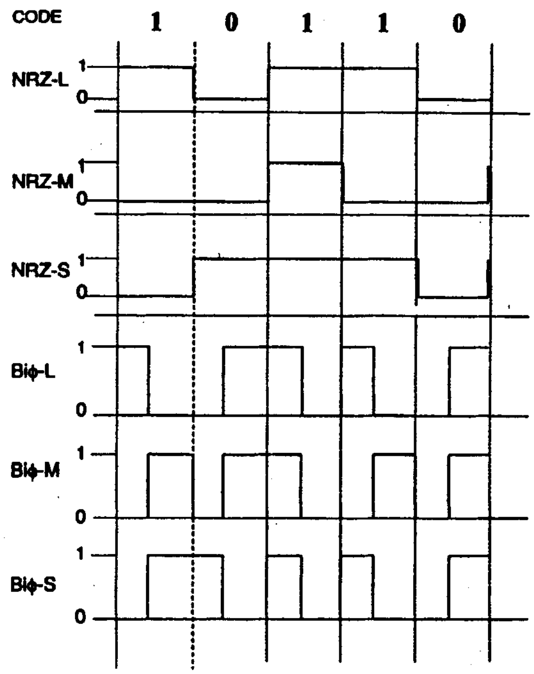

FIGURE 87.10 Different PCM codes. All lower levels in NRZ use a value different from zero. In biphase codes the information resides in the transitions rather than in the levels. In NRZ-L, a 1 is represented by the highest level, while a 0 is represented by a lower level. In NRZ-M, a 1 is represented by a change in level, while a 0 is represented by no change in level. In NRZ-S, a 1is represented by no change of level, while a 0 is represented by a change of level. In Biϕ-L, a 1 is represented by a transition to the lower level, while a 0 is represented by a transition to the higher level. In Biϕ-M, the 1 is represented by no change of level at the beginning of the bit period, while the 0 is represented by a change of level at the beginning of the bit period. In Biϕ-S, a 1 is represented by changing the level at the beginning of the bit period, while the 0 is represented by no change of level at the beginning of the bit period.

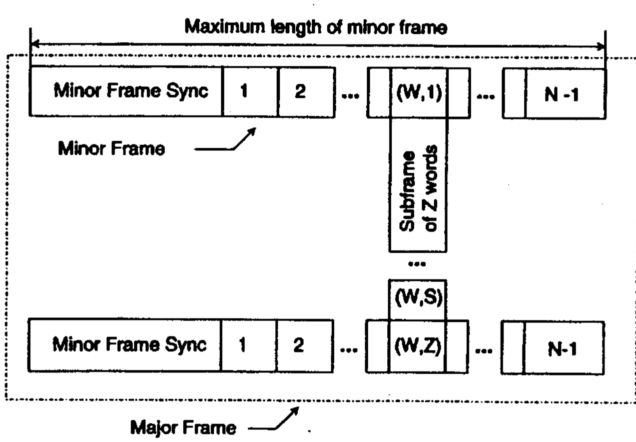

FIGURE 87.11 Structure of a PCM Frame. The maximum length of a minor frame is 8192 bits or 512 for class I and 16,284 bits for class II. A major frame contains $N \times Z$ words, where Z is the number of words in the maximum subframe, and N is the number of words in the minor frame. Regardless of its length, the minor frame synchronism is considered as one word. W is the word position in the minor frame, while S is the word position in the subframe.

Appendix C in the 106-96 IRIG Standard gives recommendations for maximal transmission efficiency in PCM telemetry. The intermediate-frequency (IF) bandwidth for PCM telemetry data receivers should be selected so that 90 to 99% of the transmitted power spectrum is within the receiver 3-dB bandwidth. The IF also has effects on the bit error probability (BEP) according to the following equation for NRZ-L PCM/FM [10]:

$$\mathrm{BEP} = 0.5e^{(k\mathrm{SNR})}$$

where $k \approx -0.7$ for IF bandwidth equal to bit rate
$k \approx -0.65$ for IF bandwidth equal to 1.2 times bit rate
$k \approx -0.55$ for IF bandwidth equal to 1.5 times bit rate

Other data codes and modulation techniques have different BEP vs. SNR performance characteristics, but in any case they will have similar trends.

The Standard also specifies the recommended frame synchronization patterns for general use in PCM telemetry. There are different lengths for synchronization patterns, but in all of them the 111 is the first bit sequence transmitted. The patterns for lengths 16 to 30 were selected in order to minimize the probability of false synchronization over the entire pattern overlap portion of the ground station frame synchronization [12]. The spectral density (S) for the NRZ and BiØ codes are

NRZ Codes $$S = \frac{\sin^2(\pi f T)}{(\pi f T)^2}$$

Biphase Codes $$S = \frac{\sin^4(\pi f T/2)}{(\pi f T/2)^4}$$

The calculation of spectral densities allows the determination of the BEP for the previous type of codes assuming perfect bit synchronization. These calculations show that for the same SNR, the lowest BEP is achieved for NRZ-L and Bi codes, followed by NRZ and BiØ mark and space codes and finally for random NRZ-L codes (RNRZ-L).

Telemetry data are usually recorded onto magnetic tape for later analysis. When recording PCM data, it is important to ensure that the tape recorder provides sufficient frequency response to capture and reproduce the PCM signal. Useful rules to calculate the maximum bit rate for various PCM codes specify that for NRZ and RNRZ codes the maximum bit rate is 1.4 times the tape recorder frequency response while for all biphase codes, the maximum rate is 0.7 times the tape recorder response. To limit the transmission bandwidth that PCM creates because it is a digital signal with sharp transitions, the PCM signal is usually passed through a premodulation filter before it is fed into the transmitter input. The filter cutoff frequency can be calculated as 0.7 times the PCM bit rate for NRZ and RNRZ codes and 1.4 times the PCM bit rate for all biphase codes.

Defining Terms

Bandwidth: The range of frequencies occupied by a signal.

Carrier: A frequency that is modulated by a signal containing information.

Channel: A subcarrier that carries information.

Constant bandwidth (CBW) channel: A channel whose bandwidth is independent of its carrier frequency.

Deviation ratio: The ratio of the maximum carrier frequency deviation to the maximum data frequency deviation.

Frequency deviation: The difference between the center frequency of a carrier and its upper or lower deviation limit.

Frequency division multiplexing (FDM): A composite signal consisting of a group of subcarriers arranged so that their frequencies do not overlap or interfere with each other.

Frequency response: The highest data frequency that can be carried by the channel.

IRIG: Inter-Range Instrumentation Group of the Range Commanders Council (RCC).

Proportional bandwidth (PBW) channel: A channel whose bandwidth is proportional to its carrier frequency.

Remote switching: Telemetry consisting only of yes/no or on/off orders.

Signaling: Telemetry consisting of binary information.

Subcarrier: A carrier combined with other carriers to create a composite signal.

Subcarrier bandwidth: The difference between the upper and lower frequencies of a modulated carrier.

References

1. B.P. Dagarin, R.K. Taenaka, and E.J. Stofel, Galileo probe battery system, *IEEE Aerospace Electron. Syst. Mag.*, 11 (6): 6–13, 1996.
2. M.W. Pollack, Communications-based signaling: advanced capability for mainline railroads, *IEEE Aerospace Electron. Syst. Mag.*, 11 (11): 13–18, 1996.
3. M.C. Shults, R.K. Rhodes, S.J. Updike, B.J. Gilligan, and W.N. Reining, A telemetry-instrumentation system for monitoring multiple subcutaneously implanted glucose sensors, *IEEE Trans. Biomed. Eng.* 41: 937–942, 1994.
4. M. Rezazadeh and N.E. Evans, Multichannel physiological monitor plus simultaneous full-duplex speech channel using a dial-up telephone line, *IEEE Trans. Biomed. Eng.*, 37: 428–432, 1990.
5. Y.S. Trisno, P. Hsieh, and D. Wobschall, Optical pulses powered signal telemetry system for sensor network application, *IEEE Trans. Instrum. Meas.*, 39: 225–229, 1990.
6. R. Pallás-Areny and J.G. Webster, *Sensors and Signal Conditioning*, New York: John Wiley & Sons, 1991, 352–379.
7. Part 90: Private land mobile radio services. Subpart J: Non voice and other specialized operations (Sec. 90.238: Telemetry), *Code of Federal Regulations*, Title 47, Telecommunication. Chapter I: Federal Communications Commission.
8. Part 15: Radio frequency devices, *Code of Federal Regulations*, Title 47, Telecommunication. Chapter I: Federal Communications Commission.
9. H. Taub and D. L. Schilling, *Principles of Communication Systems*, 2nd ed., New York: McGraw-Hill, 1986.
10. IRIG, *Telemetry Standard IRIG 106-96*, Range Commander Council, U.S. Army White Sands Missile Range, 1996.
11. K. M. Uglow, Noise and bandwidth in FM/FM radio telemetry, *IRE Trans. Telemetry Remote Control*, 19–22, 1957.
12. J.L. Maury and J. Styles, Development of optimum frame synchronization codes for Goddard space flight center PCM telemetry standards, *Proc. Natl. Telemetering Conf.*, 1964.

Further Information

E. H. Higman, Pneumatic instrumentation, in B.E. Noltingk, Ed., *Instrument Technology: Instrumentation Systems*, Vol. 4, London: Butterworths, 1987.

C.H. Hoeppner, Telemetry, in R. C. Dorf, Ed., *The Electrical Engineering Handbook*, Boca Raton, FL: CRC Press, 1993.

R.S. Mackay, *Bio-Medical Telemetry*, 2nd ed., Piscataway, NJ: IEEE Press, 1993.

Telemetry Group, *Test Methods for Telemetry Systems and Subsystems. Document 118-97*, Range Commanders Council, U.S. Army White Sands Missile Range, 1997.

Telemetry Group, *Telemetry Applications Handbook Document 119-88*, Range Commanders Council, U.S. Army White Sands Missile Range, 1988.

88

Sensor Networks and Communication

Robert M. Crovella
NVIDIA Corporation

88.1 Introduction

What Is a Communication Network?

A communication network provides a system by which multiple users may share a single communication path (or medium) to exchange information. The telephone system is an example of a system containing many communication networks, which can be considered to be a single communication network as an abstract example. Communication networks are commonly used in various industries and applications to provide an economical means to allow multiple, geographically separated users to exchange information.

Ordinary Sensors vs. Networked Sensors

A definition of the function of a sensor is to map or convert one measured variable (e.g., spatial, mechanical, electromagnetic, etc.) into another — usually electric — variable or signal. This signal may then be passed to a measurement or processing system for capture and analysis, or as a direct input to some controlled process. In this case, the measured variable is represented as an electric signal. This signal must be handled individually by the measurement system, and it may also be subject to corruption from a variety of sources, such as electromagnetic interference in the case of an electric signal.

0-8493-8347-1/99/$0.00+$.50

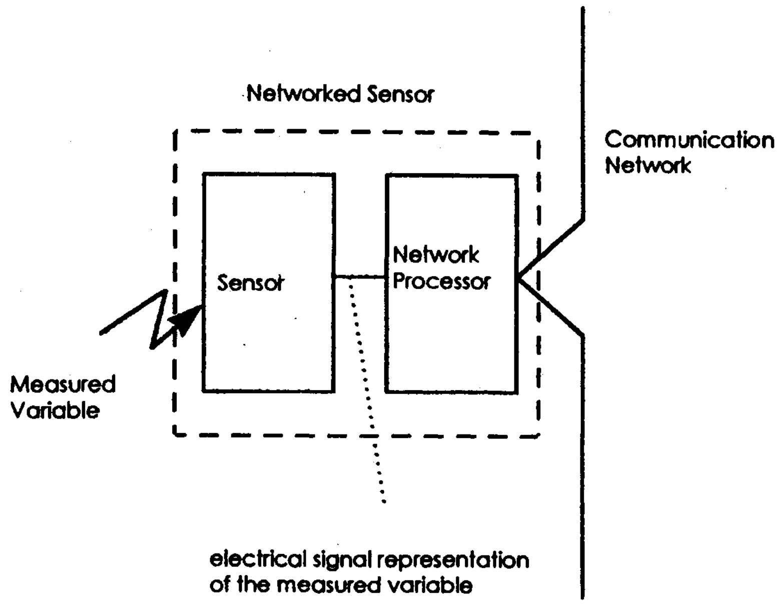

FIGURE 88.1 A networked sensor is an ordinary sensor with network communication components added.

In applications where a number of sensing devices are needed, and/or where the sensing devices are distributed geographically (or are distant from the measurement and analysis system), the application designer may wish to use a communication network to transmit sensor data from the measurement point to the measurement and analysis system. Such applications typically involve some sort of digital computing machinery at the measurement and analysis point, or as part of the control system. Figure 88.1 depicts a representative block diagram of a sensor/network system showing the relationship of the various components.

Networked sensors can be distinguished into two components: those performing the measurement function, and those components performing the communication function. In some cases, these two functions may be designed as a single unit, such that the "sensor" intrinsically includes communication capability. In other cases, an ordinary sensor may be connected to a conversion unit, which converts the output signal of the ordinary sensor into a form suitable for the network, and manages the delivery of this information on the network.

Why Use Networked Sensors?

Network communication combined with sensor technology can provide several benefits to an application, as well as to the sensor designer. The most obvious benefit of a network is the simplification of the wiring for the transmission of the signals from one place to another. For a system containing N users, the number of wires or cables T required to individually connect each user with each other user is given by Equation 88.1:

$$T = 2^{(N-1)} - 1 \tag{88.1}$$

assuming each wire or cable can carry information in both directions between the two users connected by that cable. For more than a few users, the number of cables required (T) to provide an individual connection between each pair of users is large. Sensors are often connected to a central measurement and analysis system, and may only need to communicate with the central system. In this case, the number of individual wires or cables needed is equal to the number of sensors ($N-1$). Even with this smaller number of cables, the wiring for a large number of sensors in some applications can be quite complex. A network may be able to reduce the total number of cables required to a much smaller number. In fact, in a sensor network, all of the sensors and the central measurement and analysis system can be connected to a single cable.

An indirect benefit of networking may be in its handling of the sensor signal. Because most modern networks are digital in nature, an analog sensor signal typically must be digitized before it can be transmitted on a network. With a networked sensor, the digitization will typically be carried out by circuitry in relatively close proximity to the sensor. As a result, the analog signal will have traveled a short distance before being converted to a digital signal. This can be a benefit in two ways. The first is that the analog signal will not suffer as much attenuation or degradation due to electric losses associated with carrying a signal over a great distance. The second is that once in digital form, the "signal" can be made relatively immune to the effects of distortion or degradation due to electromagnetic interference (EMI). Although digital transmission of signals is still subject to EMI, modern protocols and transmission systems can be designed to be very robust, using signaling that is resistant to EMI as well as using error control techniques. As a result, the effect of attenuation and disturbances can be essentially eliminated by digital transmission of the signal.

Another benefit of networking is the ability to communicate a much wider range of information — in both directions — when compared with a single cable carrying a sensor signal. With many modern networks suitable for networked sensing applications, a microprocessor is used at the sensor to manage the handling of the sensor signal and its transmission on the network. But there is generally no need to limit the microprocessor to this one function alone. The combination of the network and the microprocessor provides a platform upon which many additional functions and features can be incorporated into the networked sensor. For example, the signal of a sensor may need a certain calibration or correction function applied to it before it can be used in calculations. It may be beneficial to load into the networked sensor (through the network) a set of correction parameters or coefficients, and then have the microprocessor correct or calibrate the output of the sensor before transmitting it to the network. Sensors can be easily designed to have multiple sensing functions, such as temperature and pressure, for example. Each signal can be handled separately and transmitted separately on the network, with no need for additional connections. Sensors may be designed to store certain types of information, such as the name of the manufacturer, or certain calibration parameters determined by the manufacturer at the time of manufacture. This information can then be read out over the network and used for a variety of purposes. A sensor can even be designed to have "intelligent" functions, such as the ability to sense its environment and determine when certain parameters have been exceeded (such as operating temperature range), or report a special message containing an "alarm" when the sensor signal level exceeds a certain threshold. The combination of the network and the microprocessor leads to an endless variety of functions and features that can be added to the basic sensor technology.

Potential Problems with Networked Sensors

Networked sensors will generally require more complex circuitry than equivalent, nonnetworked sensors. A drawback of analog-to-digital (A/D) conversion and digital transmission of signals is the time and level quantization effect that A/D conversion can have on the analog signal. These effects can be mitigated with modern, high-speed A/D converters (to minimize the effect of time quantization, or the sampling effect) with the ability to convert in high resolution (i.e., using a large number of digital bits to represent the analog signal level). These drawbacks are not unique to networked sensors but rather to digitized sensor values and digital control whether or not it uses a network. Finally, the capacity of the network to carry information (the bandwidth) must be considered in any communication system. Putting a large number of sensors on a single network may overload the information-carrying capability of the network, resulting in queuing delays in the reception of sensor signals and, in some cases, lost data.

88.2 Communication and Networking Concepts

In order to be able to select an appropriate network technology, it is necessary to understand some basic terminology so that the features and capabilities of various networks and technologies can be categorized and compared.

Station

A station represents a single communicating element on a network system. Each user of the network must access the communication capability of the network via a station. Each station will typically have some implementation of the open systems interconnection (OSI) network reference model as the means of utilizing the network system.

Media Access

Media access is the method by which individual stations determine when they are permitted to transmit, or "use" the media. Media access control (MAC) is a function that is usually performed in the data link layer of the OSI reference model. Some well-known methods of media access control include carrier sense multiple access with collision detection (CSMA/CD) and token passing. CSMA/CD systems (such as Ethernet) allow all stations on a network equal access. Each station must "listen" to the network to determine periods of inactivity before transmitting. Any station wishing to use the network may begin transmitting providing the network is inactive when it checks for activity. If multiple stations attempt to transmit simultaneously, a collision occurs. This is detected by all transmitting stations, which all must immediately stop transmitting and each wait a randomly determined period of time, before attempting to use the network again. Controller area network (CAN), for example, uses a variant of CSMA/CD for media access. Token-passing systems have a logical "token" which is exchanged among stations via network messaging. The station that holds the token has permission to transmit. All other stations are only permitted to receive messages. Stations wishing to transmit but not having the token must wait until the station holding the token passes it on. Another commonly used method of media access control is master–slave. In this method, one station on the network (designated the master) is generally in charge of, and originates, all communications. Slaves only respond to the master, and only respond when the master initiates communications with them via sending a message to the slave. Profibus-FMS (see below) is an example of a protocol which uses both token passing (in some cases) and master–slave (in some cases) to control media access.

Bandwidth

Bandwidth may have several different definitions. For digital communication systems, bandwidth describes the capacity of the system to transport digital data from one place to another. This term may be applied to the raw capability of the physical and data link layers to transport message data (*raw bandwidth*, closely related to the bit rate concept) or it may be applied to the effective rate at which user-meaningful information is transported (*effective bandwidth*). The bandwidth of a given system is generally inversely proportional to the worst-case node-to-node distance. The smaller the network span, the higher its bandwidth can be.

Addressing

Addressing is a concept that assigns generally unique identifiers to each station in a network system. This identifier (the address) can then be used by the network for a variety of purposes, including identifying the origin and/or destination of messages, or arbitrating access to a shared communications medium. Another addressing or identifier concept assigns unique identifiers not to stations, but to unique pieces of data or signals that will be carried by the network. Stations then use an identifier according to what type of data they will be transmitting. Many, but not all networking methods require establishment of an explicit address for each network station.

Arbitration

Arbitration is a function closely related to MAC. Arbitration is used by some networks to define the procedure followed when multiple stations wish to use the network simultaneously.

Signaling

Signaling refers to the actual physical (e.g., electrical, optical, or other) representation of data as it is carried on the media. For example, in some networks, data elements may be represented by certain voltage levels or waveforms in the media. In other networks, data elements may be represented by the presence of certain wavelengths of light in the media. The association of all the representable data elements (e.g., 0/1 or on/off) with the corresponding signal representations in the media is the signaling scheme or method. An important signaling method where electric wires are used as the medium is differential signaling. Differential signaling represents a particular data element (1 or 0) as two different states on a pair of wires. Determining the data element requires measuring the voltage difference between the two wires, not the absolute level of the voltage on either wire. Different data elements are then represented by the (signed) voltage difference between the two wires. For example, RS-485 represents a digital 1 data element as a 5-V signal level on the first wire and a 0-V signal level on the second wire, and a digital 0 as a 0-V signal level on the first wire and 5-V signal level on the second wire. One of the principal benefits of differential signaling is that it is possible to determine the data being transmitted without knowing the ground reference potential of the transmitter. This allows the transmitter and receiver to operate reliably, even when they have different ground potentials (within limits), which is a common occurrence in communication systems.

Encoding

Encoding refers to the process of translating user-meaningful information into data elements or groups of data elements to be transported by the network system. A code book refers to the set of all relationships between user-meaningful information and data carried by the network. Encoding may occur at several levels within the OSI reference model, as user-meaningful information is transformed successively until it becomes an actual network message, produced by the data link layer. Decoding is the reverse process, whereby a network message is successively translated back into user-meaningful information.

Modulation

Modulation in a classical sense refers to a signaling technique by which data or information is used to control some combination of the frequency, phase, and/or amplitude of a carrier signal. The carrier signal carries the information to a remote receiver where it will be demodulated to retrieve the information. Modulated network systems are outside the scope of this chapter.

Message

A message is the fundamental, indivisible unit of information which is exchanged between stations. User-meaningful information will be grouped into one or more messages by the OSI network reference model.

Multiplexing

Multiplexing refers to the ability to use the media in a network to carry multiple messages or information streams "simultaneously." Multiplexed systems allow several communication channels to use the same physical wire or media. Each message or information stream may have different sources and destinations.

Multiplexing may be accomplished using a variety of means. Time division multiplexing (TDM) involves breaking access to the media into a series of time quanta. During each time quantum, the media carries a separate message or information stream. The close arrangement of time quanta allows the network media to carry multiple messages "simultaneously." Code division multiplexing (CDM) involves the separation of the code book (see Encoding) into sections. Each section of the code book provides all of the messages that will be used for a particular information stream. Therefore, a particular information stream within the network media is distinguished by all of the messages that belong to the section of the code book for that stream. Frequency division multiplexing (FDM) divides an available bandwidth of a communication channel into several frequency ranges, and assigns one information stream to each frequency range.

Protocols

A protocol is a defined method of information exchange. Protocols typically are defined at several levels within the OSI network reference model, such as at the application layer and at the data link layer. Protocols are used to define how the services provided by a particular layer are to be exercised, and how the results of these services are to be interpreted.

Service

A service represents a specific function or operation that is supported by a particular layer in the OSI network reference model. For example, an application layer service might be provided for the reading of or writing to a data element contained in another device (or station) on the network. This service might make use of a data link layer service which might be provided for supporting the exchange of a message with another device (or station) on the network.

Topology

Topology refers to the physical or geographic layout or arrangement of a network. Certain types of canonical topologies are commonly discussed in the context of networks, such as trunkline/branchline, star (or hub), ring, and daisy chain.

Bit Rate

Bit rate refers to the speed at which binary pieces of information (bits) are transmitted on a particular network. The raw bit rate of a network generally refers to the actual speed of transmission of bits on the network. The effective bit rate — or throughput — generally refers to the speed at which user information is transmitted. This number is less than or equal to the raw bit rate, depending on what percentage of the bits transmitted is used for carrying user information. The bits not carrying user information are overhead, used to carry protocol, timing, or other network information.

Duplex (Half and Full Duplex)

Half duplex refers to a communication system in which a station can either transmit information or receive information, but not both simultaneously. A full duplex network allows a station to transmit information and receive information simultaneously.

Error Control

Many network systems provide mechanisms to control errors. Error control has four aspects: prevention, detection, correction, and isolation. Error prevention may simply be shielding for the media to minimize electromagnetic disturbances, or it may be more complicated, such as signal sampling control to optimize the probability that a signal will be in the correct state when sampled. Error detection generally depends on detecting violations of protocol rules at various network levels, or violations of computed data added to a message for error control purposes. Some examples of error detection techniques are parity and cyclic redundancy check (CRC). Both methods involve the computation of additional bits of information based on the data that is contained in a message, and appending these bits to the message. For example, a particular protocol may require that the data link layer compute and append a CRC to a message prior to transmission. The receiver of the message may then also compute the CRC and compare it to the CRC which has been appended to the message. If a mismatch exists, then it is assumed an error has occurred. Error correction may take on a variety of forms. One of the simplest methods of error correction is to require that the data link layer of the transmitter retransmit a message which has been detected to have an error during transmission. This method is based on the assumption that the error was caused by a disturbance which is unlikely to occur again. Another method of error correction involves transmission of additional bits of information along with the user information in a message. These additional bits of information are computed by the transmitter to provide redundant information in the message. When fewer than a certain number of bit-level errors have occurred during the transmission of the message, the receiver is able to reconstruct the original user information accurately using the redundant information (bits) supplied within the message. Error isolation is a capability of some networks to localize the source of errors and isolate the sections of the network or the stations at which the errors have been localized. Error isolation allows the fault-free portions of the network to continue communicating even when other portions of the network have degraded to the point of generating errors.

Internetworking

There are occasions when communications between two or more points are best handled by multiple networks. This may be the case when a single network has limitations that prevent it from tying the points together (e.g., distance limits) or when multiple networks are required for other reasons (e.g., to carry different types of data). When multiple networks are used to provide communications, there may be a need to pass messages or information directly from one network to another.

A repeater may be used when the networks to be joined are logically identical, and the purpose is simply to extend the length of the network or extend its capabilities in some way. A repeater generally has no effect on messages, and simply carries all messages from one cable or port to another (i.e., a change of physical media). A repeater allows for connection of networks at the physical layer level.

A bridge is similar to a repeater, but allows for connection of networks at the data link layer level. Generally, a bridge will pass all messages from one network to another, by passing messages at the data link layer level.

A router usually has the function of partitioning similar networks. Two networks may be based on the same technologies and protocol, but may not be logically identical. In these cases, some, but not all, of the messages on one network may need to be carried or transported to the other network. The router has the function of determining which messages to pass back and forth based on certain rules. Functions to enable efficient, automatic routing of messages may be included in layer 3 (the network layer) of the OSI network reference model, and a router allows for connection of networks at the network layer level.

A gateway may have a function similar to a router, or it may have the function of joining dissimilar networks, i.e., networks based on dissimilar technologies and/or protocols. When functioning like a router, a gateway usually performs its discrimination at a higher protocol level than a router. When a gateway joins dissimilar networks, generally a more complex set of rules must be designed into the gateway so that message translation, mapping, and routing can occur within the gateway as it determines which messages to pass from one network to the other.

ISO/OSI Network Reference Model

The explosion in the use and types of communication networks over the last several decades has led to more precise descriptions and treatment of communication networks in general. The International

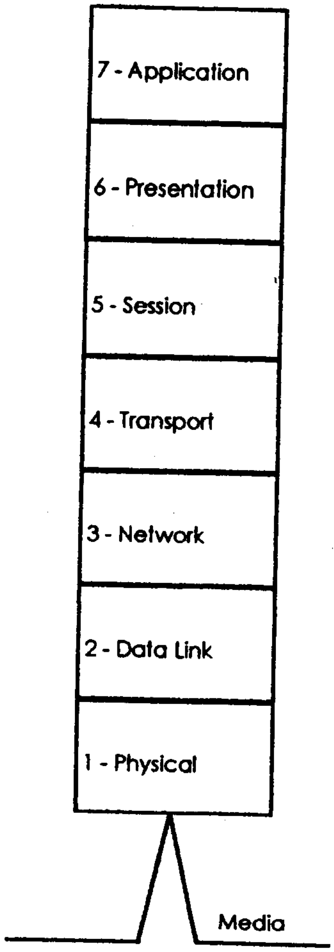

FIGURE 88.2 The ISO-OSI Seven Layer Model provides a method for segmenting communication functions

Organization for Standardization (ISO) has recognized one such method of precise description of networks, called the OSI reference model [1]. As shown in Figure 88.2, this model decomposes an arbitrary communication network into a "stack" of seven "layers." At each layer, certain types of network communication functions are described. The user of the communication system — usually another system that needs to communicate on the network — interacts with layer 7, the highest layer. The actual transmission medium (e.g., copper cable, fiber optic, free space, etc.) is connected to layer 1, the lowest layer. Most communication networks do not implement all of the layers in the reference model. In this case, formal definition, treatment, or inclusion of certain layers of the model in the actual network design are omitted. Layers 1, 2, and 7 are typically present in all networks, but the other layers may only be explicitly included or identifiable when their function is an important part of the network communications. In many sensor communication networks, the functions performed by layers 3,4,5, and 6 are "collapsed" into vestigial additions to the functions of layer 7, the application layer.

Physical Layer

The physical layer is the lowest layer of the model. This layer is responsible for converting between the symbolic or data representation of the network messages and the actual physical representation of data in the network medium. This layer specifies the behavior of the electric circuits referred to as the transmitter and the receiver. It also defines physical structures for connectors.

Data Link Layer

The data link layer, or layer 2, is responsible for several functions. This layer manages access to the network medium (MAC), structures the bits of information into well-defined groups identified as "frames" or messages, handles identification of source and destination stations on the network, and provides for error-free transmission of a message from source to destination stations, all according to the data link layer protocol. A number of standard data link layer protocols exist, which act as the basis for many of the communication networks in wide use. Ethernet, or IEEE 802.3, for example, specifies a MAC sublayer that works with the IEEE 802.2 Logical Link Control layer to form the data link layer protocol used in the majority of office information networks [2].

Network Layer

The network layer encapsulates functions related to routing of messages, both within a single network and among multiple networks. This layer typically uses addressing in a variety of forms as a key part of the functions of directing and routing messages, and the search and usage of the available communication paths.

Transport Layer

The transport layer provides any additional data transfer functions not directly provided by the data link layer for end-to-end reliable messaging. For example, some data transfer functions between stations may require the use of multiple data link layer messages to accomplish a reliable message transfer. The generation of multiple messages and the sequential disassembly, delivery, and assembly of data is accomplished by the transport layer. The transport layer also recovers from lost, duplicated and misordered messages.

Session Layer

The session layer provides for a higher level of control and management of network usage and data flow than that provided at lower layers, including opening or building up a communication channel, maintaining the channel, and closing the channel. This layer is infrequently implemented in contemporary systems.

Presentation Layer

The presentation layer provides functions to transform data from formats that are transportable by the network to the user-accessible formats that are defined in the application layer and understood in the local station.

Application Layer

The application layer, or layer 7, provides communication services directly to the user application. The usage and formatting of these services is summarized in the application layer protocol. The user interacts with the network by invoking functions and services provided by the application layer and passing data to and from the network through these services.

88.3 Network Technologies

There is a wide range of technologies in various stages of development and standardization, which address virtually all levels or layers of the ISO/OSI network reference model. One or more of the available technologies will probably suit almost any networking need. An analysis of the available technologies and their limitations will also be beneficial if it is deemed that a networking method must be designed to meet a particular application. The selection and description of technologies is by no means complete or exhaustive. The technologies presented are selected from several industries which make common use of networking to communicate sensor data. Figure 88.4 provides a comparison of selected parameters for a set of networks.

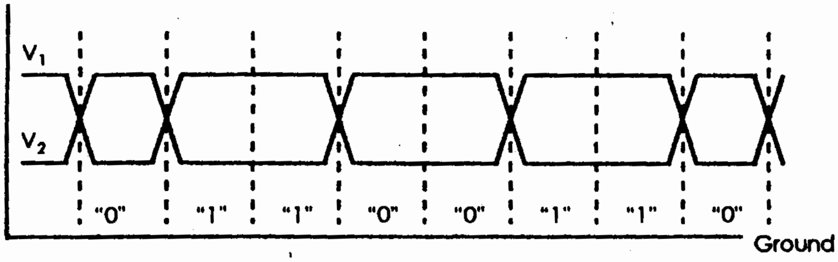

FIGURE 88.3 A sample RS-485 waveform showing voltages on differential wire pair (V_1, V_2) and superimposed bit intervals showing 0 and 1 bits. The ground reference is arbitrary within the defined signaling range.

RS-232

RS-232 (ANSI/EIA/TIA-232-E-91) is a widely used method of communication, which has been standardized in a variety of places including the Electronics Industry Association [3]. RS-232 represents elements of layer 1 of the OSI model, for communicating between two (and only two) stations. RS-232 provides a separate wire for transmission of data in each direction between the two stations, and gives the two stations different designations — data terminal equipment (DTE), and data communications equipment (DCE) — so that a method exists to distinguish which station will use which wire to transmit and receive. The signal levels for RS-232 represent a digital 1 bit as a voltage in the range of 5 to 12 V on the wire, and a digital 0 bit as a voltage of negative 5 to 12 V on the wire. RS-232 is typically implemented in a full duplex fashion, since each station can transmit to the other simultaneously using separate wires. RS-232 can be made to operate at a variety of bit rates, but typically is used at bit rates from 300 bit/s up to 115,200 bit/s.

RS-485

EIA RS-485 was made a standard in 1983, derived from the RS-422 standard. RS-485 provides for differential transmission of data on a pair of wires among 32 or more stations. Like RS-232, the standard is a layer 1 specification. RS-485 provides for half duplex communication, since a station cannot simultaneously transmit and receive independent data streams. Each station in an RS-485 system can have either a transmitter or a receiver, or both (commonly called a transceiver). Most implementations provide a transceiver. When one transceiver is transmitting, all others should be receiving (i.e., not transmitting). Which station is allowed to transmit at which time is not specified in the standard, and must be covered by a higher layer protocol (e.g., Interbus-S, Profibus-DP). Figure 88.3 shows a sample RS-485 waveform, indicating the differential nature of the signaling.

Seriplex[1]

Seriplex® is a digital, serial multiplexing system developed by Automated Process Control, Inc., in Jackson, MS. Square D Corporation purchased Automated Process Control and the rights to Seriplex in 1995, and subsequently launched Seriplex Technology Organization (STO) to manage the protocol. Seriplex is designed to be particularly efficient at handling large numbers of digital or on/off input and output points. Seriplex provides three communication wires, one for a clock signal, one for a data signal, and a ground reference. The system can be operated in two different modes (peer-to-peer and master–slave). In master–slave mode, one station is designated the master. The master synchronizes all data transmission among stations by driving a digital waveform on the clock line which all stations listen to and use for timing of transmit and receive operations. The master generates a repetitive pattern on the clock line

[1]Seriplex is a trademark of the Seriplex Technology Organization.

	Length	Stations	Bit Rate	Wires	Media	Topology
RS-232	30 m	2	115 kb/s	2	TP	P-P
RS-485	1200 m	32	10 Mb/s	2	TP	D-C
Seriplex	1500 m	256	200 kb/s	4	2STP	D-C,Free
AS-i	100 m	32	167 kb/s	2	UP	T-B
Interbus-S	25.6 km*	64	500 kb/s	6	3STP	Ring
CAN	450 m	64	1 Mb/s	4	2STP	T-B
4-20 mA	1000 m	2	-	2	STP	P-P
HART	1000 m	2(15)	1200 b/s	2	STP	P-P(D-C)
Profibus	9600 m	126	12 Mb/s	2	STP	D-C
Found. Fieldbus	1900 m	32	2.5 Mb/s	2	STP	D-C
LonWorks	1400 m	64	1.2 Mb/s	2	STP	D-C,Free

FIGURE 88.4 A comparison of selected parameters (maximum values) for various network technologies.** *Notes:* P-P = point to point; D-C = daisy-chain; T-B = trunkline-branchline; TP = twisted pair; STP = shielded twisted pair; UP = unshielded pair. * Maximum 400 m between stations. ** Maximum parameters for networks are not achievable simultaneously, and do not include repeaters, routers, or gateways. Maximum parameters are estimates based on available information.

which causes all stations to transmit and/or receive data on each cycle, or "scan" of the network. Each station is given an address, and uses the address along with the clock signal to determine when to drive the data line (in the case of an input point) or when to monitor the data line for valid output data (in the case of an output point). There are variations possible in implementation which allow for various clock speeds and bit rates (16, 100, and 200 kHz). Other protocol details allow for the handling of analog or multibit input and output points (by combining several bits on sequential scans together), bus fault detection, input redundancy, and communication error control using multiple scans of the network. Implementing the protocol in a sensor or other device typically requires using a Seriplex ASIC (Application Specific Integrated Circuit) which must be licensed from the STO [4].

AS-i

Actuator Sensor Interface, or AS-i, was developed by a consortium of primarily European companies interested in developing a low-cost, flexible method for connecting sensors and actuators at the lowest levels of industrial control systems. The system is managed by an independent worldwide organization [5]. The AS-i system provides a two-wire, nontwisted cable for interconnection of devices. Devices may draw current from the two wires (nominally at 24 V dc) for powering circuitry, and the data communications are modulated on top of the nominal dc level at a bit rate of 167 kHz, under the control of the master. A single parity bit per station is used for error detection. Similar to Seriplex, an AS-i device is typically implemented using a special ASIC which handles the communication.

Interbus-S

Interbus-S was developed by Phoenix Contact [6] and is controlled by the Interbus-S Club. The topology of the network is a ring, with data being sequentially shifted from point to point on the ring under the control of a network master. Each device in the ring acts as a shift register, transmitting and receiving data simultaneously at 500 kHz. The actual serial data transmission between stations conforms to RS-485. Interbus-S transmissions include a CRC for error detection. Interbus-S (Interbus-S Remote Bus) has also been extended to include a subprotocol called Interbus-Sensor Loop (or Interbus-S Local Bus). This subprotocol provides an alternate physical layer, with a single twisted pair carrying power and data on the same lines, and a reduction in the minimum size of the shift register in each station from 16 to 4 bits. Each Interbus sensor loop system can act as a single station on an Interbus-S network, or the sensor loop can be connected directly to a controller or master. Interbus-S devices are usually implemented with a special ASIC.

CAN

Controller Area Network (CAN) is a data link layer (layer 2) network technology developed by Robert Bosch Corporation [7], with an application target of onboard automotive networking. The technology is standardized in ISO 11898 [8], licensed to all major integrated circuit manufacturers, and is widely available — both as separate CAN controllers as well as CAN controllers integrated with microprocessors. As a result, CAN has been used in a variety of industries. As a data link layer technology, it is not a complete network definition. A number of physical layer options are usable with CAN (e.g., twisted pair, fiber optic, radio frequency wireless) and some have been subject to standardization (e.g., ISO 11898). Also, a number of application layer protocols have been developed for use with CAN, such as DeviceNet, Smart Distributed System (SDS), CANOpen [9], and SAE J1939 [10]. Both DeviceNet [11] and Smart Distributed System [12] have developed systems for creating networks of industrial field devices for the factory floor, including sensors and actuators.

4 to 20 mA Current Loop

The 4 to 20 mA current loop is a widely used method for transferring information from one station (the transmitter) to another station (the receiver). Therefore, this system allows for only two stations. A typical current loop system assigns a sensing range (e.g., 0 to 100°C) to the current range between 4 and 20 mA. A loop exists (i.e., two wires) between the transmitter and receiver. The transmitter can impress a certain current in the loop (using a controlled current source) so that the receiver can measure the current in the loop (e.g., by placing a small resistor in series with the loop and measuring the voltage drop across the resistor). After measuring the current, the receiver can then determine the present level of the sensed signal within the defined sensing range. This method uses current signaling, instead of voltage signaling, and therefore is relatively unaffected by potential differences between the transmitter and the receiver. This is similar to the benefit of differential (voltage) signaling, which also requires two wires. Another characteristic of this method is that it is not primarily digital in nature, as many other sensor communication systems are. The measured value can vary continuously in the range of 4 to 20 mA, and therefore can easily represent an analog sensing range, rather than a set of digital signals. Also, the signal is continuously variable and available. Another characteristic of this method is that the integrity of the loop can be verified. As long as the loop is unbroken and the transmitter is in good working order, the current in the loop should never fall below 4 mA. If the current approaches 0 mA, then the receiver can determine that a fault exists — perhaps a broken cable. These systems are widely used in various process control industries (e.g., oil refining) for connecting sensors (transmitters) with control computers. Because one station is always the transmitter and one station is always the receiver, this is a unidirectional, half duplex communication system.

HART[2]

HART® is a protocol which builds upon 4 to 20 mA communication systems. The basic idea is that additional data (beyond the basic sensor signal being carried in the current loop) can be transmitted by modulating a signal on top of the current flowing in the loop. The actual modulation method conforms closely to the Bell 202 standard for analog modem communications on telephone lines at 1200 bit/s. Because a 4 to 20 mA current loop carries a relatively slowly varying signal, it is easy to separate the 4 to 20 mA signal from the digital signal using filters. The Bell 202 standard uses continuous-phase frequency shift keying between two frequencies at up to 1200 shifts/s to modulate digital ones and zeros onto the 4 to 20 mA current loop. This method allows for bidirectional, full duplex communication between the two stations, on top of the 4 to 20 mA signal. It is also possible to configure HART communications on a network that is not carrying a 4 to 20 mA signal, in which case up to 15 devices can be connected together on the network. HART was developed by Fisher-Rosemount Corporation, and has been trans-

[2]HART is a trademark of the HART Communications Foundation.

ferred to an independent foundation for management [13]. Because HART is compatible with U.S. telephone systems, it can theoretically be run over the telephone line and is therefore capable of running over arbitrarily long distances.

Profibus

Profibus (PROcess FIeld BUS) is one of three networks standardized by a European standard [14]. Profibus is under the control of a global organization, PNO [15]. Profibus is an umbrella network standard which encompasses three subnetworks within the Profibus family. Profibus-DP (Distributed Periphery) is the variant which is designed specifically for communication with field devices (sensors and actuators) at the device I/O level. Profibus-PA (Process Automation) is a variant which has more capabilities designed to support the needs of device-level networking for process industries, such as oil refining. One of the capabilities of Profibus-PA is its ability to be installed in an intrinsically safe way, thus providing a higher degree of safety in environments which may be explosive or otherwise hazardous. Profibus-PA typically uses a special physical layer specification standardized under IEC 1158-2, which is used by several network systems for process automation applications. IEC 1158-2 specifies a two-wire twisted pair implementation carrying both power and data on the same two wires at 31.25 kbit/s. Profibus-FMS (Fieldbus Messaging Specification) represents the highest level implementation, which is used to link together controllers (not field or I/O devices) in a factory.

Profibus-DP systems are typically master–slave systems, where usually a single network master (the host controller) communicates with a number of slave devices (remote I/O blocks and other I/O devices). The protocol provides for cyclic exchange of I/O information as well as on-demand exchange of other types of information. Profibus-DP can be implemented on several different physical layers, including RS-485 and fiber optics, at various bit rates up to 12 Mbit/s. Profibus messages include a CRC for error detection.

Foundation Fieldbus

Foundation Fieldbus (FF) is a networking standard which has grown out of an effort within industry standards organizations, especially ISA-SP50 [16], and IEC SC65C/WG6 [17], to provide a replacement for the 4 to 20 mA analog sensor communication standard. FF provides two basic levels of networking: H1 and H2. H1 is a lower-speed system that can provide intrinsically safe (IS) operation and uses a single twisted pair to deliver both power and data communications to field devices, according to IEC 1158-2. Running at a bit rate of 31.25 kbit/s, H1 is very similar to Profibus-PA, when run on the IEC 1158-2 physical layer standard. The H1 system is designed to be able to connect hierarchically "upward" to an H2 system, which acts as the host. FF H2 can be run at either 1 or 2.5 Mbit/s on twisted pair wires, and also provides an IS option at the 1 Mbit/s rate. The H2 system can act as a network backbone in a factory environment, carrying data among various H1 systems.

WorldFIP

WorldFIP [18] is another technology of the three that were standardized in the European standard EN 50 170, running on the IEC 1158-2 physical layer. Many of the proponents of WorldFIP have embraced FF, and contributed to the development of that standard. WorldFIP is a member of the FF, and FF has incorporated many of the capabilities of WorldFIP as a result. When run on the IEC 1158-2 physical layer, WorldFIP has similar capabilities to FF.

LonWorks[3]

LonWorks® is a networking technology developed and controlled by the Echelon Corporation [19]. LonWorks is designed to be a general-purpose networking technology suitable for a variety of industries.

[3]LonWorks, LonTalk, and Neuron are trademarks of the Echelon Corporation.

LonWorks has been applied extensively in the building automation and control industry, as well as a variety of other industries. The core LonWorks technology for devices is contained in special integrated circuits — called Neuron® chips — which combine several microprocessors to manage the network, communications, and provide a general-purpose control environment. These chips are available from Motorola, Inc., and the Toshiba Corporation, which are licensees of the LonWorks technology. Echelon has also announced the possibility to license the LonTalk® protocol to other manufacturers for implementation in other microprocessors. LonWorks networks can be implemented on a variety of physical layers, including twisted pair at several bit rates and wireless options at 4800 bit/s, but the most common is a differential twisted pair system running at 78 kbit/s. Most of the networking details (the LonTalk protocol) are hidden from the user, and are encapsulated as functions within the general-purpose control environment. The user programs (using a language like the C programming language) the Neuron chip for each station to behave in a certain way and communicate various data items to other stations. Then, specialized tools are used to tie all of the stations together (handling addressing and other network details) to yield a functioning network. The system combines flexibility with a certain amount of ease of implementation, and can easily be applied to a variety of applications.

88.4 Applying Network Communications

Shielding

Many communication networks require shielding of the media (the cable). Shielding constitutes an electric conductor which completely encases the communication media (e.g., twisted pair) to provide protection against EMI. Shielding provides an electric conductive barrier to attenuate electromagnetic waves external to the shield, and provides a conduction path by which induced currents can be circulated and returned to the source, typically via a ground reference connection. Shields in communication systems are often grounded at only a single point. This single point of ground prevents the shield from participating in a "ground loop," which is an alternative path for current to flow between two points of potential difference connected to a common ground. Ground loops can lead to noise problems, and can be destructive if the stray currents are large enough, since a shield ground is usually not constructed to carry heavy currents.

Media

The most common media types for network systems fall into three categories: electric, optical, and electromagnetic. Electric media are based on conductors (e.g., copper wire), whereas optical media are based on optical waveguides, or fiber optics. Electromagnetic media consists of free space, or general electromagnetic wave-permeable materials, and are referred to as wireless systems. Within the category of electric media are a large variety of conductor configurations. The most common are unshielded pair, unshielded twisted pair (UTP), shielded twisted pair (STP), and coaxial (COAX). These conductor configurations have various properties which are significant for the transmission of electric signals, as well as varying degrees of immunity to EMI. As a rule of thumb, the quality of the transmission line characteristics (signal transmission and immunity to EMI) improves in the order listed. Twisted pair systems are generally easier to install, whereas coaxial and fiber-optic systems generally require more specialized tools and termination methods. Of course, wireless systems are easy to install, but attention must still be paid to the media. The characteristics of the free space such as distance and amount of EMI present must be considered for reliable operation of the network.

Bit Rate

Some networks provide only one choice of bit rate, whereas others provide user-selectable options for bit rate. Bit-rate options may be dependent on the type of media that is installed. As a rule of thumb,

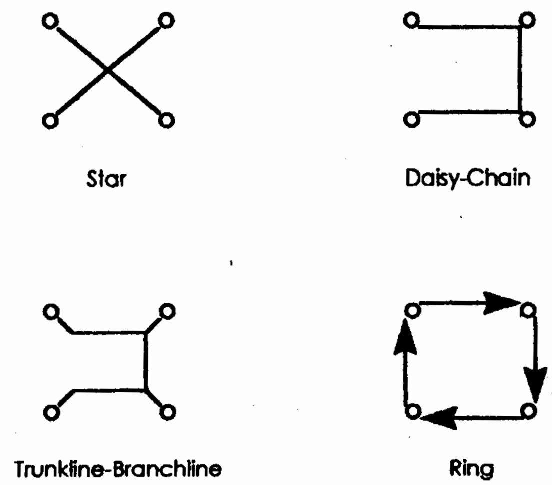

FIGURE 88.5 Some examples of the many possible network topologies using four stations.

the bit rate chosen should be the lowest possible bit rate that still supports the application requirements for speed of data transfer and overall bandwidth. This generally results in more reliable operation, and generally gives the network more immunity to minor degradations, specification violations, and EMI.

Topologies

There are a variety of network topologies that are commonly used. Topology refers to the physical arrangement and interconnection of stations by the media. Some networks can be run using several different topologies; some can only be run with a certain topology (e.g., Interbus-S requires a ring topology). The most common topologies are daisychain, trunkline–branchline, ring, and star. Variations on these exist, and networks that incorporate or can be run on highly varied topologies are sometimes called free-form, tree, or free topology networks. Figure 88.5 depicts graphically several different types of topologies. In some cases, networks require certain topologies. Deviating from these can cause degradation in network behavior (e.g., corruption of messages) or network failure.

Configuration

Most networks involve some sort of configuration. Configuration is the process of connecting stations together and assigning certain programmable parameters to each station required for proper operation of the network. The most common configurable parameter in many networks is the station address. Some networks may require other parameters to be preset, such as the communication speed, or bit rate. Some networks have the capability to autoconfigure, which means to assign parameters automatically to stations as part of the network start-up process, without explicit user intervention (e.g., Interbus-S). Many networks define various tools, which may be computer based, to assign parameters to each station in order to configure the network. In other cases, the stations may incorporate switches or other manual means to configure the necessary parameters for network operation.

88.5 Advanced Topics

Wireless Technologies

The need for networking is present even in environments where an electrical or optical cable cannot be easily distributed. This may be due to various limitations, such as difficulty in running a new cable from one building to another, or connecting to sensors in motion or on vehicles. There are two general categories of wireless communications, based on electromagnetic frequency spectra. Various wireless technologies employ the infrared spectrum. These technologies generally have transmission limited to applications that have a direct line of sight between stations. Also, the distances are generally limited to 100 m or less. Because of these limitations, there are generally no legal restrictions in employing these frequency spectra, and infrared transceivers are now becoming available from a variety of manufacturers.

The other general category of wireless communications is based on radio frequency (RF) communications. In most countries, use of these spectra is tightly controlled by governmental agencies. As a result, employing wireless networking in most of these frequency ranges requires special licensing. However, a number of frequency ranges are reserved for low-power public communications. Within these frequency ranges, devices are allowed to communicate in an unlicensed fashion as long as they transmit according to certain rules about transmitted power output. RF-based wireless systems are generally not limited to line-of-sight applications, and can be designed to cover greater distances than infrared-based systems.

Wireless technologies can be viewed as simply another choice for the physical layer media, i.e., free space. As such, it is possible to consider, in some cases, a wireless media for implementation of a variety of protocols. For example, both CAN and LonWorks systems could be candidates for wireless networking.

Fiber Optics

Another physical layer media choice is fiber-optic media. Fiber-optic media employs pulses of light delivered along a tubular waveguide (glass or plastic fiber) to transmit information from one station to another. Fiber optics enjoy some benefits over traditional copper wiring. First, attenuation of light within fiber optics is generally about an order of magnitude less than attenuation of an electric signal within a copper wire. Second, fiber-optic transmission systems can be modulated (or pulsed) at much higher frequencies, yielding greater potential bandwidths and bit rates than copper media. Finally, fiber-optic systems are generally immune to the traditional sources of EMI that can cause trouble for copper media systems. There are also limitations in present implementations of fiber-optic systems. One of the limitations is that special tools and termination techniques must be used to connect a fiber-optic cable to a sensor or field device. Second, fiber-optic "taps" are not easily created. Therefore, most fiber-optic systems are implemented in a point-to-point fashion. When multiple devices are involved in a network, each device usually acts as an optical repeater, with a fiber-optic input and a fiber-optic output port.

Network Design Considerations

Designing a network communication system from the ground up can be a lengthy undertaking, and should not be considered unless a careful review of available technologies has yielded no solutions to the particular requirements of the application. The designers must take into account a number of fundamental questions to shape the capability of the network. One topic mentioned frequently in the area of networking for control applications is the subject of determinism. This refers to the ability of the network to behave in a predictable fashion under any given set of stimuli, or external conditions. Many networks do not exhibit this characteristic. Another question to be resolved is the subject of priority, and media access. The designers must determine the conditions under which any particular station is allowed to transmit, and if multiple stations are attempting to transmit, how it will be determined which station will be given priority to transmit first. Media access methods often impact the ability of a network to behave in a deterministic fashion.

Integrating Sensors with Communications — IEEE P1451

A recent interesting development in the area of sensor networks is an effort being sponsored by the IEEE [20] out of its TC-9 committee, called IEEE P1451. This activity is working toward the development of a standard to define sensor (or transducer) interfaces to networks generically. The first part of the proposed standard, IEEE P1451.1, includes definitions for the interface between the device and the network (refer to Figure 88.1). The second part, IEEE P1451.2, includes definitions for the interface between the transducer (or sensor) and the network interface block within the device. P1451.2 includes a definition for a transducer electronic data sheet, or TEDS, which defines a summary set of information pertinent to the sensor, allowing for standardized exchange of data on the network. The proposed standard has the potential benefits to make it easier to connect a sensor to a variety of networks, and to allow similar sensors from different manufacturers to be handled in a similar fashion on the network.

References

1. *ISO/IEC 7498-1:1994 Information Technology — Open Systems Interconnection — Basic Reference Model: The Basic Model,* International Organization for Standardization (ISO), 1, rue de Varembé, Case postale 56, CH-1211 Genève 20, Switzerland, [online]. Available http://www.iso.ch/.
2. *8802-3: 1996 (ISO/IEC) [ANSI/IEEE Std 802.3, 1996 Edition] Information Technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Specific requirements — Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications,* Institute of Electrical and Electronics Engineers, 445 Hoes Lane, Piscataway, NJ 08855-1331, [online]. Available http://www.ieee.org/.
3. *ANSI/EIA/TIA-232-E-91,* Electronic Industries Association, 2500 Wilson Boulevard, Arlington, VA 22201-3834, [online]. Available http://www.eia.org/.
4. *Distributed, Intelligent I/O for Industrial Control and Data Acquisition… The SERIPLEX Control Bus, Bulletin No. 8310PD9501R4/97,* Seriplex Technical Organization, P.O. Box 27446, Raleigh, NC 27611-7446, [online]. Available http://www.seriplex.org/.
5. AS-Interface U.S.A., 5010 East Shea Blvd., Suite C-226, Scottsdale, AZ 85254, [online]. Available http://www.as-interface.com/.
6. *Interbus-S Protocol Structure, Data Sheet 0005C,* Phoenix Contact, P.O. Box 4100, Harrisburg, PA 17111, [online]. Available http://www.ibsclub.com/.
7. *CAN Specification, Version 2.0, 1991,* Robert Bosch GmbH, Postfach 50, D-7000 Stuttgart 1, Germany.
8. *ISO 11898:1993 Road vehicles—Interchange of digital information—Controller area network (CAN) for high-speed communication,* International Organization for Standardization (ISO), 1, rue de Varembé, Case postale 56, CH-1211 Genève 20, Switzerland, [online]. Available http://www.iso.ch/.
9. *CiA Draft Standard 301 (Version 3.0), CANopen Communication Profile for Industrial Systems,* CiA Headquarters, Am Weichselgarten 26, D-91058 Erlangen, Germany, [online]. Available http://www.can-cia.de/.
10. *SAE J 1939 — Recommended Practice for Serial Control and Communications Vehicle Network,* Society of Automotive Engineers, 400 Commonwealth Drive, Warrendale, PA 15096, [online]. Available http://www.sae.org/.
11. *DeviceNet Specification v1.4,* Open DeviceNet Vendors Association, 8222 Wiles Road, Suite 287, Coral Springs, FL 33067, [online]. Available http://www.odva.org/.
12. *Smart Distributed System Application Layer Protocol Specification v2.0, 1996,* Honeywell MICRO SWITCH Division, 11 West Spring Street, Freeport, IL 61032, [online]. Available http://www.sensing.honeywell.com/sds/.
13. HART Communication Foundation, 9390 Research Boulevard, Suite I-350, Austin, TX, 78759 [online]. Available http://www.ccsi.com/hart/hcfmain.html.
14. *EN 50 170 — Volume 2,* CENELEC Central Secretariat, 35, rue de Stassart, B-1050 Brussels, Belgium.
15. PROFIBUS Trade Organization U.S.A., 5010 East Shea Blvd., Suite C-226, Scottsdale, AZ 85254-4683, [online]. Available http://www.profibus.com/.
16. ISA, the International Society for Measurement & Control, P.O. Box 12277, Research Triangle Park, NC 27709, [online]. Available http://www.isa.org/.
17. *IEC 61158-2(1993-12), Fieldbus standard for use in industrial control systems — Part 2: Physical layer specification and service definition,* International Electrotechnical Commission, 3, rue de Varembé, PO Box 131, 1211 Geneva 20, Switzerland, [online]. Available http://www.iec.ch/.

18. WorldFIP Headquarters, 2, rue de Bone, 92160 Antony, France, [online]. Available http://www.world-fip.org/.
19. *005-0017-01 Rev C, LonTalk Protocol*, Echelon Corporation, 4015 Miranda Avenue, Palo Alto, CA 94304, [online]. Available http://www.echelon.com/.
20. *P1451.2, Draft Standard for a Smart Transducer Interface for Sensors and Actuators — Transducer to Microprocessor Communication Protocols and Transducer Electronic Data Sheet (TEDS) Formats*, Institute of Electrical and Electronics Engineers, 445 Hoes Lane Piscataway, NJ 08855-1331, [online]. Available http://www.ieee.org/.

Further Information

P. Z. Peebles, Jr., *Digital Communications Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1987, provides a good general text on communication.

B. Svacina, *Understanding Device Level Buses*, Minneapolis, MN: TURCK Inc. (3000 Campus Dr., Minneapolis, MN 55441), 1996, is an in-depth study of the subject of communication networks for industrial field devices.

J. D. Gibson, Ed., *The Communications Handbook*, Boca Raton, FL: CRC Press, 1997, includes recent material on communication techniques.