SECTION 7

CONTROL COMMUNICATIONS*

B. A. Loyer

Systems Engineer, Motorola, Inc., Phoenix, Arizona (Local Area Networks—prior edition)

Howard L. Skolnik

Intelligent Instrumentation, Inc. (a Burr-Brown Company), Tucson, Arizona (Data Signal Handling in Computerized Systems; Noise and Wiring in Data Signal Handling)

---

[*] *Persons who authored complete articles or subsections of articles, or otherwise cooperated in an outstanding manner in furnishing information and helpful counsel to the editorial staff.*

DATA SIGNAL HANDLING IN COMPUTERIZED SYSTEMS

by Howard L. Skolnik*

Prior to the advent of the digital computer, industrial instrumentation and control systems, with comparatively few exceptions, involved analog, rather than digital, signals. This was true for both the outputs from sensors (input transducers—strain gages, thermocouples, and so on) and the inputs to controlling devices (output transducers—valves, motors, and so on). In modern systems many transducers are still inherently analog. This is important because computers can operate with only digital information. Therefore a majority of contemporary systems include analog-to-digital (A/D) and digital-to-analog (D/A) converters.

An important feature of data-acquisition products is how they bring together sophisticated functions in an integrated, easy-to-use system. Given the companion software that is available, the user can take advantage of the latest technology without being intimately familiar with the internal details of the hardware. When selecting a system, however, it is useful to have a basic understanding of data-acquisition principles. This article addresses how real-world signals are converted and otherwise conditioned so that they are compatible with modern digital computers, including personal computers

---

* Intelligent Instrumentation, Inc. (a Burr-Brown Company), Tucson, Arizona.

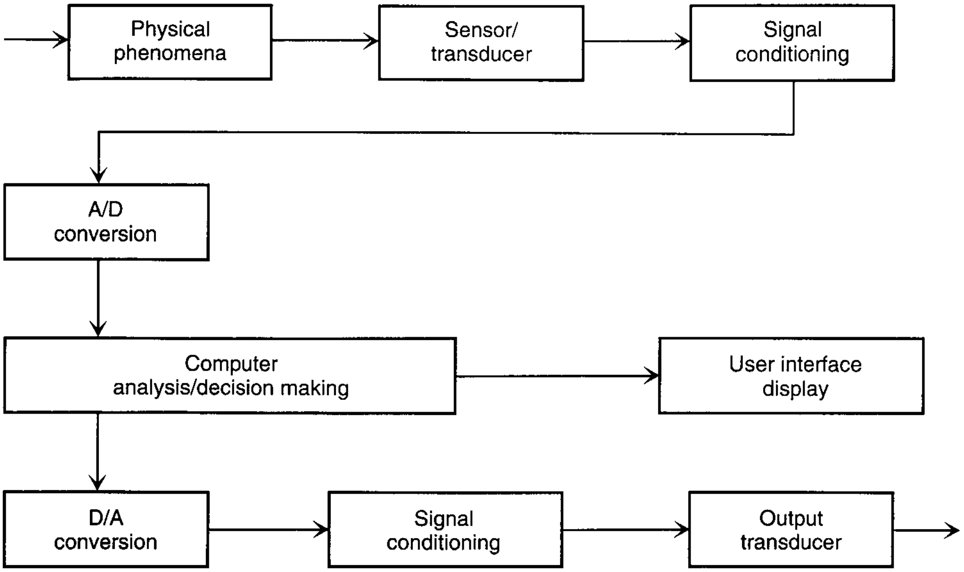


FIGURE 1 Flow diagram of modern computer-based data-acquisition and control system. The measured variable may be any physical or chemical parameter, such as temperature, pressure, flow, liquid level, chemical concentration, dimension, motion, or position. The sensor/transducer is the measuring device (primary element) that converts the measured variable into an electrical quantity. Common transducers include thermocouples, strain gages, resistance temperature devices, pH cells, and switches. The signal from a transducer can be in the form of a voltage, current, charge, resistance, and so on. Signal conditioning involves the manipulation of the raw transducer's output into a form suitable for accurate analog-to-digital (A/D) conversion. Signal conditioning can include filtering, amplification, linearization, and so on. Data conversion provides the translation between the real world (mostly analog) and the digital domain of the computer, where analysis, decision making, report generation, and user interface operations are easily accomplished. To produce analog output signals from the computer (for stimulus or control) digital-to-analog (D/A) conversion is used. Signal conditioning and output transducers provide an appropriate interface to the outside world via power amplifiers, valves, motors, and so on. (*Intelligent Instrumentation, Inc.*)

(PCs). The techniques suggested here are specifically aimed at PC-based measurement and control applications. These generally involve data-acquisition boards that plug directly into an expansion slot within a PC. References to specific capabilities and performance levels are intended to convey the current state of the art with respect to PC-based products (Fig. 1).

SIGNAL TYPES

Signals are often described as being either analog, digital, or pulse. They are defined by how they convey useful information (data). Attributes such as amplitude, state, frequency, pulse width, and phase can represent data. While all signals can be assumed to be changing with time, analog signals are the only ones to convey information within their incremental amplitude variations. In instrumentation and control applications most analog signals are in the range of −10 to +10 volts or 4 to 20 mA. Some of the differences between analog and digital signals are suggested in Fig. 2. Digital and pulse signals have binary amplitude values, that is, they are represented by only two possible states—low and high. While low and high states can be represented by any voltage level, transistor-transistor-logic (TTL) levels are most often used. TTL levels are approximately 0 and 5 volts. The actual allowable ranges

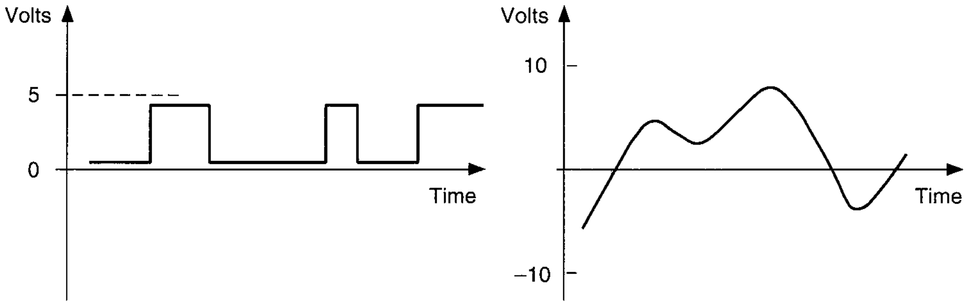


FIGURE 2 Comparison of digital signal (left) and analog signal (right). (*Intelligent Instrumentation, Inc.*)

for TTL signals are

$$\text{Low level} = 0 \text{ to } 0.8 \text{ volt}$$
$$\text{High level} = 2.0 \text{ to } 5.0 \text{ volts}$$

Thus with analog it is important how high the signal is, while with digital it matters only whether the signal is high or low (on or off, true or false). Digital signals are sometimes called discrete signals. While all digital signals have the potential of changing states at high speed, information is usually contained in their static state at a given point in time. Digital inputs can be used to indicate whether a door is open, a level is beyond a set point, or the power is on. Digital outputs can control (on or off) power to a motor, a lamp, an alarm, and the like. In contrast, analog inputs can indicate how high a level is or how fast a motor is turning. Analog outputs can incrementally adjust the position of a valve, the speed of a motor, the temperature of a chamber, and so on. Pulse signals are similar to digital signals in many respects. The distinction lies in their time-dependent characteristics. Information can be conveyed in the number of state transitions or in the rate at which transitions occur. Rate is referred to as frequency (pulses per second). Pulse signals can be used to measure or control speed, position, and so on.

TERMINATION PANELS

Termination panels are usually the gateway to a data-acquisition system. Screw terminals are provided to facilitate easy connection of the field wiring. Figure 3 suggests two of the many termination panel styles. Some models are intended for standard input or output functions, while others are designed to be tailored for unique customized applications. Mounting and interconnection provisions are provided for resistors, capacitors, inductors, diodes, transistors, integrated circuits, relays, isolators, filters, connectors, and the like. This supports a wide range of signal interface and conditioning capabilities. In most cases termination panels are located outside, but adjacent to, the data-acquisition system's host PC. Many mounting and enclosure options are available to suit different applications. Because the actual data-acquisition board is located inside the PC, short cables (normally shielded ribbon cables) are used to connect the termination panel's signals.

FIELD SIGNALS AND TRANSDUCERS

Whatever the phenomenon detected or the device controlled, transducers play a vital role in the data-acquisition system. It is the transducer that makes the transition between the physical and the electrical world. Data acquisition and control can involve both input and output signals. Input signals

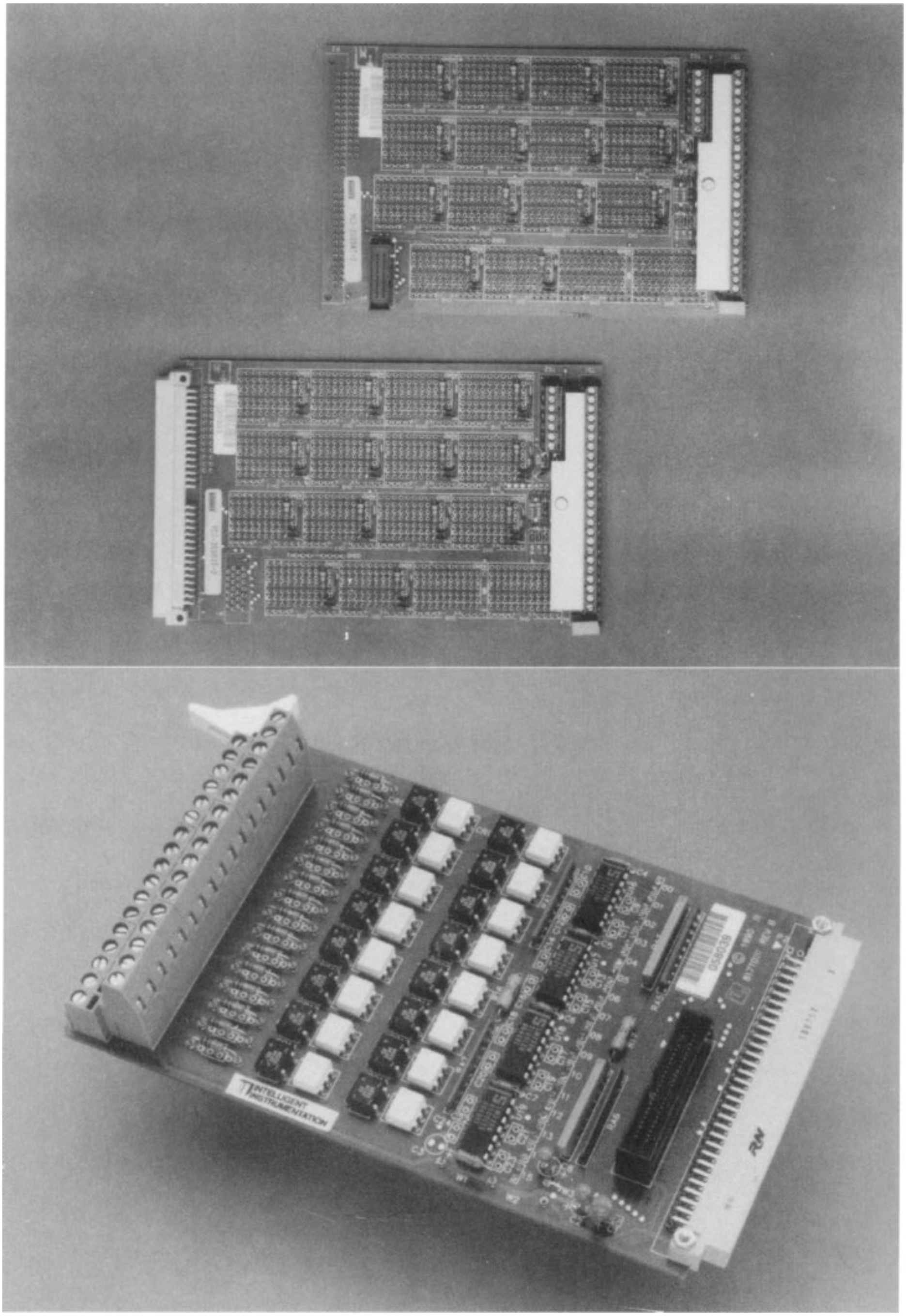


FIGURE 3 Representative termination panel styles. (*Intelligent Instrumentation, Inc.*)

can represent force, temperature, flow, displacement, count, speed, level, pH, light intensity, and so on. Output signals can control valves, relays, lamps, horns, and motors, to name a few. The electrical equivalents produced by input transducers are most commonly in the form of voltage, current, charge, resistance, or capacitance. As shown later, the process of signal conditioning will further convert these basic signals into voltage signals. This is important because the interior blocks of the data-acquisition system can only deal with voltage signals.

Thermocouples

Thermocouples are used widely to measure temperature in industry and science. Temperatures in the range of −200 to +4000°C can be detected. Physically a thermocouple is formed by joining together wires made of two dissimilar metals. The resulting junction produces a voltage across the open ends of the wires that is proportional to temperature (the Seebeck effect). The output voltage is usually in the range of −10 to +50 mV and has an average sensitivity of 10 to 50 μV/°C, depending on the metals used. However, the output voltage is very nonlinear with respect to temperature. Many different thermocouple types are in wide use. For convenience, alphabetic letter designations have been given to the most common. These include the following:

J	Iron-constantan (Fe-C)
K	Chrome-Alumel (Ch-Al)
T	Copper-constantan (Cu-C)

Tungsten, rhodium, and platinum are also useful metals, particularly at very high temperatures.

Thermocouples are low in cost and very rugged. Still, they are not without their limitations and applications problems. In general, accuracy is limited to about 1 to 3 percent due to material and manufacturing variations. Response time is generally slow. While special thermocouples are available that can respond in 1 to 10 ms, most units require several seconds. In addition to the thermocouple's nonlinear output, compensation must also be made for the unavoidable extra junctions that are formed by the measuring circuit.

As mentioned previously, a single thermocouple junction generates a voltage proportional to temperature:

$$V = k(t) \tag{1}$$

where k is the Seebeck coefficient defining a particular metal-to-metal junction, and t is in degrees/kelvin.

Unfortunately the Seebeck voltage cannot be measured directly. When the thermocouple wires are connected to the terminals of a voltmeter or data-acquisition system, new thermoelectric junctions are created. For example, consider the copper-constantan (type T) thermocouple connected to a voltmeter shown in Fig. 4. It is desired that the voltmeter read only V_1 (of J_1), but the act of connecting the voltmeter creates two more metallic junctions, J_2 and J_3. Since J_3 is a copper-to-copper junction, it

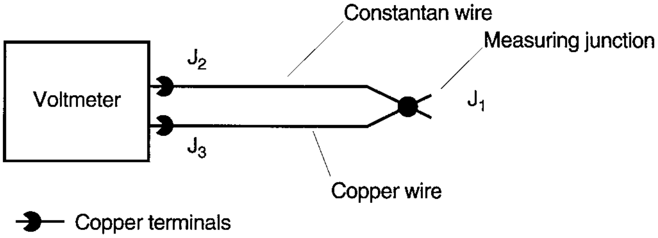


FIGURE 4 Essence of the thermocouple measurement problem. (*Intelligent Instrumentation, Inc.*)

creates no thermal voltage ($V_3 = 0$), but J_2 is a copper-to-constantan junction that will add a voltage V_2 in opposition to V_1. As a result, the voltmeter reading V_v will actually be proportional to the temperature difference between J_1 and J_2. This means that determining the temperature at J_1 requires a knowledge of the temperature at J_2. This junction is referred to as the *reference junction*, or *cold junction*. Its temperature is the reference temperature t_{ref}. Note that V_2 equals V_{ref}. Therefore it follows from Eq. (1),

$$V_v = V_1 - V_{\mathrm{ref}} = k(t_1 - t_{\mathrm{ref}}) \tag{2}$$

It is important to remember that k is highly nonlinear with respect to temperature. However, for measurement purposes it is not necessary to know the value of k. Tables have been compiled by the U.S. National Bureau of Standards (now the National Institute of Standards and Technology, NIST) that take variations in k into account and can provide t_1 directly in terms of V_v, assuming that t_{ref} is at 0°C. Separate tables were made for each thermocouple type. J_2 (and J_3) was physically placed in an ice bath, forcing its temperature to 0°C. Note that even under these conditions [see Eq. (1)], V_{ref} is not 0 volts. The Seebeck relationship is based on the kelvin (absolute zero) scale. In computer-based applications the thermocouple tables are transformed into polynomial equations for ease of use. Depending on the thermocouple type and the accuracy (compliance with the NIST tables) desired, between fifth- and ninth-order polynomials are used.

The copper-constantan thermocouple used in this example is a special case because the copper wire is the same metal as the voltmeter terminals. It is interesting to look at a more general example using iron-constantan (type J). The iron wire increases the number of dissimilar metal junctions in the circuit as J_3 becomes a Cu-Fe thermocouple junction. However, it can be shown that if the Cu-Fe and the Cu-C junctions (at the termination panel) are at the same temperature, the resulting voltage is equivalent to a single Fe-C junction. This allows the use of Eq. (2). Again, it is very important that both parasitic junctions be held at the same (reference) temperature. This can be aided by making all connections on an isothermal (same temperature) block.

Clearly, the requirement of an ice bath is undesirable for many practical reasons. Taking the analysis to the next logical step, Eq. (2) shows that t_{ref} need not be at any special temperature. It is only required that the reference temperature be accurately known. If the temperature of the isothermal block (the reference junction) can be measured independently, this information can be used to compute the unknown temperature t_1.

Devices such as thermistors, resistive temperature detectors, and semiconductor sensors can provide a means of independently measuring the reference junction. (Semiconductor sensors are the most popular for the reasons described hereafter.) A thermocouple temperature measurement, under computer control, could proceed as follows:

1. Measure t_{ref} and use the thermocouple polynomial to compute the equivalent thermocouple voltage V_{ref} for the parasitic junctions.
2. Measure V_v and *add* V_{ref} to find V_1.
3. Compute t_1 from V_1 using the thermocouple polynomial.

Solid-State Temperature Sensors

These devices are derived from modern silicon integrated-circuit technology, and are often referred to as Si sensors. They consist of electronic circuits that exploit the temperature characteristics of active semiconductor junctions. Versions are available with either current or voltage outputs. In both cases the outputs are directly proportional to temperature. Not only is the output linear, but it is of a relatively high level, making the signal interpretation very easy. The most common type generates 1 μA/K (298 μA at 25°C). This can be externally converted to a voltage by using a known resistor. The usable temperature range is -50 to 150°C. The stability and the accuracy of these devices are good enough to provide readings within ±0.5°C. It is easy to obtain 0.1°C resolution. Si sensors are ideal reference junction monitors for thermocouple measurements.

Resistance Temperature Detectors

Resistance temperature detectors (RTDs) exhibit a changing resistance with temperature. Additional detailed information on using RTDs can be found in the section on signal conditioning for resistive devices. Several different metals can be used to produce RTDs. Platinum is perhaps the most common for general applications. Yet at very high temperatures, tungsten is a good choice. Platinum RTDs have a positive temperature coefficient of about 0.004 Ω/°C. The relationship has a small nonlinearity that can be corrected with a third-order polynomial. Many data-acquisition systems include this capability. Platinum RTDs are usually built with 100-ohm elements. These units have sensitivities of about +0.4 Ω/°C. Their useful temperature range is about −200 to about +600°C.

Most RTDs are of either wire-wound or metal-film design. The film design offers faster response time, lower cost, and higher resistance values than the wire-wound type. The more massive wire-wound designs are more stable with time. High resistance is desirable because it tends to reduce lead-wire induced errors. To convert resistance into a voltage, an excitation current is required. Care must be taken to avoid current levels that will produce errors due to internal self-heating. An estimate of the temperature rise (in °C) can be found by dividing the internal power dissipation by 80 mW. This is a general rule that applies to small RTDs in a conductive fluid such as oil or water. In air the effects of self-heating can be 10 to 100 times higher.[1]

SAMPLED-DATA SYSTEMS

Modern data-acquisition systems use sampled-data techniques to convert between the analog and digital signal domains. This implies that while data may be recorded on a regular basis, they are not collected continuously, that is, there are gaps in time between successive data points. In general there is no knowledge of the missing information, and the amplitude of missing data points cannot be predicted. Yet under special circumstances it can be assumed that missing data fall on a straight line between known data points.

Fourier analysis reveals that signals, other than pure sine waves, consist of multiple frequencies. For example, a pulse waveform contains significant frequency components far beyond its fundamental or repetition rate. Frequencies extending to approximately $0.3/t_r$ are often important, where t_r is the pulse rise time. Step functions suggest that frequency components extend to infinity.

The Nyquist theorem defines the necessary relationship between the highest frequency contained in a waveform and the *minimum* required sampling speed. Nyquist states that the sample rate must be *greater* than two times the highest frequency component contained within the input signal. The danger of undersampling (sampling below the Nyquist rate) is erroneous results. It is not simply a matter of overlooking high-frequency information, but of reaching totally wrong conclusions about the basic makeup of the signal. See Fig. 5 for an example. Note that sampling a pure sine wave (containing only the fundamental frequency) at a rate in violation of the Nyquist criterion leads to meaningless results. This example suggests the presence of a totally nonexistent frequency. This phenomenon is known as aliasing.

If it were possible to sample at an infinite rate, aliasing would not be a problem. However, there are practical limits to the maximum sampling speed, as determined by the characteristics of the particular A/D converter used. Therefore action must be taken to ensure that the input signal does not contain frequency components that cause a violation of the Nyquist criterion. This involves the use of an input low-pass filter (antialiasing filter) *prior* to the A/D converter. Its purpose is to limit the measured waveform's frequency spectrum so that no *detectable* component equals or exceeds half of the sampling rate. Detectable levels are determined by the sensitivity of the A/D converter and the attenuation of the antialiasing filter (at a given frequency).

---

[1] Additional information on temperature sensors will be found in Section 3 of this handbook.

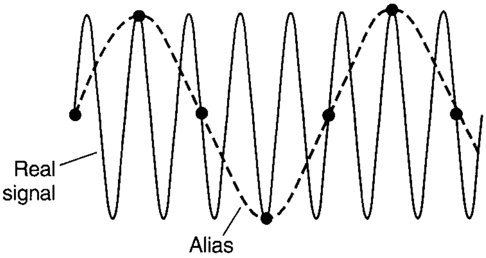


FIGURE 5 Aliasing because of insufficient sampling rate. (*Intelligent Instrumentation, Inc.*)

Sequential Scanning

Systems are usually designed to collect data from more than one input channel. To reduce cost, most PC-based systems share significant components, including the A/D converter, with all of the channels. This is accomplished with a multiplexer (electronic scanning switch). However, when there is only one A/D converter, only one input channel can be acquired at a given point in time. Each channel is read sequentially, resulting in a time skew between readings. Techniques for minimizing time skew will be described.

ANALOG INPUT SYSTEMS

The fundamental function of an analog input system is to convert analog signals into a corresponding digital format. It is the A/D converter that transforms the original analog information into computer-readable data (a digital binary code). In addition to the A/D converter, several other components may be required to obtain optimum performance. These can include a sample/hold, an amplifier, a multiplexer, timing and synchronization circuits, and signal-conditioning elements.

A good starting point when choosing an A/D converter (or a system using an A/D converter) is to consider the characteristics of the input transducer. What is its *dynamic range*, maximum signal level, signal frequency content, source impedance, accuracy, and so on? A match in characteristics between the A/D converter and the transducer is usually desired. It is also important to consider possible sources of external interfering noise. This will have a bearing on the choice of A/D converter and the required signal conditioning. Some sensors have a very wide dynamic range. Dynamic range defines the span of input stimulus values that correspond to *detectable* output values. It is often expressed as the ratio of the maximum full-scale output signal to the lowest *meaningful* output signal. In this context, dynamic range and signal-to-noise ratio are the same. Caution! There is not necessarily a good correlation between sensor dynamic range and *accuracy*. Accuracy refers to how close the measured output corresponds to the *actual* input. For example, a transducer could have a dynamic range of 5000:1 and an accuracy of 1 percent. If the full-scale range is 100°C, a change as small as 0.02°C can be detected. Still, the actual temperature is only known, in absolute terms, to 1°C. 1°C is the accuracy; 0.02°C is the sensor's sensitivity. The proper choices of signal conditioning and A/D converter are essential to preserving the performance of the transducer. This example suggests the need for an A/D converter with more than 12 bits of resolution. Applications using other transducers may require 14-, 16-, or even 18-bit resolution.

Analog-to-Digital Converters

Many different types of A/D converters exist. Among these, a few stand out as the most widely used—successive-approximation, integrating, and flash (parallel) converters. Each converter has a set of unique characteristics that makes it better suited for a given application. These attributes include speed, resolution, accuracy, noise immunity, cost, and the like.

Industrial and laboratory data-acquisition tasks usually require a resolution of 12 to 16 bits. 12 bits is the most common. As a general rule, increasing resolution results either in increased cost or in reduced conversion speed. Therefore it makes sense to consider the application requirements carefully before making a resolution decision.

All A/D converters accomplish their task by partitioning the full analog input range into a fixed number of discrete digital steps. This is known as digitizing or quantizing. A different digital code corresponds to each of the assigned steps (analog values). Digital codes consist of N elements, or bits. Because each bit is binary, it can have one of two possible states. Thus the total number of possible steps is 2^N. N is often referred to as the converter's *resolution* (that is, an N-bit converter). Given the number of steps S, it follows that $N = \log S/\log 2$. Caution is required when using the term "resolution" because it can also be expressed in other ways. For example, a 12-bit system divides its input into 2^{12}, or 4096, steps. Thus if the A/D converter has a 10-volt range, it has a resolution of 2.4 mV (10 volts ÷ 4096). This refers to the minimum detectable signal level. One part in 4096 can also be expressed as 0.024 percent of full-scale (FS). Thus the resolution is 0.024 percent FS. These definitions apply *only* to the ideal, internal capabilities of the A/D converter alone.

In contrast to this 12-bit example, 16 bits correspond to one part in 65,536 (2^{16}), or approximately 0.0015 percent FS. Therefore increasing the resolution has the *potential* to improve both dynamic range and overall accuracy. On the other hand, system performance (effective dynamic range and accuracy) can be limited by other factors, including noise and errors introduced by the amplifier, the sample/hold, and the A/D converter itself.

Flash-type A/D converters can offer very high-speed operation, extending to about 100 MHz. Conversion is accomplished by a string of comparators with appropriate reference voltages, operating in parallel. To define N quantizing steps requires $2^N - 1$ comparators (255 and 4095 for 8 and 12 bits, respectively). Construction is not practical beyond about 8 or 10 bits. In contrast, most data-acquisition and control applications require more than 10 bits.

12-bit devices that run in the 5- to 15-MHz range are available using a *subranging* or half-flash topology. This uses two 6-bit flash encoders, one for coarse and the other for detailed quantizing. They work in conjunction with differencing, amplifying, and digital logic circuits to achieve 12-bit resolution. It is essential that the input signal remain constant during the course of the conversion, or very significant errors can result. This requires the use of a sample/hold circuit, as described hereafter. There is also a speed trade-off compared with a full flash design, but the cost is much lower. Still, the size and power requirements of these flash converters usually prevent them from being used in PC-based internal plug-in products. These factors (cost, size, and power) generally limit flash converters to special applications.

Successive approximation converters are the most popular types for general applications. They are readily available in both 12- and 16-bit versions. Maximum sampling speeds in the range of 50 kHz to 1 MHz are attainable. These converters use a single comparator and are thus relatively low cost and simple to construct. An internal D/A converter (described below) systematically produces binary weighted guesses that are compared to the input signal until a "best" match is achieved. A sample/hold is required to maintain a constant input voltage during the course of the conversion.

Integrating converters can provide 12-, 16-, and even 18-bit resolution at low cost. Sampling speeds are typically in the range of 10 to 500 Hz. This converter integrates the unknown input voltage, V_x for a specific period of time T_1. The resulting output e_1 (from the integrator) is then integrated back down to zero by a known reference voltage V_{ref}. The time required T_2 is proportional to V_x:

$$V_x = V_{\text{ref}} \frac{T_2}{T_1}$$

Noise and other signal variations are effectively averaged during the integration process. This characteristic inherently smooths the input signal. The dominant noise source in many data-acquisition applications is the ac power line. By setting the T_1 integration period to a multiple of the ac line frequency, significant noise rejection is achieved. This is known as normal- or series-mode rejection. In addition, linearity and overall accuracy are generally better than with a successive approximation

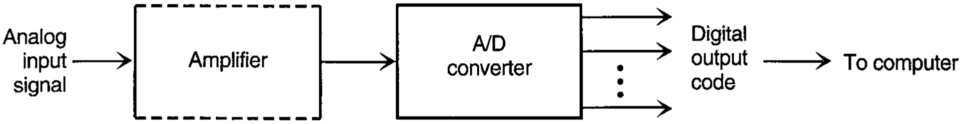


FIGURE 6 Simple analog input stage. *Note*: Amplifier may not be required in every application. (*Intelligent Instrumentation, Inc.*)

converter. These factors make the integrating converter an excellent choice for low-level signals such as thermocouples and strain gages.

Amplifiers

A simple analog input stage is shown in Fig. 6. This circuit can accommodate only one input channel. For multiple channels, several parallel stages can be used. However, the use of a multiplexer to share common resources can provide significant cost savings. This is suggested in Fig. 7. The amplitude of analog input signals can vary over a very wide range. Signals from common transducers are between 50 μV and 10 volts. Yet most A/D converters perform best when their inputs are in the range of 1 to 10 volts. Therefore many systems include an amplifier to boost possible low-level signals to the desired amplitude. Note that adding a fixed-gain amplifier increases sensitivity, but does not increase dynamic range. While it extends low-level sensitivity, it reduces the maximum allowable input level proportionally (the ratio remains constant). Ideally an input amplifier would have several choices of

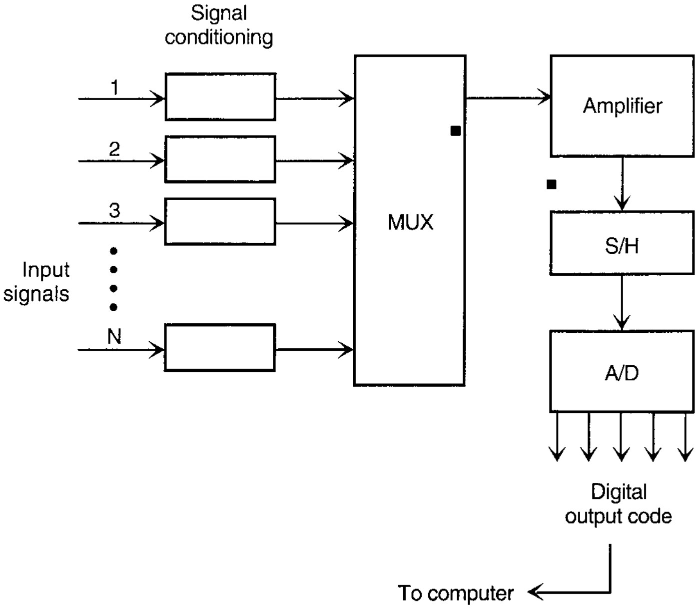


FIGURE 7 Complete analog input subsystem in which a multiplexer (MUX) is used to handle multiple signals. Software can control switches to select any one channel for processing at a given time. Since the amplifier and A/D converter are shared, the speed of acquisition is reduced. To a first approximation, the rated speed of the amplifier and A/D converter will be divided by the number of input channels serviced. Throughput rate is often defined as the per-channel speed multiplied by the number of channels. (*Intelligent Instrumentation, Inc.*)

gain, all under software control. Such a device is called a programmable gain amplifier (PGA). Gains of 1, 2, 4, and 8 or 1, 10, 100, and 200 are common. Unlike a fixed amplifier, a PGA can increase the system's effective dynamic range. For example, consider the following:

- A 12-bit converter (which can resolve one part in 4096) with a full-scale range of 10 volts
- A PGA with gains of 1, 10, and 100—initially set to 1
- An input signal that follows a ramp function from 10 volts down toward zero

Under these conditions, the minimum detectable signal will be 2.4 mV (10 volts ÷ 4096). However, software can be written to detect when the signal drops below 1 volt and to reprogram the PGA automatically for $G = 10$. This extends the minimum sensitivity to 240 μV while increasing a 0.999-volt signal to 9.99 volts. As the signal continues to drop (below 0.1 volt), the PGA can be set for $G = 100$, extending the sensitivity to 24 μV. The effect of making these corresponding changes in gain, to track signal variations, is to increase the *system's* dynamic range. Dynamically adding a gain of 100 is like adding close to 7 bits to the converter. Thus a 12-bit A/D converter can support a range of more than 18 bits. This technique is restricted to systems that have a PGA and to applications where the sample rate is slow enough to permit the time required for the autoranging function.

Most PGAs are differential input instrumentation amplifiers. They present a very high input impedance at both their + and − terminals. The common-mode rejection characteristic of this type of amplifier can attenuate the effects of ground loops, noise, and other error sources. Thus differential inputs are especially useful for measuring low-level signals. Most analog input systems have provisions for configuring the input multiplexer and amplifier for either single-ended or differential use.

Single-Ended versus Differential Signals

Single-ended inputs all share a common return or ground line. Only the high ends of the signals are connected through the multiplexer to the amplifier. The low ends of the signals return to the amplifier through the system ground connections, that is, both the signal sources and the input to the amplifier are referenced to ground. This arrangement works well for high-level signals when the difference in ground potential is relatively small. Problems arise when there is a large difference in ground potentials. This is usually caused by current flow (a ground loop) through the ground conductor. This is covered in further detail in the next article of this handbook section.

A differential arrangement allows both the noninverting (+) and the inverting (−) inputs of the amplifier to make direct connections to both ends of the actual signal source. In this way any ground-loop-induced voltage appears as a common-mode signal and is rejected by the differential properties of the amplifier. While differential connections can greatly reduce the effects of ground loops, they consume the equivalent of two single-ended inputs. Thus a 16-channel single-ended system can handle only eight differential inputs.

Ideally the input impedance, common-mode rejection, and bandwidth of the system's amplifiers would be infinite. In addition, the input currents and offset voltage would be zero. This would provide a measuring system that does not load or alter the signal sources. In contrast, real amplifiers are not perfect. Offset voltage V_{os} appears as an output voltage when the inputs are short-circuited (input voltage is zero). V_{os} can sometimes be compensated for in software. Input (bias) current can be more of a problem. This is the current that flows into (or out of) the amplifier's terminals. The current interacts with the signal source impedance to produce an additional V_{os} term that is not easy to correct. A resistive path must be provided to return this current to ground. It is necessary that the resistance of this path be small enough so that the resulting V_{os} (bias current × source resistance) does not degrade the system's performance significantly. In the extreme case where the inputs are left floating (no external return resistance), the amplifier is likely to reside in a nonlinear or otherwise unusable state. As a general rule, single-ended inputs do not require attention to the bias current return resistance. This is because the path is often provided by the signal source. In contrast, differential connections almost always require an external return resistor. Normally the system's termination panel will provide these resistors. Typically, values of 10 or 100 kΩ are used.

TABLE 1 Relationship between CMRR Expressed in dB and Common-Mode Error in LSB for Hypothetical Input System

Signal gain	Common-mode rejection ratio, dB	Error_{CM} absolute, LSB	$\text{Error}_{\text{CM}}/\text{V}_{\text{CM}}$, LSB/V
1	80	0.4	0.04
10	90	1.3	0.13
100	100	4.1	0.41
1000	110	13	1.3

Common-Mode Rejection

The ability of a differential-input amplifier to discriminate between a differential mode (desired input signal) and a common mode (undesired signal) is its common-mode rejection ratio (CMRR), expressed in decibels (dB). For a given amplifier the CMRR is determined by measuring the change in output that results from a change in common-mode input voltage. CMRR (dB) $= 20 \log(dV_{\text{out}}/dV_{\text{CM}})$. In a data-acquisition system the output signal from the input amplifier includes any error due to the finite CMRR. The A/D converter cannot discriminate between true and error portions of its input signal. Thus the relationship between the magnitude of the error and the sensitivity of the A/D converter is significant. This sensitivity is often referred to as the A/D converter's resolution, or the size of its least-significant bit (LSB). If the error exceeds 1 LSB, the A/D converter responds. As suggested, CMRR is used to measure the analog output error produced by a common-mode input. This makes sense because it is the ratio of two analog signals. However, a complete data-acquisition system (including an A/D converter) has a digital output. Therefore it is more meaningful to express the system's common-mode error in terms of LSBs. This is done by dividing the common-mode error voltage dV_{out} by the sensitivity of the A/D converter (1 LSB on the given range). Amplifier gain G_{diff} must be taken into account. Sensitivity is equal to the converter's full-scale range (FSR) divided by its resolution (number of steps):

$$\text{Error}_{\text{CM}}(\text{LSB}) = \frac{dV_{\text{CM}} \cdot G_{\text{diff}} \cdot 10^{-\text{CMRR}/20}}{\text{FSR}/\text{resolution}} \tag{3}$$

Table 1 shows the relationship between CMRR in decibels and common-mode error in least-significant bits for a hypothetical input system. A 12-bit A/D converter on a 10-volt range (0 to 10 volts or ±5 volts) is assumed in this comparison. A 10-volt common-mode signal is applied to the short-circuited (connected together) input terminals of the system. Dividing the common-mode error (in LSB) by the common-mode voltage yields a direct (useful) figure of merit for the complete data-acquisition system, not just the input amplifier.

Note that independently increasing (improving) CMRR at a *given* gain improves the system's performance. However, the increase in CMRR that accompanies an *increase* in gain actually results in a decrease of the system's overall accuracy. This is because the increase in CMRR (note the log relationship) has less effect than the direct increase in common-mode error ($dV_{\text{CM}} \cdot G_{\text{diff}}$).

Sample/Hold System

There is a distinct time interval required to complete a given A/D conversion. Only the integrating converter can tolerate input amplitude changes during this period. For the other converters a detectable change will result in significant errors. In general, an analog signal can have a continuously changing amplitude. Therefore, a sample/hold (S/H) is used as a means of "freezing" or holding the input constant during the conversion period. Fundamentally the S/H consists of a charge storage device (capacitor) with an input switch. When the switch is closed, the voltage on the capacitor tracks the input signal (sample mode). Before starting an A/D conversion, the switch is opened (hold mode), leaving the last instantaneous input value stored on the capacitor. This is maintained until the conversion is complete. In all practical applications, both successive approximation and subranging converters must use an S/H at their inputs. While a full flash A/D converter does not normally require an S/H, there are applications where it will improve its spurious-free dynamic range.

Multiplexers

The multiplexer (MUX) is simply a switch arrangement that allows many input channels to be serviced by one amplifier and A/D converter (Fig. 7). Software or auxiliary hardware can control this switch to select any one channel for processing at a given time. Because the amplifier and A/D converter are shared, the channels are read sequentially, causing the overall speed of the system to be reduced. To a first approximation, the rated speed of the amplifier and A/D converter will be divided by the number of input channels serviced. The throughput rate is defined as the sample rate (per-channel speed) multiplied by the total number of channels.

The user must be careful not to be misled by the speed specifications of the individual components in the system. Conversion time defines only the speed of a single A/D conversion. Software overhead, amplifier response time, and so on, can greatly reduce a system's performance when reading multiple channels.

In an ideal system all of the input channels would be read at the same instant in time. In contrast, multiplexing inherently generates a "skew," or time difference, between channels. In some cases the system may be fast enough to make it "appear" that the channels are being read at the same time. However, some applications are very sensitive to time skew. Given the fastest A/D converters available, there are still many applications that cannot tolerate the time difference between readings resulting from multiplexing. In critical applications the technique of *simultaneous* S/H can reduce time skew by a factor of 100 to 1000 (Fig. 8).

The simultaneous S/H architecture is ideal when the phase and time relationships of multiple input channels are critical to a given application. For example, assume the system in Fig. 7 is sequentially scanning four analog inputs at a throughput rate of 100 kHz. The elapsed time between conversions would be 10 μs. About 40 μs would be required to digitize all four channels. If the input signals are each 10 kHz sine waves, there will be an apparent phase shift of 144° between the first and fourth channels (40 μs/100 μs $\cdot$ 360°). In contrast, the simultaneous S/H system in Fig. 8 can capture all four channels within a few nanoseconds of each other. This represents a phase shift of about 0.01°.

This technique is particularly useful for preserving time and phase relationships in applications where cross-correlation functions must be calculated. Prime examples include speech research,

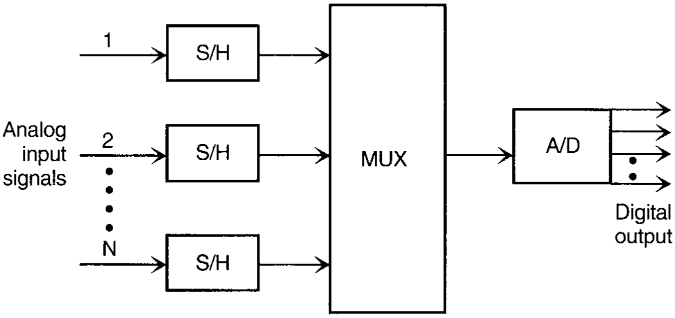


FIGURE 8 Simultaneous sample/hold system. The function of the S/H system is to grab the present value of the signal just before the beginning of an A/D conversion. This level is held constant, despite a changing input, until the A/D conversion is complete. This feature allows the accurate conversion of high-frequency signals. (*Intelligent Instrumentation, Inc.*)

materials and structural dynamics testing, electrical power measurements, geophysical signal analysis, and automatic test equipment (ATE) on production lines.

Analog Signal Conditioning

Analog input systems, based on the components described previously, are representative of most PC-based plug-in products. These boards are usually designed to accept voltage inputs (only) in the range of perhaps ±1 mV to ±10 V. Other signal ranges and signal types generally require preprocessing to make them compatible. This task is known as signal conditioning. The type of conditioning used can greatly affect the quality of the input signal and the ultimate performance of the system. Signal conditioning can include current-to-voltage conversion, surge protection, voltage division, bridge completion, excitation, filtering, isolation, and amplification. The required components can be physically located either remotely, at the signal source (the transducer), or locally, at the data-acquisition board (the host PC). Remote applications use *transmitters* that include the required components. They generally deliver high-level conditioned signals to the data-acquisition system by means of a twisted-pair cable. Local transducers usually connect directly to termination panels that include the required components.

Filtering

Of all the signal-conditioning categories, filtering is the most widely needed and most widely misunderstood. Simply stated, filtering is used to separate desired signals from undesired signals. Undesired signals can include ac line frequency pickup and radio or TV station interference. All such signals are referred to here as *noise*. Filtering can be performed, prior to the A/D conversion, using "physical" devices consisting of resistors, capacitors, inductors, and amplifiers. Filtering can also be accomplished, after conversion, using mathematical algorithms that operate on the digital data within the PC. This is known as digital signal processing (DSP).

Averaging is a simple example of DSP. It is a useful method for reducing unwanted data fluctuations. By averaging a series of incoming data points, the signal-to-noise ratio can be effectively increased. Averaging will be most effective in reducing the effects of random nonperiodic noise. It is less effective in dealing with 50- or 60-Hz or other periodic noise sources. When the desired signal has lower frequency components than the error sources, a low-pass filter can be used. This includes the case where the "real" input signal frequency components can equal, or exceed, half the sampling rate. Here the filter is used to prevent sampled-data aliasing. Aliasing results in the generation of spurious signals within the frequency range of interest that cannot be distinguished from real information. Hence serious errors in the interpretation of the data can occur. Noise-filtering techniques, whether implemented in hardware or software, are designed to filter specific types of noise. In addition to low-pass filters, high-pass and notch (band-reject) filters also can be used. For example, if the frequency band of interest includes the ac line frequency, a notch filter could be used to selectively remove this one component.

Signal termination panels are available that have provisions for the user to install a variety of filters. The most common types of filters are represented by the one- and two-pole *passive* filters shown in Fig. 9. The main difference between these passive filters and the *active* filters, mentioned hereafter, is the addition of amplifiers. Figure 9*b* is an example of an effective single-ended double-pole circuit to attenuate 50/60-Hz noise. The filter has a −6-dB cutoff at about 1 Hz while attenuating 60 Hz about 52 dB (380 times).

Figure 10 suggests a differential two-pole low-pass filter. In contrast to the circuits in Fig. 9, this can be used in balanced applications. Note that any mismatch of the attenuation in the top and bottom paths will result in the generation of a differential output signal that will degrade the system's common-mode rejection ratio. Therefore the resistors and capacitors should be matched carefully to each other. If it is given that all of the resistors and capacitors are of equal values, the pole position f_1

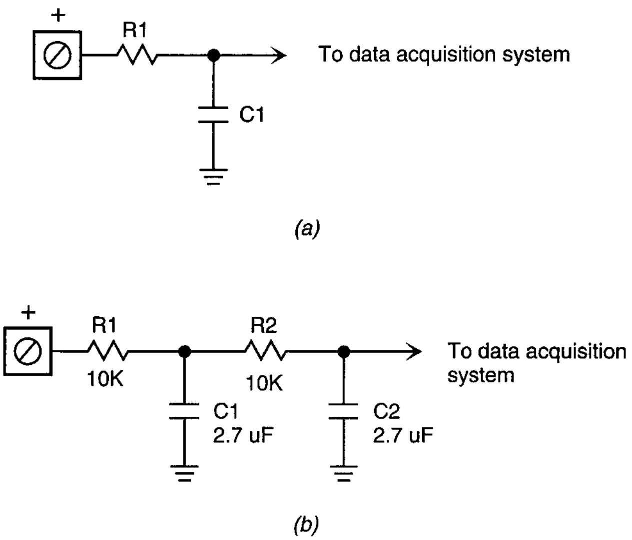


FIGURE 9 Low-pass filters. (*a*) One pole. (*b*) Two poles. (*Intelligent Instrumentation, Inc.*)

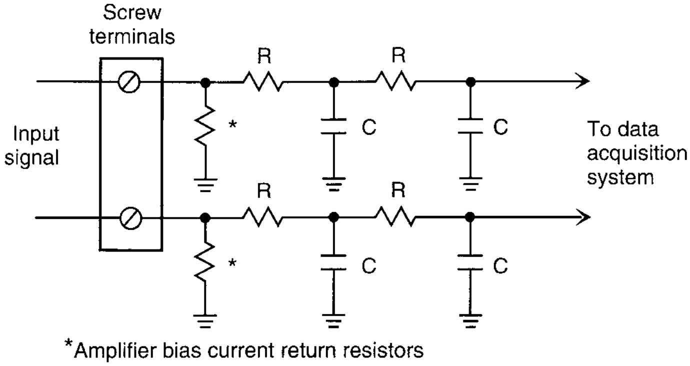


FIGURE 10 Two-pole differential low-pass filter. (*Intelligent Instrumentation, Inc.*)

for this *differential* two-section filter is

$$f_1 = \frac{0.03}{R \cdot C} \tag{4}$$

and the approximate attenuation ratio ($r = V_{\text{in}}/V_{\text{out}}$), at a given frequency f_x, is

$$r = \left(\frac{f_x \cdot R \cdot C}{0.088} + 1\right)^2$$

$$= \left(\frac{0.3 f_x}{f_1} + 1\right)^2 \tag{5}$$

$$\text{dB} = 20 \log r$$

The equations for a single-ended single-pole filter are

$$f_1 = \frac{0.159}{R \cdot C}$$

$$r = \sqrt{\left(\frac{f_x \cdot R \cdot C}{0.159}\right)^2 + 1} \tag{6}$$

and also

$$r = \sqrt{\left(\frac{f_x}{f_1}\right)^2 + 1}$$

The foregoing equations assume that the source impedance is much less than *R* and that the load impedance is much larger than *R*.

For filter applications, monolithic ceramic-type capacitors have been found to be very useful. They possess low leakage, have low series inductance, have very high density (small in size for a given capacitance), and are nonpolarized. Values up to 4.7 μF at 50 volts are commonly available.

In the ideal case a perfect low-pass filter could be built with infinite rejection beyond its cutoff frequency f_1. This would allow f_1 to be set just below one-half the sampling rate, providing maximum bandwidth without danger of aliasing. However, because perfect filters are not available, some unwanted frequencies will "leak" through the filter. Nyquist requires that the "2 times" rule be applied to the highest frequency that can be resolved by the A/D converter. This may not be the same as the highest frequency of interest f_1. Therefore the margin between the highest frequency of interest and the sampling rate must be adjusted. This could involve increasing the sampling rate or possibly forcing a reduction in signal bandwidth. In applications using simple passive filters, the attenuation rate might only be -20 to -40 dB per decade (one and two poles, respectively). This could require the sampling frequency to be 10 to 1000 times the filter corner frequency. The exact factor depends on the resolution of the A/D converter and the amplitude of the highest frequency component. Using high-order active filters (seven to nine poles) might require a factor of only 1.5 to 3 (relative to the original 2× Nyquist rule).

Complete ninth-order elliptic designs are available in a number of configurations. These filters have very steep rolloff (approximately –100 dB per octave), while maintaining nearly constant gain in the passband (±0.2 dB is common). In selecting elliptic filters care must be taken to choose a unit that has a stopband attenuation greater than the resolution of the system's A/D converter. For example, a 12-bit converter has a resolution of one part in 4096, which corresponds to 72 dB. The filter used should attenuate all undesired frequencies by more than 72 dB. Likewise, a 16-bit converter would require a 96-dB filter. Fixed-frequency filter modules, as well as switch- and software-programmable units, can be purchased from various manufacturers. Several of these modular filters can be installed directly on the system's input termination panels. Complete programmable filter subsystems are also available in the form of boards that plug directly into an expansion slot within the data-acquisition PC. The advantage of high-order active filters in antialiasing applications is now clear. Yet while offering excellent performance, they are physically large and expensive compared to simpler filters.

In summary, filtering is intended to attenuate *unavoidable* noise and to limit bandwidth to comply with the Nyquist sampling theorem which prevents aliasing. An antialiasing filter must be a physical analog filter. It cannot be a digital filter that operates on the data after A/D conversion. Noise suppression can often be accomplished, or assisted, with either an analog or a digital filter. Still, filters are not intended as substitutes for proper wiring and shielding techniques. Ground loops, along with capacitively or inductively coupled noise sources, require special attention.

Analog Signal Scaling

As indicated above, most A/D converters are designed to operate with high-level input signals. Common A/D conversion ranges include 0 to 10, ±5, and ±10 volts. When the maximum input signal

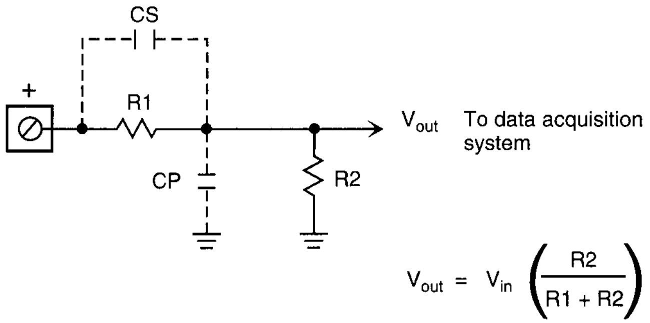


FIGURE 11 Resistive voltage divider to reduce large analog input signals to below 10 volts. (*Intelligent Instrumentation, Inc.*)

is below 1 volt, accuracy is degraded. Under these circumstances it is often appropriate to amplify the signal before the A/D converter. Some A/D converter boards have amplifiers built in. If needed, external amplifiers can be added as part of the signal-conditioning circuitry on the termination panels. In addition to an input signal being too small, it is possible that it might be too large. Remember that most converters accept a maximum of 10 volts at their input. Signals could be 12, 48, or 100 volts (or more). Fortunately it is a simple matter to reduce excessive levels with a resistive voltage divider network. Figure 11 is appropriate for most analog signals. In selecting R_1 and R_2 there are practical factors to consider. Making R_1 large can introduce limitations on signal bandwidth, due to the low-pass filter produced by R_1 and the parasitic capacitance C_p in parallel with R_2. In some applications the network bandwidth can be extended by placing a capacitor C_s across R_1. The value should be selected to make the time constant $R_1 \cdot C_s$ equal to $R_2 \cdot C_p$. The equation assumes that the source (signal) impedance is very low compared with the series combination of R_1 and R_2, that is, $R_1 + R_2$. From this perspective, R_1 and R_2 should be as large as possible.

Input Buffering

The input characteristics of most data-acquisition boards are suitable for general applications. Yet in some cases the input resistance is too low or the bias current too high to allow accurate measurements. Input capacitance can also be an important factor. This is because some transducers, including piezoelectric and pH cells, exhibit a very high output impedance. Under these conditions, direct connection to the data-acquisition system can cause errors. These applications can be satisfied by adding a high input impedance buffer amplifier to the signal-conditioning circuitry. Figure 12 suggests one type of buffer circuit that can be used.

Resistance Signals

Resistance signals arrive at the data-acquisition system from primary sensors, such as strain gages and RTDs. Resistance is changed to a voltage by exciting it with a known current ($V_{\text{out}} = IR$). Figure 13 shows the simplest way to measure resistance. As suggested in Fig. 13*a*, the parasitic (unwanted) resistance of the two connecting lead wires can introduce significant errors. This is because the excitation current flows through the signal measurement wires. Figure 13*b* uses four connecting wires and a differential input connection to the data-acquisition system to minimize the lead-wire effects. This is known as a four-terminal (or kelvin) measurement. The extra wires allow the direct sensing of the unknown resistance. In both cases the resistance of the wires going to the data-acquisition system has little effect. This is because very little current flows in these leads. However, this technique is not well suited to RTD and strain-gage applications because of the very small change in measured voltage

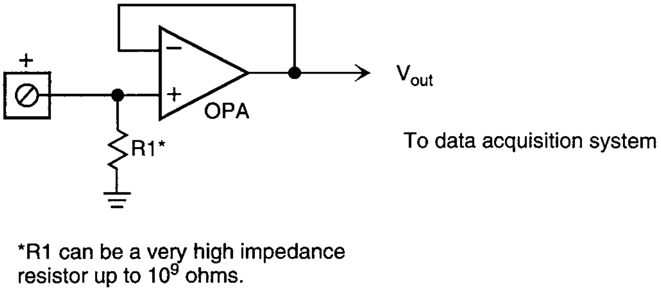
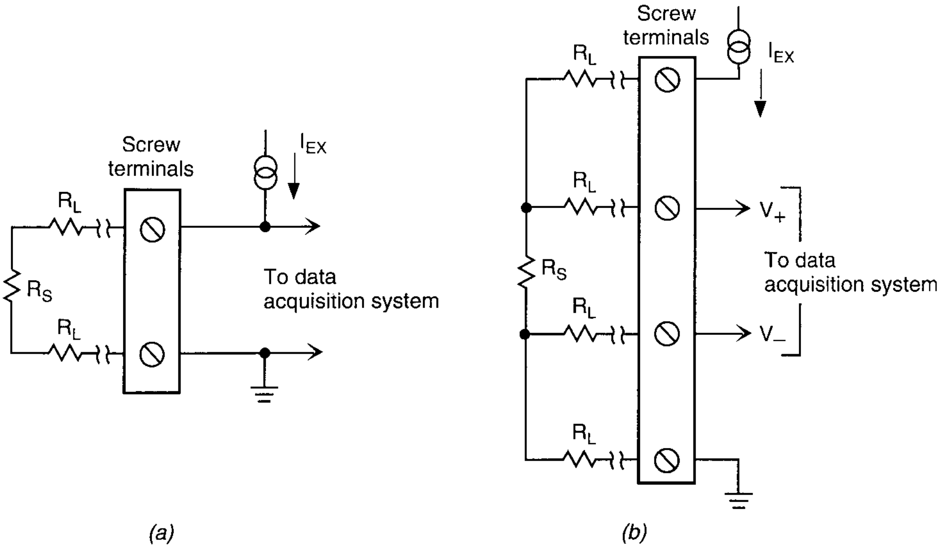


FIGURE 12 High-input impedance buffer circuit. (*Intelligent Instrumentation, Inc.*)



(a) (b)

FIGURE 13 Measurement system for resistive device. R_L—lead-wire resistance. (*Intelligent Instrumentation, Inc.*)

compared to the steady-state (quiescent) voltage. The large quiescent voltage prevents the use of an amplifier to increase the measurement sensitivity.

It is usually better to measure resistive sensors as part of a Wheatstone bridge. A bridge is a symmetrical four-element circuit that enhances the system's ability to detect small changes in the sensor. In Fig. 14 the sensor occupies one arm of the bridge. The remaining arms are completed with fixed resistors equal to the nominal value of the sensor. In general, however, the sensor can occupy one, two, or four arms of the bridge, with any remaining arms being filled with fixed resistors. Note that the differential output from the bridge is zero when all of the resistors are equal. When the sensor changes, an easily amplified signal voltage is produced. Bridge-completion resistors should be of very high precision (typically 0.05 percent). However, stability is actually more important. Initial inaccuracies can be calibrated out, but instability always appears as an error.

A 100-ohm platinum RTD can be used to compare the merits of the approaches in Figs. 13 and 14. To control internal self-heating, the excitation level will be limited to 2 mA. Given that the sensitivity of this type of device is about +0.4 Ω/°C, the output will be about 0.8 mV/°C. This is indeed a small

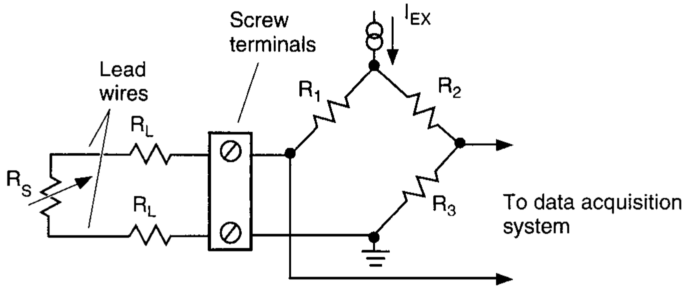


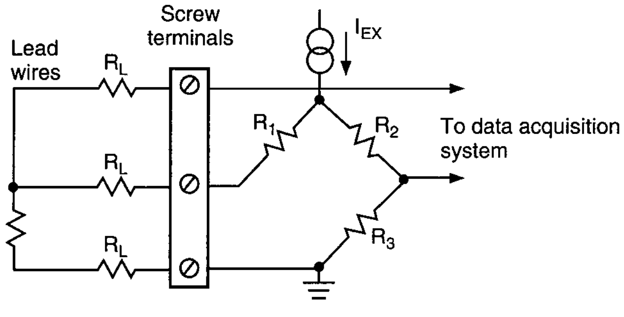
FIGURE 14 Two-wire bridge configuration. (*Intelligent Instrumentation, Inc.*)

signal that will require amplification. It would be useful to multiply the signal by 100 to 1000 to make best use of the A/D converter's full-scale range (typically 5 or 10 volts). However, the quiescent voltage across the RTD is 2 mA · 100 Ω = 0.2 volt. In Fig. 13 this limits the maximum gain to 10. Thus in a 12-bit system, the smallest detectable temperature change will be about 0.5°C. In contrast, the bridge circuit in Fig. 14 balances out the fixed or quiescent voltage drop, allowing greater magnification of the difference signal. This allows the detection of changes as small as 0.005°C.

Figure 14 has the same lead-wire resistance problem that the simple circuit had. The lead-wire resistances are indistinguishable from the transducer's resistance. So while this bridge circuit has high sensitivity, it is not suitable for precision applications.

Figure 15 shows a means of correcting for lead-wire effects. While the three-wire bridge requires an additional wire to be run to the sensor, several very important advantages are gained. If the (reasonable) assumption is made that the two wires bringing current to the sensor are of the same material and length, many of the potential error terms cancel. Comparing Figs. 14 and 15 shows that the additional signal wire has moved the measurement point directly to the top of the sensor. Again, the resistance of this wire is not important because current flow to the data-acquisition system is very low. Moving the measurement point has the effect of locating one current-carrying lead resistance in the top arm of the bridge while the other remains in the lower arm. The current in each is the same, so their voltage drops tend to cancel. While the compensation is not perfect, it does offer a significant performance improvement.

Transducer excitation and bridge-completion components are normally installed on the system's signal termination panels. While both voltage and current excitation can be used, current excitation is



R_L is lead-wire resistance

FIGURE 15 Three-wire bridge configuration. (*Intelligent Instrumentation, Inc.*)

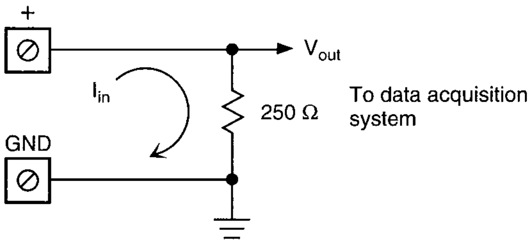


FIGURE 16 4- to 20-mA input conversion circuit (single-ended). (*Intelligent Instrumentation, Inc.*)

generally more desirable. This is because current excitation provides a more linear output response, making the data interpretation easier.

Current Conversion

The need to measure signals in the form of currents is quite common in a data-acquisition system. The outputs from remote transducers are often converted to 4- to 20-mA signals by two-wire transmitters. At the data-acquisition system, current is easily converted back to a voltage with a simple resistor. Figure 16 shows how this is done. For a 4- to 20-mA signal a resistor value of 250 ohms can be used to provide a voltage output of 1 to 5 volts. As a general rule, the largest resistor that does not cause an overrange condition should be used. This ensures the maximum resolution. Stability of the resistor is essential, but the exact value is not important. Most systems have software provisions for calibrating the measurement sensitivity of each channel at the time of installation. Low-cost 0.1 percent metal-film resistors are usually adequate. For larger currents it is a simple matter to scale the resistor down to yield the desired full-scale voltage ($R = V/I$, where V is the full-scale voltage range and I is the maximum current to be read).

The technique of using only a resistor to convert from current to voltage does have limitations. If, for example, a 1-μA level is to be measured, a resistor of approximately 5 MΩ will be required. Unfortunately the use of high-value resistors leads to potentially large errors due to noise and measuring system loading. As suggested before, the data-acquisition system has a small but finite input current. This bias current (typically around 10 nA) will also flow through the conversion resistor and will be indistinguishable from the signal current. Therefore when very low currents must be measured, a different technique is used. Figure 17 suggests an active circuit that utilizes a precision field effect transistor (FET) amplifier to minimize the bias current problem. Both the simple resistor and the FET amplifier circuits require the same resistor value for a given current level. However, in the latter case the data-acquisition system's bias current is supplied by the amplifier and does not affect the measurement accuracy. A wide range of low-bias-current amplifiers are available for special applications. With the amplifier shown, currents as low as 10 pA can be read reliably.

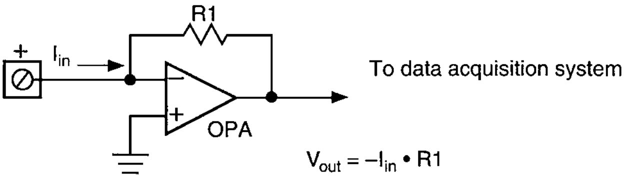


FIGURE 17 Current-to-voltage converter circuit suitable for very low current levels. (*Intelligent Instrumentation, Inc.*)

Transmitters

When low-level signals (below 1 volt) are located remotely from the data-acquisition system, special precautions are suggested. Long wire runs with small signals usually result in poor performance. It is desirable to preamplify these signals first to preserve maximum signal-to-noise ratio. Two-wire transmitters provide an ideal way of packaging the desired signal-conditioning circuitry. In addition to signal amplification, transmitters can also provide filtering, isolation, linearization, cold-junction compensation, bridge completion, excitation, and conversion to a 4- to 20-mA current. Transmitters are ideal for thermocouples, RTDs, and strain gages. Current transmission allows signals to be sent up to several thousand feet (1500 meters) without significant loss of accuracy. While voltage signals are rapidly attenuated by the resistance of the connecting wires, current signals are not. In a current loop, the voltage drop due to wire resistance is compensated by the compliance of the current source, that is, the voltage across the current source automatically adjusts to maintain the desired current level. Note that power for the transmitter is conveyed from the data-acquisition system over the same two wires that are used for signal communications. No local power is required.

In addition to analog transmitters, there are also digital devices. These provide most of these features, except that the output signal is in a digital form instead of 4 to 20 mA. The output protocol is usually a serial data stream that is RS-232, RS-422, or RS-485 compatible. This is accomplished by including an A/D converter and a controller (computer) inside the transmitter. In many cases the output signal can be connected directly to a serial port on the PC without additional hardware. Two possible disadvantages of a digital transmitter are that it requires local power and, because of the added complexity, is generally more expensive.

Surge Protection

When a system can be subjected to unintentional high-voltage inputs, it is prudent to provide protection to avoid possible destruction of the equipment. High-voltage inputs can be induced from lightning, magnetic fields, static electricity, and accidental contact with power lines, among other causes.

Figure 18 suggests two different protection networks. Both circuits offer transient (short- duration) as well as steady-state protection. The circuit in Fig. 18*a* can tolerate continuous inputs of up to about 45 volts. When the overload disappears, the signal path automatically returns to normal. The circuit in Fig. 18*b* protects against continuous overloads of up to about 280 volts. In contrast, sustained

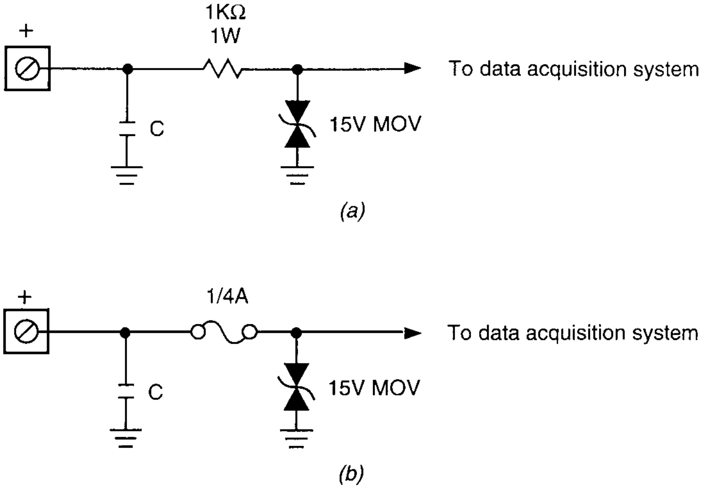


FIGURE 18 Representative input protection networks. (*Intelligent Instrumentation, Inc.*)

overloads to this circuit will cause the fuse to open (protecting the protection circuit). A disadvantage of this network is that the fuse must be replaced before the signal path is active again. In either case, signal flow is interrupted during the overload period. The resistor (or fuse) and the metal-oxide varistor (MOV) form a voltage clamp to ensure that transients will not get to the input of the data-acquisition system. MOVs are semiconductor devices that can react very quickly to absorb high-energy spikes. The 15-volt rating shown is high enough to pass all normal signals, but low enough to protect the data-acquisition system's input. Consideration should be given to the fact that even below the MOV's threshold voltage a small leakage current flows. If the series R is too large, the leakage could appear as a significant temperature-dependent error voltage (IR).

The optional capacitor can help suppress high-frequency transients. In some applications it must be rated for high voltage. For example, transients in power stations or other noisy environments can exceed 1000 volts. The capacitance value should be as large as physically possible, and the capacitor should be positioned as close as possible to the signal entry point of the system. Capacitors with low series impedance at high frequencies should be selected. This requirement eliminates electrolytic-type capacitors. If the input signal can change polarity, polarized capacitors must be avoided.

Analog Isolation

Isolators can be used to protect people and equipment from contact with high voltage. They usually provide the same protection as MOVs with the addition of one very important extra feature. Isolators can block overloads (protect) while simultaneously passing a desired signal. Applications include the breaking of ground loops, patient monitoring, and the removal of large common-mode signals. For example, if a thermocouple is connected to a motor winding, it could be in contact with 240 volts ac. Yet the thermocouple output voltage might be only 30 mV. The 30 mV (the actual signal) is seen as a differential signal while the 240 volts appears as a common-mode signal. The isolator operates in a way that is similar to a differential amplifier (described earlier). Its common-mode rejection capabilities block the effects of the unwanted portion of the signal. While standard differential amplifiers are generally limited to a ±10-volt common-mode signal, isolators are available with ratings beyond 5000 volts.

A family of industry-standard 5B signal-conditioning modules is available. These complete plug-in units are designed to provide a wide range of input and output capabilities. Each module supports a single channel, allowing the flexibility to mix the various types when configuring a system. Isolation, rated at 1500 volts, provides high-voltage separation between the signals and the data-acquisition system. Input modules are available for most voltage ranges, current ranges, thermocouples, RTDs, and strain gages. All of the required conditioning functions are included: protection, filtering, linearization, cold-junction compensation, bridge completion, and excitation. Output modules support 4- to 20-mA current loops. Standard termination panels accommodate up to 16 modules.

ANALOG OUTPUTS

Digital-to-Analog Converters

Analog outputs are required in many test and industrial automation applications. For example, they can be used to generate inputs (stimuli) to a device under test and to operate valves, motors, and heaters in closed-loop feedback control systems. A D/A converter is used to transform the binary instructions from the digital computer (PC) to a variable output level. Common analog output ranges include ±5, ±10, and 0 to 10 volts and 4 to 20 mA.

A popular type of D/A converter consists internally of N binary weighted current sources. The values (the levels can be scaled to suit speed and output requirements) correspond to 1, $^1/_2$, $^1/_4$, $^1/_8$, . . . , $^1/_2{}^{N-1}$. N is also the number of digital input lines (bits). Each source can be turned on or off independently by the computer. By summing the outputs of the sources together, 2^N current combinations are produced. Thus a 12-bit converter can represent an analog output range with 4096 discrete steps. A current-to-voltage converter is included in voltage output models.

Faithful generation of a complex signal requires that the conversion rate (clock rate) of the D/A converter be very high compared with the repetition rate of the waveform. Ratios of 100 to 1000 points per cycle are common. This suggests that a “clean” 1-kHz output could require a 100-kHz to 1-MHz converter. This is pushing the current state of the art in PC-based data-acquisition products.

When operating in the voltage output mode, most D/A converters are limited to supplying around 5 mA of load current. This implies that most D/A converters will use some kind of signal conditioning when interfacing to real-world devices (transducers). When large loads such as positioners, valves, lamps, and motors are to be controlled, power amplifiers or current boosters need to be provided. Most data-acquisition systems do not include these high-power analog drivers internally.

Output Filtering

A D/A converter attempts to represent a continuous analog output with a series of small steps. The discontinuities inherent in a digitally produced waveform represent very high frequencies. This is seen as noise or distortion that can produce undesired effects. A low-pass filter is often used at the output of a D/A converter to attenuate high frequencies and, thus, “smooth” the steps.

DIGITAL INPUTS AND OUTPUTS

Most data-acquisition systems are able to accept and generate TTL level signals. These are binary signals that are either high or low (on or off). The low state is represented by a voltage near 0 volts (generally less than 0.8 volt), while a high state is indicated by a voltage near 5 volts (generally greater than 2 volts). Levels between 0.8 and 2 volts are not allowed. The output levels are intended to drive other “logic” circuits rather than industrial loads. As a result, drive capabilities are generally under 24 mA. Still, digital signals in many real-world applications are not TTL-compatible. It is common to encounter 24-volt, 48-volt, and 120/240-volt ac levels as digital input signals. High voltage and current outputs are often required to operate solenoids, contactors, motors, indicators, alarms, and relays.

Many types of digital signal termination panels are available to facilitate the connection of field wires to the data-acquisition system. In addition to screw terminals, the panels have provisions for signal conditioning, channel status indicators (such as light-emitting diodes), voltage dividers, and isolators. Thus the monitoring and the control of high dc levels, along with ac line voltage circuits, are readily accomplished.

Pulse and Frequency Inputs and Outputs

A variety of counting, timing, and frequency-measuring applications exists. Other applications require that devices be turned on and off for precise time periods. All of these functions can be provided by counter/timer circuits. The system’s counter/timers are optimized for pulse applications, including frequency measurement and time-base generation. Counters are characterized by the number of input events that can be accumulated and by their maximum input frequency. Several independent counters are usually provided. They can be used to count events (accumulate), measure frequency, measure pulse width, or act as frequency dividers. Counting can be started from a defined initial value, and most counters can be configured to reset automatically to this value after it has been read. Software can easily interpret the counter’s data as a sum or difference from an arbitrary starting point. Pulse generators (rate generators) are software programmable over a very wide range of frequencies and duty cycles. A rate generator is often used to provide the precise time base required for accurate data acquisition. Most systems use 16-bit counters that can accumulate pulses at frequencies up to 8 MHz. Up to 65,536 (2^{16}) events can be accumulated before the counter overflows. Two counters can generally be cascaded to provide 32-bit capability (more than 4 billion counts). Most counters accept only TTL level signals. Other levels require signal conditioning.

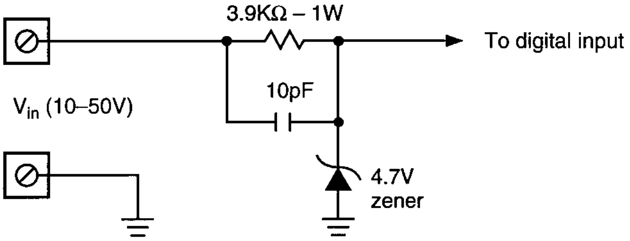


FIGURE 19 Circuit to convert large digital signals to TTL-compatible levels. (*Intelligent Instrumentation, Inc.*)

Frequency measurements using counters can be accomplished in different ways, depending on the application. When the unknown frequency is a TTL signal, it can be applied directly to the counter circuit. Analog signals can be converted to TTL levels with comparator circuits available from some manufacturers. Voltage dividers using resistors, zener diodes, or optoisolators can be used to scale down high-level signals. When using any kind of signal conditioner before a counter input, consideration should be given to possible speed limitations.

Two distinct options exist for measuring high or low frequencies. The first method counts a known clock generator for the period of the unknown input signal. This provides high resolution for low-frequency signals, while minimizing the time required for the measurement. Generally this is used for frequencies below 10 Hz. The second method counts cycles of the unknown input signal for a fixed time interval. The advantage of this technique is that it allows measurements up to the limit of the counter's speed (typically 8 MHz). It is easy to implement an auto-ranging software algorithm that optimizes resolution over a very wide frequency range.

Digital Signal Scaling

For large digital signals, the circuit in Fig. 19 can be used to produce TTL-compatible levels. Most digital circuits (digital input ports and counters) require fast input level transitions to ensure reliable operation. Steps faster than 10 μs are usually adequate. Parasitic capacitance at the input to the data-acquisition system can interact with the series resistor to degrade input steps. The 10-pF capacitor in Fig. 19 is included to help correct this problem. When the input is not fast enough, it can be made TTL-compatible with the Schmitt trigger circuit shown in Fig. 20.

Digital Isolation

A family of industry standard signal-conditioning modules is available. These complete plug-in units are designed to provide a wide range of input and output capabilities. Each module supports a single

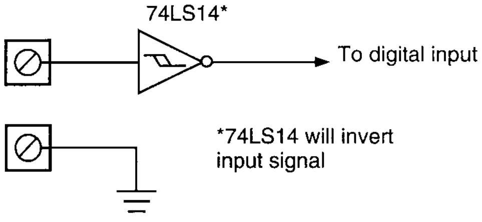


FIGURE 20 Schmitt trigger circuit to "speed up" slow input signals. Input levels must be TTL-compatible. (*Intelligent Instrumentation, Inc.*)

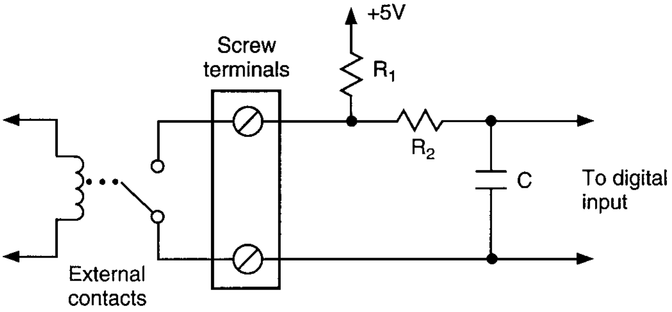


FIGURE 21 Contact sensing and wetting. (*Intelligent Instrumentation, Inc.*)

channel, allowing the flexibility to mix the various types when configuring a system. Optical isolation, rated at 4000 volts, provides high-voltage separation between the signals and the data-acquisition system. This is useful for safety, equipment protection, and ground-loop interruption. Output modules use power transistors or triacs to switch high-voltage high-current ac or dc loads. Loads up to 60 volts dc or 280 volts ac at 3 amperes can be accommodated. Input modules convert digital signals between 10 and 280 volts to TTL levels. Standard termination panels accommodate up to 16 modules.

Contact Sensing

As shown in Fig. 21, contact sensing can be implemented on a signal termination panel. When interfacing to relay or switch contacts, a pull-up current must be provided. The pull-up current converts the opening and closing of the contacts to TTL level voltages. Because all metal surfaces tend to oxidize with time, poor relay contacts can result. Both level generation and contact wetting can be accomplished by connecting a resistor between the input line and the +5-volt power supply. When the switch is open, the input system sees +5 volts. When the switch is closed, the input is 0 volts. This satisfies the TTL requirements of the data-acquisition system. A value of 250 ohms for R_1 will provide 20 mA of wetting current, which is usually enough to keep most contacts free of oxide buildup. R_2 and C_1 function as a very simple debounce filter to reduce erroneous inputs due to the mechanical bouncing of the contacts. Care must be taken to avoid slowing the signal transition so much that false triggering occurs. If needed, a Schmitt trigger can be added, as shown in Fig. 20. Digital filtering techniques can also be used to eliminate the effects of contact bounce.

Relay Driving

Figure 22 shows how a TTL output from a data-acquisition board can be connected to drive an external 5-volt relay coil. The digital output must be able to switch the coil current. The specifications of digital output ports vary considerably between models. However, most can support 16 to 24 mA. When large relays, contactors, solenoids, or motors are involved, an additional driver or intermediate switching network can be used. The diode D_1 protects the internal circuitry against the inductive kickback from the relay coil. Without the diode, the resulting high-voltage spikes will damage the digital port. Note that the direction (polarity) of the diode must be as shown in the diagram. Protection diodes must be able to respond very quickly and absorb the coil's energy safely. Most standard switching diodes fill these needs.

Motor Control

Many different types of motors are in common use today. When it comes to controlling these devices, specialized circuits are often required. Some applications, however, require only on-off operations.

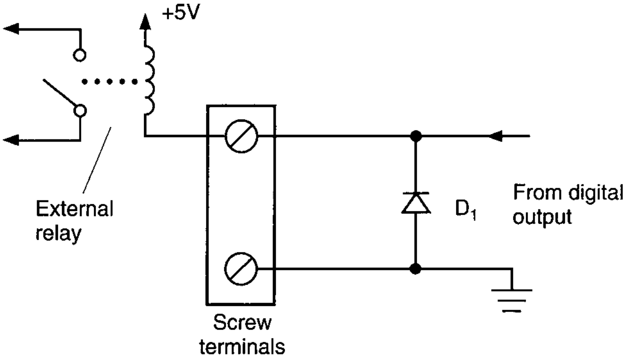


FIGURE 22 Relay driving circuit. (*Intelligent Instrumentation, Inc.*)

These can simply be driven by digital output ports, usually through optical isolators (loads of up to 3 amperes) or with various types of contactors (relays).

In general, when variable speed is desired, either analog or digital outputs from the data-acquisition system are used to manipulate the motor through an external controller. A wide range of both ac and dc controllers is available. Motor controls are discussed in more detail in Section 9 of this handbook.

Stepper-type motors are of particular interest in robotics, process control, instrumentation, and manufacturing. They allow precise control of rotation, angular position, speed, and direction. While several different types of stepping motors exist, the permanent-magnet design is perhaps the most common. The permanent magnets are attached to the rotor of the motor. Four separate windings are arranged around the stator. By pulsing direct current into the windings in a particular sequence, forces are generated to produce rotation. To continue rotation, current is switched to successive windings. When no coils are energized, the shaft is held in its last position by the magnets. In some applications these motors can be driven directly (via opto relays) by one of the data-acquisition system's digital output ports. The user provides the required software to produce the desired pulses in proper sequence. The software burden can be reduced by driving the motor with a specially designed interface device. These units accept a few digital input commands representing the desired speed, rotation, direction, and acceleration. A full range of motors and interfaces is available. Stepper motors are discussed in more detail in Section 9 of this handbook.

NOISE AND WIRING IN DATA SIGNAL HANDLING

by Howard L. Skolnik*

Signals entering a data-acquisition and control system include unwanted noise. Whether this noise is troublesome depends on the signal-to-noise ratio and the specific application. In general it is

---

* Intelligent Instrumentation, Inc. (a Burr-Brown Company), Tucson, Arizona.

desirable to minimize noise to achieve high accuracy. Digital signals are relatively immune to noise because of their discrete (and high-level) nature. In contrast, analog signals are directly influenced by relatively low-level disturbances. The major noise-transfer mechanisms include conductive, inductive (magnetic), and capacitive coupling. Examples include the following:

- Switching of high-current loads in nearby wiring can induce noise signals by magnetic coupling (transformer action).
- Signal wires running close to ac power cables can pick up 50- or 60-Hz noise by capacitive coupling.
- Allowing more than one power or signal return path can produce ground loops that inject errors by conduction.

Conductance involves current flowing through ohmic paths (direct contact), as opposed to inductance or capacitance.

Interference via capacitive or magnetic mechanisms usually requires that the disturbing source be close to the affected circuit. At high frequencies, however, radiated emissions (electromagnetic signals) can be propagated over long distances.

In all cases, the induced noise level will depend on several user-influenced factors:

- Signal source output impedance
- Signal source load impedance (input impedance to the data-acquisition system)
- Lead-wire length, shielding, and grounding
- Proximity to noise source or sources
- Signal and noise amplitude

Transducers that can be modeled by a current source are inherently less sensitive to magnetically induced noise pickup than are voltage-driven devices. An error voltage coupled magnetically into the connecting wires appears in series with the signal source. This has the effect of modulating the voltage across the transducer. However, if the transducer approaches ideal current-source characteristics, no significant change in the signal current will result. When the transducer appears as a voltage source (regardless of impedance), the magnetically induced errors add directly to the signal source without attenuation.

Errors also are caused by capacitive coupling in both current and voltage transducer circuits. With capacitive coupling, a voltage divider is formed by the coupling capacitor and the load impedance. The error signal induced is proportional to $2\pi fRC$, where R is the load resistor, C is the coupling capacitance, and f is the interfering frequency. Clearly, the smaller the capacitance (or frequency), the smaller is the induced error voltage. However, reducing the resistance only improves voltage-type transducer circuits.

Example. Assume that the interfering signal is a 110-volt ac 60-Hz power line, the equivalent coupling capacitance is 100 pF, and the terminating resistance is 250 ohms (typical for a 4- to 20-mA current loop). The resulting induced error voltage will be about 1 mV, which is less than 1 least-significant bit in a 12-bit 10-volt system.

If the load impedance were 100 kΩ, as it could be in a voltage input application, the induced error could be much larger. The equivalent R seen by the interfering source depends on not only the load impedance but also the source impedance and the distributed nature of the connecting wires. Under worst-case conditions, where the wire inductance separates the load and source impedances, the induced error could be as large as 0.4 volt. This represents about an 8-percent full-scale error.

Even though current-type signals are usually converted to a voltage at the input to the data-acquisition system, with a low-value resistor this does not improve noise performance. This is because both the noise and the transducer signals are proportional to the same load impedance.

It should be pointed out that this example does not take advantage of, or benefit from, shielding, grounding, and filtering techniques.

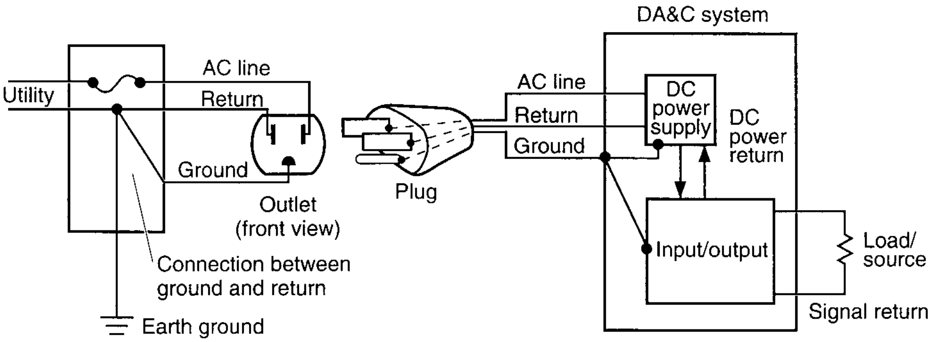


FIGURE 1 Differences between ground and return conductors. (*Intelligent Instrumentation, Inc.*)

GROUNDING AND SHIELDING PRINCIPLES

Most noise problems can be solved by giving close attention to a few grounding and shielding principles:

- Do not confuse the definitions of ground and return paths. Ground = safety; return = current-carrying.
- Minimize wiring inductance.
- Limit antennas.
- Maintain balanced networks wherever possible.

The foregoing directions appear simple, but what really is involved?

For a beginning, one should redefine some common terms. A ground is *not* a signal or power supply return path. A ground wire connects equipment to earth for safety reasons—to prevent accidental contact with dangerous voltages. Ground lines do not normally carry current. Return lines are an active part of a circuit—carrying power or signal currents (Fig. 1). Care should be taken to distinguish between grounds and returns and to avoid more than one connection between the two.

To be effective, return paths should have the lowest possible impedance. Someone once said that the shortest distance between two points is a straight line. But in geography this is not true, and it is not generally true in electronics either. Current does not take the shortest path; rather it takes the path of least resistance (really, of least impedance). Return impedance is usually dominated by the path inductance. Wiring inductance is proportional to the area inside the loop formed by the current-carrying path. Therefore impedance is minimized by providing a return path that matches or overlaps the forward signal path. Note that this may not be the shortest or most direct route. This concept is fundamental to ensuring proper system interconnections.

Three different grounding and connection techniques are suggested in Figs. 2, 3, and 4. The circuit in Fig. 2 allows the signal return line to be grounded at each chassis. This may look like a good idea from a safety standpoint. However, if a difference in potential exists between the two grounds, a ground current must flow. This current, multiplied by the wire impedance, results in an error voltage e_e. Thus the voltage applied to the amplifier is not V_1, but $V_1 + e_e$. This may be acceptable in those applications where the signal voltage is much greater than the difference in the ground potentials.

When the signal level is small and a significant difference in ground potentials exists, the connection in Fig. 3 is more desirable. Note that the return wire is not grounded at the amplifier and ground current cannot flow in the signal wires. Any difference in ground potential appears, to the amplifier, as a common-mode voltage. In most circuits the effects of common-mode voltage are very small, as long as the sum of signal voltage plus common-mode voltage is less than 10 volts. (Ten volts is the linear

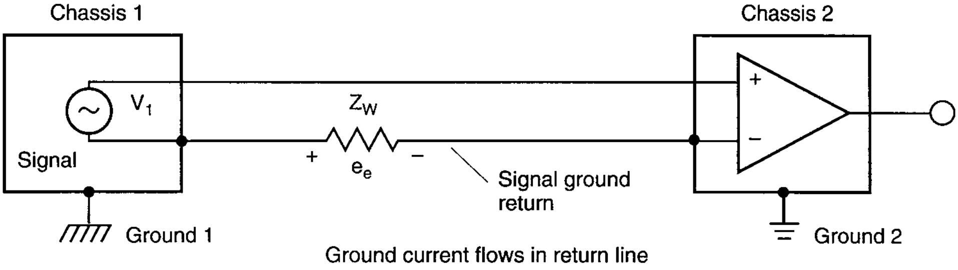


FIGURE 2 Single-ended connection. (*Intelligent Instrumentation, Inc.*)

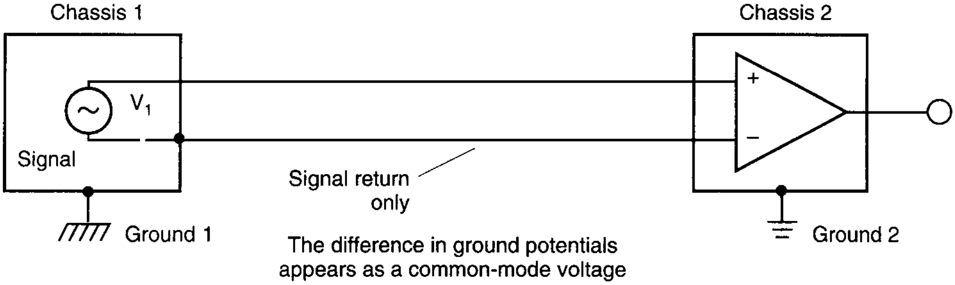


FIGURE 3 Differential connection. (*Intelligent Instrumentation, Inc.*)

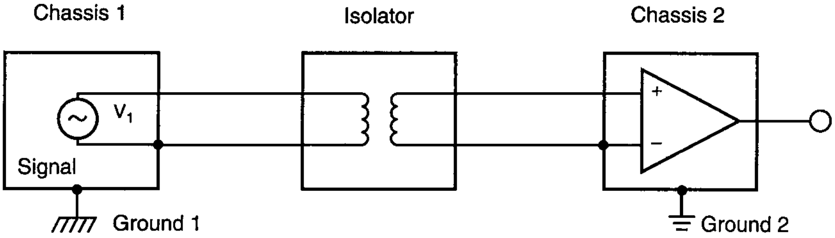


FIGURE 4 Isolated connection. (*Intelligent Instrumentation, Inc.*)

range for most amplifiers.) Additional information about common-mode rejection and single-ended versus differential amplifiers can be found in a prior article in this handbook.

If cost is not a limitation, Fig. 4 offers the highest performance under all conditions. Injecting an isolator into the signal path faithfully conveys V_1 to the amplifier while interrupting all direct paths. In this configuration multiple ground connections can be tolerated along with several hundred volts between input and output circuits. Additional information on both analog and digital isolators can be found in a prior article in this handbook.

Cable Types

What kind of wire should be used to interconnect a system? First, it must be emphasized that a single piece of wire is not generally useful. Circuits consist of complete paths, so pairs of wires are referred to in this discussion. Basically four kinds of wire are fundamental: (1) plain pair, (2) shielded pair,

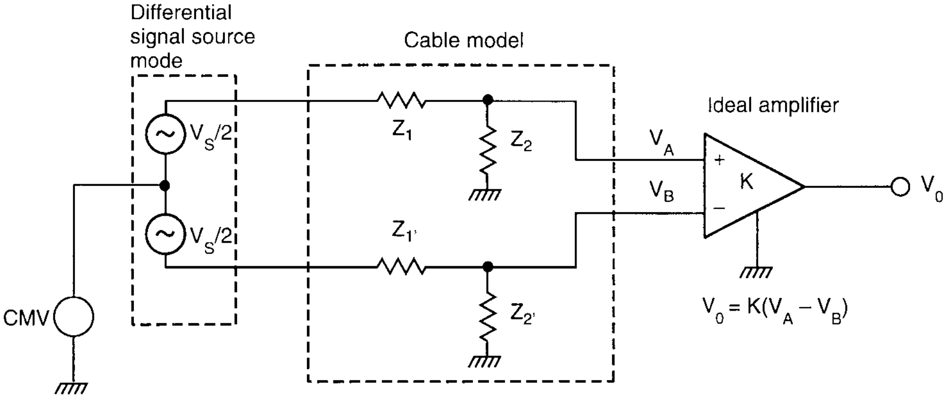


FIGURE 5 Influence of cable connectors on common-mode signal performance. (*Intelligent Instrumentation, Inc.*)

(3) twisted pair, and (4) coaxial cable. All but the coaxial wires are said to be balanced. Coaxial cable differs from the others in that the return line surrounds the central conductor.

Technically, the outer conductor should not be called a shield because it carries signal current. It is significant that the forward and return path conductors do not have exactly the same characteristics. In contrast, a shielded pair is surrounded by a separate conductor (properly called a shield) that does not carry signal current.

Figure 5 suggests a simple model for a differential signal connection. The attributes of the signal source have been split to model the influence of a common-mode voltage. Let us focus on the effect of forward and return path symmetry in the cable. Assuming that the amplifier is perfect, it will respond only to the difference between V_A and V_B. The technique of superposition allows us to analyze each half of the cable model separately and then to add the results. Z_1 is usually dominated by series inductance, while Z_2 is dominated by parallel capacitance. In any case, Z_1 and Z_2 form a voltage divider. If the dividers in both legs of the cable are identical, $V_A - V_B$ will not be influenced by common-mode voltage. If, however, the capacitance represented by Z_2 is different in the two paths, a differential voltage will result and the amplifier will be unable to distinguish the resulting common-mode error from a change in V_S.

Coaxial cable offers a very different capacitance between each of its conductors and ground. Not only does the outer conductor surround the inner, it is also connected to ground. Thus coaxial cable is intended for single-ended applications only. Note that even perfectly balanced cables can still attenuate differential signals.

Sometimes even a single-ended source is best measured with a differential amplifier. Refer again to Fig. 3. To maintain a high rejection of any ground difference potential, balanced cables are required.

TROUBLESHOOTING GUIDE FOR NOISE

One method of reducing errors due to capacitive coupling is to use a shield. Generally there is little that can be done to reduce the actual capacitance. (Wire length and physical location are factors, however.) Nevertheless, placing a conductive material (at ground potential) between the signal wires and the interference source is very useful. The shield blocks the interfering current and directs it to the ground. Depending on how complete the shield is, attenuations of more than 60 dB are attainable. When using shielded wire, it is very important to connect only one end of the shield to ground. The connection should be made at the data-acquisition system end of the cable (such as input amplifier). Connecting both ends of the shield can generate significant error by inducing ground-loop currents.

A shield can work in three different ways:

- Bypassing capacitively coupled electric fields
- Absorbing magnetic fields
- Reflecting radiated electromagnetic fields

Another approach is to use twisted pairs. Twisted-pair cables offer several advantages. Twisting of the wires ensures a homogeneous distribution of capacitances. Capacitances both to ground and to extraneous sources are balanced. This is effective in reducing capacitive coupling while maintaining high common-mode rejection. From the perspective of both capacitive and magnetic interference, errors are induced equally into both wires. The result is a significant error cancellation.

The use of shielded or twisted-pair wire is suggested whenever low-level signals are involved. With low-impedance sensors the largest gage-connecting wires that are practical should be used to reduce lead-wire resistance effects. On the other hand, large connecting wires that are physically near thermal sensing elements tend to carry heat away from the source, generating measurement errors. This is known as thermal shunting, and it can be very significant in some applications.

The previous discussion concentrated on cables making single interconnections. Multiconductor cables, for connecting several circuits, are available in similar forms (such as twisted pairs and shielded pairs). Both round and flat (ribbon) cables are used widely. Because of the close proximity of the different pairs in a multiconductor cable, they are more susceptible to crosstalk. Crosstalk is interference caused by the inadvertent coupling of internal signals via capacitive or inductive means.

Again, twisted pairs are very effective. Other methods include connecting alternate wires as return lines, running a ground plane under the conductors, or using a full shield around the cable.

Still another noise source, not yet mentioned, is that of triboelectric induction. This refers to the generation of noise voltage due to friction. All commonly used insulators can produce a static discharge when moved across a dissimilar material. Fortunately the effect is very slight in most cases. However, it should not be ignored as a possible source of noise when motion of the cables or vibration of the system is involved. Special low-noise cables are available that use graphite lubricants between the inner surfaces to reduce friction.

The key to designing low-noise circuits is recognizing potential interference sources and taking appropriate preventive measures. Table 1 can be useful when troubleshooting an existing system.

After proper wiring, shielding, and grounding techniques have been applied, input filtering can be used to further improve the signal-to-noise ratio. However, filtering should never be relied upon as a fix for improper wiring or installation.

Cable-Length Guidelines

What is the maximum allowable cable length? There is no direct answer to this question. The number of factors that relate to this subject is overwhelming. Signal source type, signal level, cable type, noise source types, noise intensity, distance between cable and noise source, noise frequency, signal frequency range, and required accuracy are just some of the variables to consider. However, experience can yield some “feel” for what often works, as per the following examples:

Analog Current Source Signals. Given 4- to 20-mA signal, shielded wire, bandwidth limited to 10 Hz, required accuracy 0.5 percent, and average industrial noise levels. Cable lengths of 1000 to 5000 feet (300 to 1500 meters) have been used successfully.

Analog Voltage Source Signals. Given ±1- to ±10-volt signal, shielded wire, bandwidth limited to 10 Hz, required accuracy 0.5 percent, and average industrial noise levels. Cable lengths of 50 to 300 feet (15 to 90 meters) have been used successfully.

Analog Voltage Source Signals. Given 10-mV to 1-volt signal, shielded wire, bandwidth limited to 10 Hz, required accuracy 0.5 percent, and average industrial noise levels. Cable lengths of 5 to 100 feet (1.5 to 30 meters) have been used successfully.

Digital TTL Signals. Given ground-plane-type cable and average industrial noise levels. Cable lengths of 10 to 100 feet (3 to 30 meters) have been used successfully.

TABLE 1 Troubleshooting Guide for Noise

Observation	Subject	Possible solution	Notes
Noise a function of cable location	Capacitive coupling	Use shielded or twisted pair.	*a*
	Inductive coupling	Reduce loop area; use twisted pair or metal shield.	*b*
Average value of noise:			
Not zero	Conductive paths or ground loops	Faulty cable or other leakage. Eliminate multiple ground connections.	*c*
Zero	Capacitive coupling	Use shielded or twisted pair.	
Shield inserted:			
Ground significant	Capacitive coupling	Use shielded or twisted pair.	*a*
Ground insignificant	Inductive coupling	Reduce loop area; use twisted pair or metal shield.	*b*
Increasing load:			
Reduces error	Capacitive coupling	Use shielded or twisted pair.	*a*
Increases error	Inductive coupling	Reduce loop area; use twisted pair or metal shield.	*b*
Dominant feature:			
Low frequency	60-Hz ac line, motor, etc.	1. Use shielded or twisted pair. 2. Reduce loop area; use twisted pair or metal shield. 3. Faulty cable or other leakage; eliminate multiple connections.	
High frequency		Complete shield.	*d*
Noise a function of cable movement	Triboelectric effect	Rigid or lubricated cable.	
Noise is white or 1/*f*	Electronic amplifier, etc.	Not a cable problem.	

a. Complete shield to noise-return point and check for floating shields.
b. Nonferrous shields are good only at high frequencies. Use MuMetal shields at low frequencies.
c. Could be capacitive coupling with parasitic rectification, such as nonlinear effects.
d. Look for circuit element whose size is on the order of the noise wavelength (antennas). Openings or cracks in chassis or shields with a dimension bigger than the noise wavelength/20 should be eliminated.
Source: Intelligent Instrumentation, Inc.

Ground-plane cable reduces signal reflections, ringing, and RFI (radio frequency interference). Special termination networks may be required to maintain signal integrity and minimize RFI. If squaring circuits (e.g., Schmitt triggers) are used to restore the attenuated high-frequency signals, improved performance can be realized.

This information is given only as a typical example of what might be encountered. The actual length allowed in a particular application could be quite different.

The following relationships are offered as an aid to visualizing the influence of the most significant factors determining cable length. These relationships show how the various parameters affect cable length. These relationships are *not equations*, and will not allow the calculation of cable length.

For Current Source Signals:
Allowable length is proportional to

$$\frac{I_s D_n C_f}{f_n A N_i}$$

For Voltage Source Signals:
Allowable length is proportional to

$$\frac{V_s D_n C_f}{f_n A N_i R_L}$$

where I_s, V_s = signal level

C_f = coupling factor, which is inversely proportional to effectiveness of any shielding or twisting of wires

D_n = distance to noise source

f_n = noise frequency

A = required accuracy

N_i = noise source intensity

R_L = equivalent resistance to ground at signal input

INDUSTRIAL CONTROL NETWORKS

Commencing in the late 1960s and continuing through 1999 to the 2000s, industrial data and control networks will capture the ingenuity of control engineers, computer scientists, and, of course, communications specialists.[1] The well-justified thirst for information from top management down the industrial hierarchy will continue undiminished. To date, communication system designs have survived past difficult periods in an effort to find the best network systems for given applications, always with an eye toward finding the "universal" answer. Open systems hold some promise along these lines. Millions of hours of effort have gone into defining optimal protocols, improved communication media, practical and cost-effective bus configurations, and the reconfiguration of controls, such as programmable logic controllers (PLCs), distributed control systems, and personal computers (PCs), in an effort to make them increasingly "network friendly." There are other objectives and there have been tough decisions and roadblocks, but the leading technical societies have mustered strength through special committees in defining terms and establishing standards. This work will continue apace.

EARLY NETWORKING CONCEPTS

Early communication needs were served with point-to-point data links (Fig. 1). Very early standards, such as TTY (teletypewriter) current loops and RS-232, which allow different equipment to interface with one another, appeared and were accepted. From that, the star topology (Fig. 2) was developed so that multiple computers could communicate. The central, or master, node uses a communications port with multiple drops, as shown in Fig. 3. In this system the master is required to handle traffic from all the nodes attached, poll the other nodes for status, and, if necessary, accept data from one node to be routed to another. The heavy software burden on the master is also shared to a lesser degree among all the attached nodes. In addition, star topologies are inflexible as to the number of nodes that can

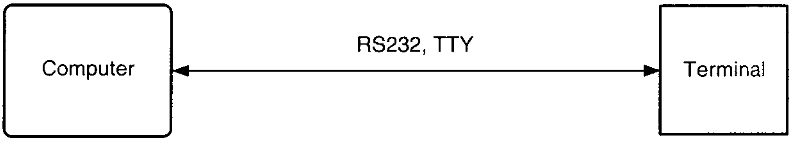


FIGURE 1 Point-to-point communication.

---

[1] Industrial control networks also are discussed in other portions of this handbook. See, in particular, the articles, "Distributed Control Systems," "Programmable Controllers," and "Distributed Numerical Control and Networking" in Section 3.

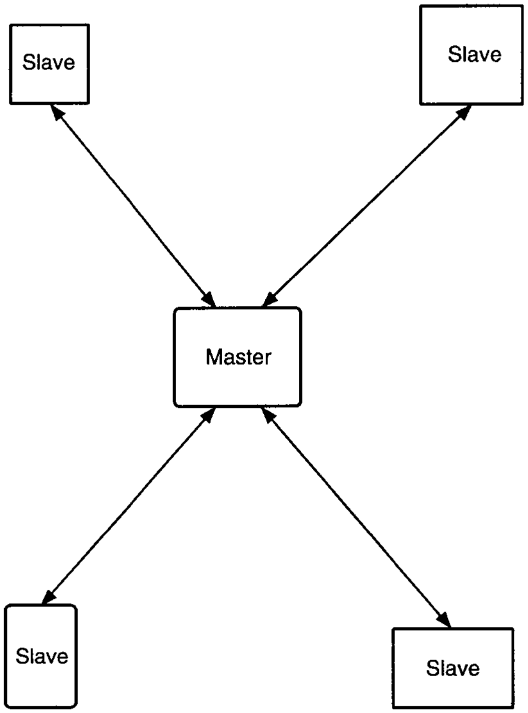


FIGURE 2 Star topology.

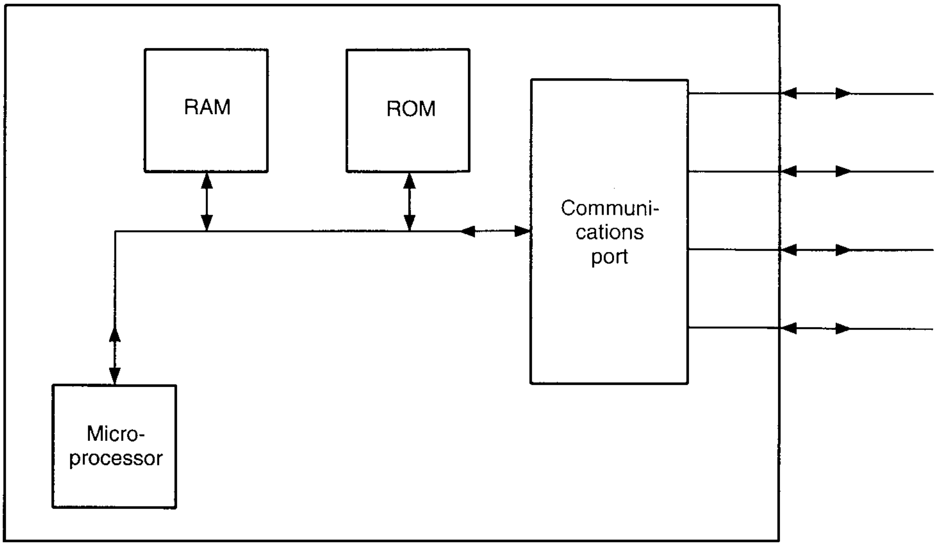


FIGURE 3 Master node for star topology.

be attached. Either one pays for unused connections (for future expansion), or a system results that cannot grow with demands.

To overcome some of these shortfalls, multidrop protocols were established and standardized. Data loop, such as SDIC (synchronous data link control), were developed as well as other topologies, including buses and rings (Fig. 4). The topology of these standards makes it easy to add (or subtract) nodes on the network. The wiring is also easier because a single wire is routed to all nodes. In the case of the ring and loop, the wire also is returned to the master. Inasmuch as wiring and maintenance are major costs of data communications, these topologies virtually replaced star networks. These systems, however, have a common weakness—one node is the master, with the task of determining which station may transmit at any given time. As the number of nodes increases, throughput becomes a problem because (1) a great deal of "overhead" activity may be required to determine which may transmit and (2) entire messages may have to be repeated because some protocols allow only master-slave communications, that is, a slave-to-slave message must be sent first to the master and then repeated by the master to the intended slave receiver. Reliability is another problem. If the master dies, communications come to a halt.

The need for multinode networks without these kinds of problems and restraints led to the development of the initial local area networks (LANs) using peer-to-peer communications. Here no one node is in charge; all nodes have an equal opportunity to transmit. An early LAN concept is shown schematically in Fig. 5.

In designing LAN architecture, due consideration had to be given to the harsh environment of some manufacturing and processing areas. Design objectives included the following.

Noise. Inasmuch as a LAN will have long cables running throughout the manufacturing space, the amount of noise pickup can be large. Thus the LAN must be capable of working satisfactorily in an electrically noisy area. The physical interface must be defined to provide a significant degree of noise rejection, and the protocol must be robust to allow easy recovery from data errors. (See preceding article in this handbook section, "Noise and Wiring in Data Signal Handling.")

Response. The LAN in an industrial situation should have an assured maximum response time, that is, the network must be able to transmit an urgent message within a specified time frame. The real-time aspect of industrial control demands this.

Priority Message. On the factory floor, both control and status, when carried over the same network, should recognize the higher priority of the control message.

Early Data Highways

In 1972 the very first serial data communications highways were introduced. At that time the only purpose of the data highway was to allow host computers to adjust set points or, in some cases, perform direct digital control (DDC), while providing measurement data to the host computer. With such radically altered control concepts, in designing a data highway, great emphasis was placed on proposed data highways and their ability to operate at sufficiently high rates. There was concern that, during process upsets, many alarm conditions could suddenly change and these had to be reported to the entire system quickly so that remedial action could be taken. There also was major concern over start-up and shutdown procedures that cause heavy communication loads. Security also was a major concern.

In 1973 the distributed control system (DCS) appeared. It represented a major departure in control system architecture and impacted on the configuration of the data highway.

Ethernet. The first LAN, developed by Xerox Corporation, has enjoyed years of application. The network uses CSMA/CD (carrier sense multiple access with collision) and is a baseband system with a bus architecture (Fig. 4*b*). Baseband is a term used to describe a system where the information being sent over the wire is not modulated.

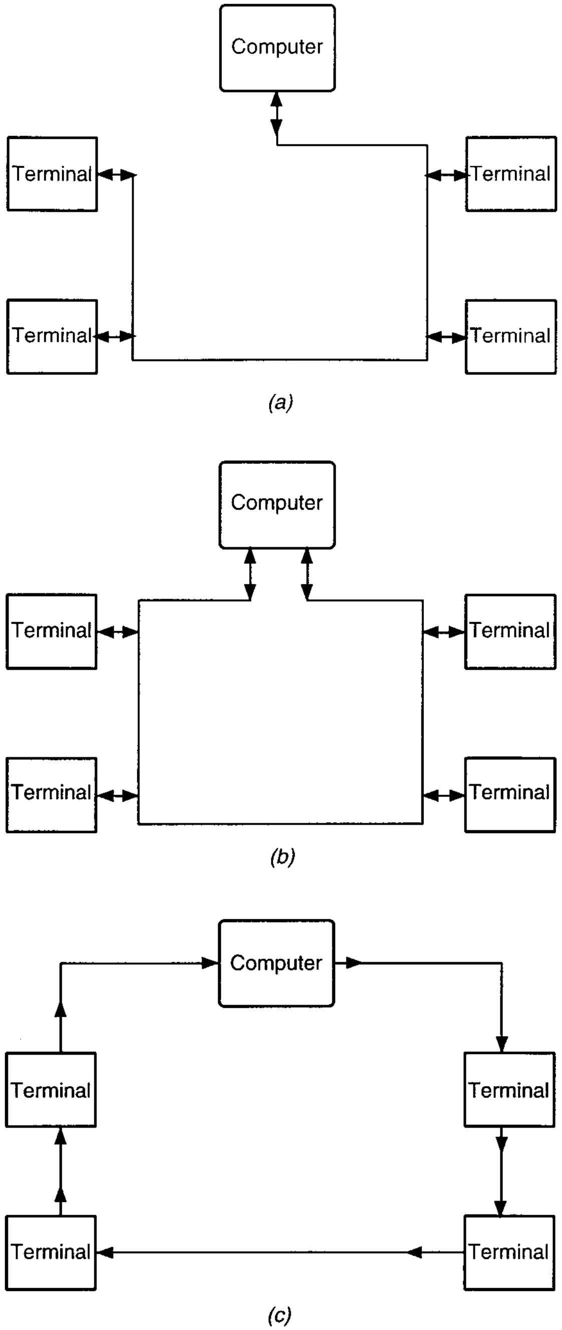


FIGURE 4 Basic communication standards. (*a*) Data-loop topology. (*b*) Bus topology. (*c*) Ring topology.

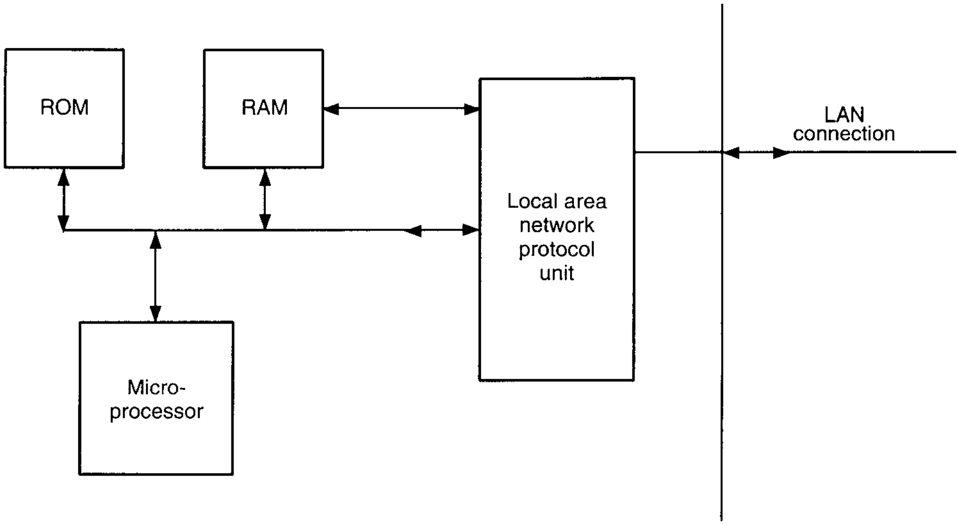


FIGURE 5 Early LAN concept shown schematically.

DECnet. DEC (Digital Equipment Corporation) computers, operating on the factory floor in the early 1970s, were linked by DECnet. This was a token passing technique, described later under "Network Protocols."

DECnet/Ethernet. In 1980, with an aim to support high-speed local LANs, DECnet and Ethernet were used together to form DECnet/Ethernet, a network that has been used widely over the years. One of the major advantages of combining the two concepts is that Ethernet's delay in one node's response to another's request is much shorter than that of a token passing protocol. Users generally found that these networks provide good real-time performance. Ethernet is inherently appropriate for transmitting short, frequent messages and it effectively handles the irregular data transfers typical of interactive terminal communications (Figs. 6 and 7).

CATV Cable. In 1979 a CATV (community antenna television) cable system was announced. This system also had a central point of control and a multimaster protocol, but it used CATV cable at 1 Mbit/s. At these data rates, even in baseband, the transceiver design was based on radio-frequency (RF) technology. The topology of the network used a local star cluster with the CATV interconnecting all clusters. The data communication within the cluster was bit serial, byte parallel.

Later networks offered dual redundant mechanisms so that if one data highway failed, a second data highway would take over. The second highway was unused except for integrity diagnostics. To make certain that these data highways were in good order, elaborate mechanisms were implemented apart from the data communications to ensure cable and station integrity. CATV is mentioned later under MAP protocol.

NETWORK PROTOCOLS

A data communication protocol may be defined as "a set of conventions governing the format and relative timing of message exchange between two (or more) communications terminals," or, restated, "the means used to control the orderly communication of information between stations on a data link."

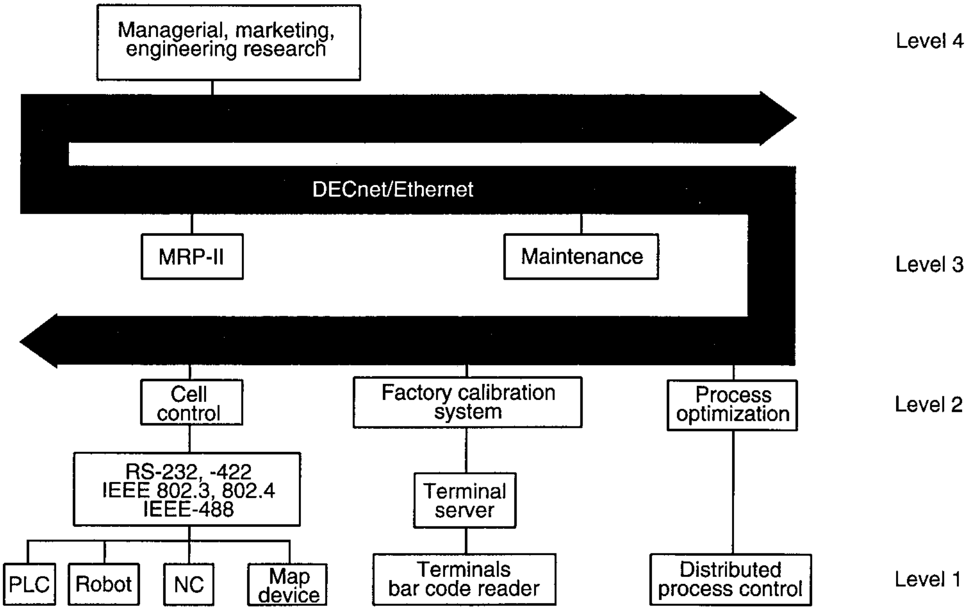


FIGURE 6 LAN for manufacturing applications combines features of DECnet and Ethernet. It may be defined as a multilevel functional model in which distributed computers can communicate over a DECnet/Ethernet backbone. Commencing in the 1980s, large numbers of these networks have been installed.

The communication protocol is vital to equipment design and must be defined accurately if all elements in a system are to work together in harmony in the absence of a lot of local "fixes." Thus much international effort has been made over several years by various technical society committees to accurately identify, define, and refine protocols. Because of the heavy investment involved, a protocol usually is debated for at least a few years before the various committees officially approve standards.

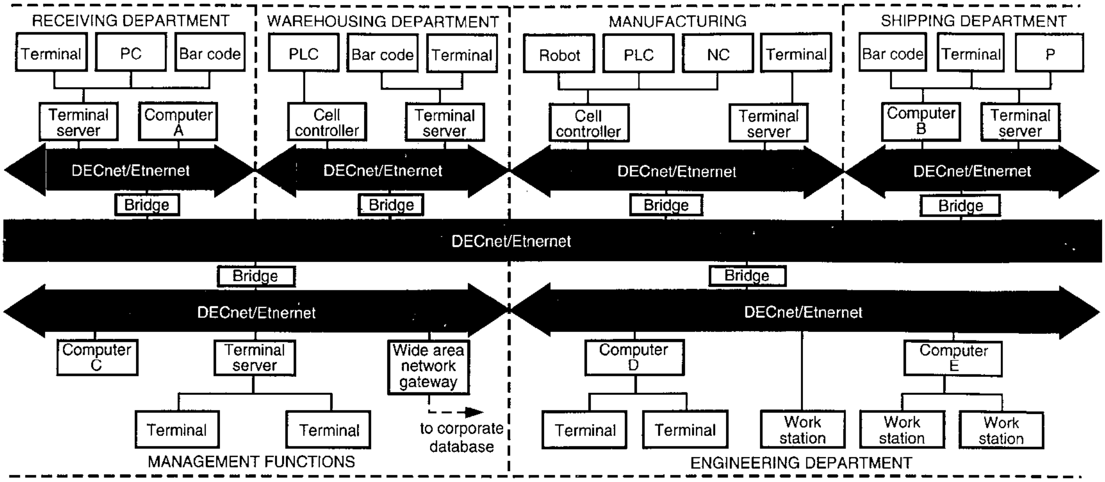


FIGURE 7 Schematic diagram shows how a DECnet/Ethernet baseband cable backbone can extend a manufacturing network to numerous plant areas, including a corporate database.

CSMA/CD Protocol (IEEE 802.3)

"Carrier sense multiple access with collision" is a baseband system with a bus architecture. Normally only one station transmits at any one time. All other stations hear and record the message. The receiving stations then compare the destination address of the message with their address. The one station with a "match" will pass the message to its upper layers, while the others will throw it away. Obviously, if the message is affected by noise (detected by the frame check sequence), all stations will throw the message away.

The CSMA/CD protocol requires that a station listen before it can transmit data. If the station hears another station already transmitting (carrier sense), the station wanting to transmit must wait. When a station does not hear anyone else transmitting (no carrier sense), it can start transmitting. Since more than one station can be waiting, it is possible for multiple stations to start transmitting at the same time. This causes the messages from both stations to become garbled (called a collision). A collision is not a freak accident, but a normal way of operation for networks using CSMA/CD. The chances of collision are increased by the fact that signals take a finite time to travel from one end of the cable to the other. If a station on one end of the cable starts transmitting, a station on the other end will "think" that no other station is transmitting during this travel time interval and that transmission can be resumed. After a station has started transmitting, it must detect when another station is also transmitting. If this happens (collision detection), the station must stop transmitting. Before quitting transmitting, however, the station must make sure that every other station is aware that the last frame is in error and must be ignored. To do this, the station sends out a "jam," which is simply an invalid signal. This jam guarantees that the other colliding station also detects the collision and quits transmitting. Each station that was transmitting must then wait before trying again. To make sure that the two (or more) stations that just collided do not collide again, each station picks a random time to wait. The first station to time out will look for silence on the cable and retransmit its message again.

Token Bus Protocol (IEEE 802.4)

This standard was developed with the joint cooperation of many firms on the IEEE 802 Committee. Since becoming a standard, it was selected by General Motors (GM) for use in its manufacturing automation protocol (MAP) as a local area network to interconnect GM factories.

This is also a bus topology, but differs in two major ways from the CSMA/CD protocol: (1) The right to talk is controlled by passing a "token," and (2) data on the bus are always carrier-modulated.

In the token bus system, one station is said to have an imaginary token. This station is the only one on the network that is allowed to transmit data. When this station has no more data to transmit (or it has held the token beyond the specific maximum limit), it "passes" the token to another station. This token pass is accomplished by sending a special message to the next station. After this second station has used the token, it passes it to the next station. After all the other stations have used the token, the original station receives the token again.

A station (for example, *A*) will normally receive the token from one station (*B*) and pass the token to the third station (*C*). The token ends up being passed around in a logical token ring (*A* to *C* to *B* to *A* to *C* to *B* . . .). The exception to this is when a station wakes up or dies. For example, if a fourth station, *D*, gets in the logical token ring between stations *A* and *C*, *A* would then pass the token to *D* so that the token would go *A* to *D* to *C* to *B* to *A* to *D*. . . . Only the station with the token can transmit, so that every station gets a turn to talk without interfering with anyone else. The protocol also has provisions that allow stations to enter and leave the logical token ring.

The second difference between token bus and CSMA/CD, previously mentioned, is that with the token bus, data are always modulated before being sent out. The data are not sent out as a level, but as a frequency. There are three different modulation schemes allowed. Two are single-channel and one is broadband. Single-channel modulation permits only the token bus data on the cable. The broadband method is similar to CATV and allows many different signals to exist on the same cable, including video and voice, in addition to the token bus data. The single-channel techniques are simpler, less costly, and easier to implement and install than broadband. Broadband is a higher-performance network, permitting much longer distances and, very important, satisfying the present and future communications needs by allowing as many channels as needed (within the bandwidth of the cable).

This protocol still is evolving. Considerations have been made to incorporate the standard of an earlier standards group, known as the PROWAY (process data highway), which was discussed in the mid-1970s by several groups worldwide, including the International Electrotechnical Committee (IEC), the International Purdue Workshop on Industrial Computer Systems (IPWICS), and the Instrument Society of America (ISA). It was also at about this time that the Institute of Electrical and Electronics Engineers (IEEE) became very active with its 802 Committee. PROWAY also is a token bus concept.

Technical committee standards efforts progress continuously, but often quite slowly, so that sometimes users with immediate requirements must proceed with approaches that have not been "officially" standardized. Association and society committees meet at regular intervals (monthly, quarterly, and annually). Thus the individual who desires to keep fully up to date must obtain reports and discussions directly from the organizations involved, or attend meetings and consult those periodicals which assiduously report on such matters.

Benefits to the token bus protocol gained from PROWAY are several.[2]

Token Ring Protocol (IEEE 802.5)

Originally token ring and token bus used the same protocol with different topologies. IBM suggested a different token ring protocol to IEEE. The proposal was accepted and became the basis for the token ring protocol, thereby forming two token protocols.

The topology is that of a ring (Fig. 4*c*). Any one node receives data only from the "upstream" node and sends data only to the "downstream" node. All communication is done on a baseband point-to-point basis. The "right to talk" for this network also is an imaginary token. Token ring has simplicity in that the station with the token simply sends it to the next downstream station. This station either uses the token or lets it go on to the next station.

A general summary of the three aforementioned protocols is given in Table 1.

COMMUNICATION MODELS AND LAYERS

Before describing more recent network protocols, such as MAP and MMS, it may be in order to mention tools and models that have been developed to assist in defining the various "layers" of data communications required from the factory floor or process to top-level corporate management. Computer-integrated manufacturing (CIM), for example, requires excellent communication at all levels before the promises of the concept can be achieved.

---

[2] PROWAY advantages include:

1. The immediate acknowledgment of a frame. If station *A* has the token and uses it to send a message to station *B*, station *B* can be requested to send an acknowledge message back to station *A*. Station *B* does not wait until it gets the token, but instead, station *B* "uses" station *A*'s token for this one message. Station *B*, in essence, is telling station *A* that it received the message without any errors. This idea is allowed in 802.4.

2. Capability to initialize and control a station by sending a message over the network. (Every token bus protocol handler would have an input to control the rest of the node.)

3. Every station has some predefined data that are available upon demand. Any other station can request the data, and this station would send the information out immediately. (The station sending the response is using the other station's token.)

4. Acess time of the network is deterministic and on upper bound must be predictable for any given system. This means that if a station on a network wants to send a message, one should be able to predict the maximum possible delay before it is sent out. (The original 802.4 specified a maximum access time per node, but no upper bound on the number of nodes.)

5. It provides a method of monitoring membership in the logical token ring. If a station died, every other station would be aware of it. Every station in the logical ring would have a list of all stations that are in the ring.

6. It provides for the accumulation of network performance statistics.

TABLE 1 General Characteristics of Basic Protocols

CSMA/CD protocol	Designed for a lot of short messages. Works well if there are not a lot of collisions. With heavy traffic there is a lot of overhead because of the increased number of collisions. Probabilistic in nature—every station has a finite chance of hitting some other station every time it tries to send. Thus there is no guarantee that a message may not be held up forever. This could be catastrophic. Baseband configuration means that digital data are represented as discrete levels on the cable, thus reducing noise effects and allowing longer cable lengths. Probably the most cost-effective for most applications.
Token bus protocol	Robust, able to recover from errors easily. Cable length is limited only by attenuation of signal in cable. Allows three different modulation schemes. Can carry multiple voice and video channels at the same time as data.
Token ring protocol	Deterministic access and priorities on messages. Can use fiber-optic cables. Maximum physical length of cable can be large. Requires more complex wiring than a bus because the last station must be connected to the first station to form a ring. Redundant cabling may be needed for fault tolerance because a single break or down station can stop all data transfer.

Experts have placed communication systems in levels ranging from three to seven in number. Examples of a three-level and a seven-level model are given.

Three-Level Concept. This represented the state of the art prior to the development of the OSI reference model in the early 1970s.

Lowest Level. Links groups of people and machines. The link involves the flow of information among different workstations at a department level. These local networks must have a fast payback because it is in this area where most manufacturing and processing alterations are made.

Middle Level. Facilitywide networks that permit all departments within a facility to share data. Examples may include (1) obtaining employee vacation data at a moment's notice so that shift assignments can be made without delay, (2) tracing the history of a quality control problem immediately when a failure is noted, and (3) permitting a service supervisor to check the current readiness of production equipment.

Highest Level. Corporatewide communications where all multifactories and departments can exchange information and report to a central information processing site. Manufacturing automation, such as CAD/CAM, materials management, and automated machine tools, can be linked to management information processing, including such functions as financial data, sales reporting, statistical quality control (SQC), statistical process control (SPC), and corporate planning, among many other functions. This arrangement enables managers to check the flow of production and product lead times as only one of many examples that could be given. This level of networking is more complex than the local area networks (LANs) or the facilitywide networks (WANs), in part because of the heavy information-exchange load.

OSI Reference Model. The OSI (open system interconnections) model was developed in the 1970s as a joint effort of several groups, including the International Standards Organization (ISO), the American National Standards Institute (ANSI), the Computer and Business Equipment Manufacturers Association (CBEMA), and the National Institute of Standards and Technology (NIST), formerly the National Bureau of Standards (NBS). It has been estimated that this work represents tens of thousands of hours of effort by experts worldwide.

Initially the OSI reference model was oriented to telephony systems. Some of its achievements have included MAP (manufacturing automation protocol) and MMS (manufacturing message service). The seven distinct layers of OSI are shown in Table 2.

TABLE 2 Open-System Interconnections (OSI) Model

Layer	Name	Uses and applications
1	Physical	Electrical, mechanical, and packaging specifications of circuits. Functional control of data circuits.
2	Link	Transmission of data in local network-message framing, maintain and release data links, error and flow control.
3	Network	Routing, switching, sequencing, blocking, error recovery, flow control. System addressing and wide-area routing and relaying.
4	Transport	Transparent data transfer, end-to-end control, multiplexing, mapping. Provides functions for actual movement of data among network elements.
5	Session*	Communications and transaction management. Dialog coordination and synchronization. Administration and control of sessions between two entities.
6	Presentation†	Transformation of various types of information, such as file transfers; data interpretation, format, and code transformation.
7	Application‡	Common application service elements (CASE); manufacturing message service (MMS); file transfer and management (FTAM); network management; directory service.

[*] The session layer provides functions and services that may be used to establish and maintain connections among elements of the session, to maintain a dialog of requests and responses between the elements of a session, and to terminate the session.

[†] The presentation layer provides the functions, procedures, services, and protocol selected by the application layer. Functions may include data definition and control of data entry, data exchange, and data display. This layer comprises CASE (common application services), SAS (specific application services), and management protocols required to coordinate the management of OSI networks in conjunction with management capabilities that are embedded within each of the OSI layer protocols.

[‡] The application layer is directly accessible to, visible to, and usually explicitly defined by users. This layer provides all of the functions and services needed to execute user programs, processes, and data exchanges. For the most part, the user interacts with the application layer, which comprises the languages, tools (such as program development aids, file managers, and personal productivity tools), database management systems, and concurrent multiuser applications. These functions rely on the lower layers to perform the details of communications and network management. Traditionally, network vendors have provided a proprietary operating system for handling functions in the upper layers of the OSI model. These unique features have been the source of interconnection difficulties.

Manufacturing Automation Protocol

In the early 1970s the management of several industrial firms in the discrete-piece manufacturing industries realized that industrial computers and the networks that serve them were the key tools for achieving production automation on a grand scale as contrasted with the low-key efforts of the past, such as "isolated," or islands of, robotics and computerized numerical control of machines. Particularly in the United States the need to automate and pursue the ambitious goals of new production concepts, such as CIM, MRP I and II (materials requirement planning), FMSs (flexible manufacturing systems), and just in time (delivery), was precipitated by severely threatening competition from abroad. An excellent example of this was the decision by the management of GM to develop a program that would hasten the achievement of CIM, recognizing the pitfalls that existed then in attempting to utilize the control products of numerous manufacturers, that is, products which could not easily be connected and orchestrated in a practical operating network. Not necessarily the first, but certainly the most illuminated effort was the recommendation and demand of GM for simplification and implementation of control and communication components. Thus the MAP program was initiated. The U.S. government, also at about this time, created a special section of NBS (later NIST) to assist industry in the development of improved automation techniques, including networking. Again, the motive was a serious concern over the country's diminishing leadership in manufacturing. These actions placed great emphasis on LANs and the general concept of interconnectability.

GM, prior to the formation of the MAP plan, had for a number of years used a network of CATV cable (described previously) for closed-circuit television. To this had been added several channels of low-speed serial data communications, which involved RF modems to operate in the television broadband spectrum. To get a plantwide high-speed information network standard established for GM, the firm formed an internal committee called MAP. This well-publicized committee invited proposals from many vendors and circulated several papers on GM requirements.

The primary purpose of the plantwide (“backbone”) network was not process control, but rather to allow the two-way flow of high-level production data. Otherwise the whole bandwidth of the backbone network would easily be consumed by local traffic. GM had defined a true hierarchical network environment and had clearly endorsed token bus and CATV, but not for process control. Also, during this period, a lot of interest was shown in TOP (technical office protocol).

The MAP concept developed at a good rate for several years, progressing to MAP 3.0. Scores of MAP networks have been installed in the United States and abroad, with the largest number located in large discrete-manufacturing firms, including, of course, GM, but also Ford Motor and Boeing. In June 1987 a 6-year freeze was imposed on the MAP 3.0 specification. This intended to allow manufacturers to build, and users to install, MAP networks without concern that the specifications continue to change over relatively short intervals of time. A leader in the MAP field indicated that the freeze would not affect the addition of functionality, compatible backward and forward, but that technology would not be added that would make obsolete what has been installed already.

Acceptance of MAP probably peaked in 1990. A large base, including Ethernet, DECnet, ARCNET, and others, remains in place. Thus, unfortunately, evaluations and guides to system selection are well beyond the province of a “permanent” handbook reference. However, a few general suggestions may be in order:

1. Are there severe noise conditions?
   - YES—Seriously consider fiber-optic cables.
   - NO—Hard-wired.
2. Are there time-critical throughputs—monitoring, data collection, control?
   - YES—Provide equal access by all nodes.
   - NO—If transmission distance is under 1 mile (1.9 km), consider twisted pair. If over this distance, consider telephone.
3. Is equal access by all nodes a requirement?
   - YES—Use single high speed.
   - NO—Consider combination of high-speed access and twisted pair.
4. Are there plans for future expansion?
   - YES—Use standards-based network.
   - NO—Use proprietary or standards-based network.

The general characteristics of PLC-based LANs are listed in Table 3 for various network types.

Open Systems

Although not necessarily fully accredited to the developers of the OSI model described previously, that group placed early emphasis on the concept of open-system architecture. The footnotes included with Table 2, which describes the OSI model, aptly define the presentation and application layers in open-system architecture. Also, the importance of MMS is listed under the applications layer of the OSI model. MMS is accepted internationally as a standard communications protocol for integrating mixtures of unlike devices that are used on the factory floor or by processing areas. The MMS reduces development costs for adding new devices to a network and diminishes the requirement for custom software and hardware for diverse device interfacing. Within the last few years it is estimated that some 40 major suppliers have recognized the MMS protocol.[3]

As of 1993 the topic of open-system architecture remains quite fluid. Some major firms are refining their most recent offerings of “open” networks that use the term as part of a proprietary trade name.

---

[3] It is interesting to note that 10 of these firms participated in an unusual demonstration at the ISA 1991 Exhibit in Anaheim, California, for the purpose of showing how MMS can handle data transfer between PLCs, PCs, NC, robots, and others on the plant floor.

TABLE 3 General Characteristics of PLC-Based LANs*

Very Good	Good	Fair	Poor
Noisy environments			
RHF FO	HC	TP	RTRL
Speed			
HC FO RHF		TP	RTRL
Throughput			
	HC FO RHF	TP	RTRL
Purchase price			
	TP RTRL	HC FO	RHF
Lifetime cost			
	RHF	FO	HC TP RTRL
Expandability			
HC FO RTRL RHF		TP	

*FO—fiber-optic; HC—hard-wired coaxial; RHF = redundant hard-wired/fiber; RTRL—remote telephone/radio link; TP = twisted pair.

Fieldbus

Since electronic measurement and control systems essentially replaced pneumatic systems several decades ago, industry has depended heavily on the 4- to 20-mA transmission standard, that is, until the *near* future! The ever-increasing use of microprocessor technology in sensors, transmitters, and control devices has created the need for a digital replacement of the 4- to 20-mA standard. Notably, this applies to “smart” transmitters.

A new fieldbus standard has been in preparation since 1985, sponsored essentially by the same society and institutional groups that have done admirable work in preparing other standards, network models, and so on. As of early 1992, completed portions of the new (SP 50) standard for both process control and factory automation include the physical layer and function block requirements. Field installations are under way in the United States and internationally. See Section 10 for an overview of the ISA SP50 fieldbus standard.

One portion of the standard (H1) specifies a low-speed powered link as a digital replacement for 4- to 20-mA transmission. When implemented, microprocessors embedded in smart transmitters will be able to communicate directly with digital control systems. Another portion (H2) specifies a high-speed unpowered link to operate at 1 Mbaud.

Other tests completed to date have included what have been described as the “worst of conditions,” such as inserting a bad message, a missing terminator, and an open spur. Still other tests included adding and removing a device on line, adding crosstalk, adding a wide frequency range of white noise, and placing a walkie-talkie within 2 feet (30 cm) from the open cable for RF interference tests.

It can be safely forecast that the proposed field bus will be the subject of innumerable papers and discussions over the next few years.

FIBER-OPTIC CABLES AND NETWORKS

Fiber-optic technology has been used for well over a decade in telephony. The first large-scale demonstration was made by AT&T in 1980 in connection with the Olympic Games held at Lake Placid, New York. This test installation was only 4 km (2.5 miles) long, but tested out very successfully. The first actual commercial installations were made between Washington and New York and New York and Boston. As early as 1982 Leeds & Northrup offered an earlier version of the network, as shown in Fig. 8. This is a redundant electrical-optical highway which has been used in hundreds of installations worldwide.

Fiber-optics offers many advantages and relatively few limitations as a networking medium. Some of the advantages include the following:

- Not affected by electromagnetic radiation (RMI). For example, the cable can be installed in existing high-voltage wireways, near RF sources, and near large motors and generators. No shielding is required.
- With experience, fiber cable is easy to install and at less labor cost. Only a few special tools are needed.
- Not affected by lightning surges.
- Resists corrosion.
- Intrinsically safe.
- Compatibility makes fiber cables easy to integrate with existing platforms. Fiber cable is inherently suited to open systems.
- Increased security because of immunity to "bugging."

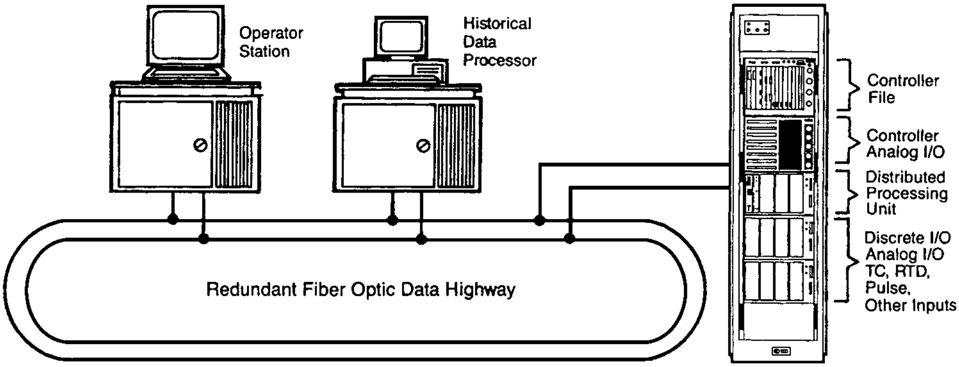


FIGURE 8 Redundant fiber-optic data highway for advanced data acquisition and plantwide control. This is an updated version of fiber-optic system first offered to industry in 1982, with ensuing hundreds of installations. Data highway consists of two redundant fiber-optic loops with repeaters to provide digital data communications among large numbers of multiloop controllers, operator stations, and computers. The optical data highway loops from one cluster to another, eventually returning to a control room with operator stations. One loop transmits digital data in a clockwise direction, while the other transmits counterclockwise. This ensures that communication between any two stations will be maintained no matter where a fault occurs. (*Leeds & Northrup*.)

- Higher data rates. The next generation of fiber-optic network protocol (FDDI) can transfer data at a 100-Mbit/s rate on the same cable fiber that now offers a 10-Mbit/s rate.

For reasons of caution and lack of better understanding of fiber technology, coupled with a continuing (but decreasing) cost differential with other media, optic cables still are in the lower portion of their growth curve. Predictions for increased use are very optimistic.

Characteristics of Optical Fibers and Cables

As in electrical transmission systems, the transmission sequence of a light-wave system begins with an electric signal. This signal is converted to a light signal by a light source, such as a light-emitting diode (LED) or a laser. The source couples the light into a glass fiber for transmission. Periodically, along the fiber, the light signal may be renewed or regenerated by a light-wave repeater unit. At its destination, the light is sensed by a special receiver and converted back to an electric signal. Then it is processed like any signal that has been transmitted in electrical form.

The system comprises transmitter circuitry that modulates or pulses in code the light from a light source. An optical fiber waveguide conducts the light signal over the prescribed distance, selected because of its particularly good transmission capability at the wavelength of the light source. The terminal end of the waveguide is attached to a detector, which may be a *pn* junction semiconductor diode or an avalanche photodiode, to accept the light and change the signal into an electromagnetic form for the receiver circuitry. The latter decodes the signal, making it available as useful electronic analog or digital output. When two-way communication is needed, the system is fully duplexed and two circuit links are needed.

Optical Fibers. Glasses of many compositions can be used for optical fibers, but for intermediate- and low-loss applications the options become increasingly limited. Multicomponent glasses containing a number of oxides are adequately suited for all but very low-loss fibers, which are usually made from pure fused silica doped with other minor constituents. Multicomponent glasses are prepared by fairly standard optical melting procedures, with special attention given to details for increasing transmission and controlling defects resulting from later fiber drawing steps. In contrast, doped fused silica glasses are produced by very special techniques that place them almost directly in a form from which fibers may be drawn.

Digital Light-Wave Systems. Much research has been directed toward light-wave systems that are digital. In a digital system, the light source emits pulses of light of equal intensity, rather than a continuous beam of varying intensity (analog approach). Each second is divided into millions of slices of time. The light source inserts 1 bit of information into each time slot, which flashes on briefly or remains off. The receiver looks for 1 bit in each slot. If the receiver senses a pulse, it registers a 1; if the absence of a pulse, a 0. Eight such bits of information make up a digital word. From a series of such words, other elements of the transmission system can reconstruct the original signal.

The capacity of a digital light-wave system is the maximum rate at which pulses can be sent and received. The maximum pulse rate is limited by how much the signal is distorted by dispersion as it travels along the fiber. *Dispersion* means that a pulse is spread out in time, so that some of the pulses arrive in the wrong time slot. If enough is lost from the proper slot, the receiver may not sense a pulse that was sent. If enough is received in an adjoining slot, the receiver may sense a pulse when none was sent. The greater the dispersion, the longer the time slots must be for the receiver to sense accurately.

Basic Fiber Types. Dispersion is of two kinds: (1) modal and (2) chromatic. Modal dispersion is the spreading of light as it traverses a length of fiber along different paths or modes (Fig. 9). Each path is a different length, and thus light takes a different time to travel through each. The highest-capacity fiber has only a single mode, so it has no modal dispersion. However, such fibers are much smaller, more difficult to couple light into, and harder to splice and connect with other types of fibers.

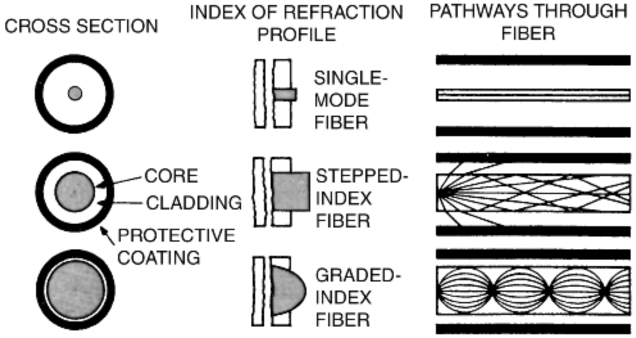


FIGURE 9 Schematic sectional views of fiber-optic cable. Not all layers are shown. Structure of the fiber determines whether and how the light signal is affected by modal dispersion. A single-mode fiber permits light to travel along only one path. Therefore there is no modal dispersion. In contrast, a step-index fiber provides a number of pathways of different lengths, but only one index of refraction boundary between layers, which bends the light back toward the center. Here modal dispersion is high. A graded-index fiber has many layers. The resulting series of graded boundaries bends the various possible light rays along paths of nominally equal delays, thus reducing modal dispersion.

The more common type of fiber is multimode, either step index or graded index. These fibers have wider-diameter cores than single-mode fibers and accept light at a variety of angles. As light enters at these different angles, it travels through the fiber along different paths. A light beam passing through a step-index fiber travels through its central glass core and in the process ricochets off the interface of the cladding adhering to and surrounding the core. The core-cladding interface acts as a cylindrical mirror that turns light back into the core by a process known as total internal reflection. To ensure that total internal reflection occurs, fibers are usually made from two glasses: core glass, which has a relatively higher refractive index, and clad glass, or possibly a plastic layer surrounding the core, which has a somewhat lower refractive index. When the seal interface between core glass and clad glass is essentially free of imperfections and the relative refractive indexes of the glasses used are correct, many millions of internal reflections are possible and light can travel through many kilometers of fiber and emerge from the far end with only a modest loss in brightness or intensity. A step-index fiber has just a single composition inside the cladding. Light must travel to this boundary before it is bent toward the center. The paths in this type of fiber disperse the pulse more than in a graded-index fiber.

In a graded-index fiber, light is guided through it by means of refraction or bending, which refocuses it about the center axis of the fiber core. Here each layer of glass from the center of the fiber to the outside has a *slightly* decreased refractive index compared to that of the layer preceding it. This type of fiber construction causes the light ray to move through it in the form of a sinusoidal curve rather than in the zigzag fashion of the step-index variety. With this type of fiber, when the physical design is correct and the glass flaws are limited, light can also be conducted over very long distances without severe loss because it is trapped inside and guided in an efficient manner. The fiber core is the portion of an optical fiber that conducts the light from one end of the fiber to the other. Fiber core diameters range from 6 to $\sim$250 μm.

Fiber Cladding. To help retain the light being conducted within the core, a layer surrounding the core of an optical fiber is required. Glass is the preferred material for the cladding, although plastic-clad silica fibers are common in less demanding applications. The cladding thickness may vary from 10 to $\sim$150 μm, depending on the particular design.

Index of Refraction. This is the ratio of the velocity of light passing through a transparent material to the velocity of light passing through a vacuum using light at the sodium D line as a reference.

The higher the refractive index of a material, the lower the velocity of the light passing through the material and the more the ray of light is bent on entering it from an air medium.

Numerical Aperture. For an optical fiber this is a measure of the light capture angle and describes the maximum core angle of light rays reflected down the fiber by total reflection. The formula from Snell's law governing the numerical aperture (NA) for a filter is

$$\mathrm{NA} = \sin\theta = \sqrt{n_1^2 - n_2^2}$$

where n_1 is the refractive index of the core and n_2 the refractive index of the clad glass.

Most optical fibers have NAs between 0.15 and 0.4, and these correspond to light acceptance half-angles of about 8° and 23°. Typically, fibers having high NAs exhibit greater light losses and lower bandwidth capabilities.

Light Loss or Attenuation through a Fiber. This is expressed in decibels per kilometer. It is a relative power unit according to the formula

$$\mathrm{dB} = 10 \log \frac{I}{I_0}$$

where I/I_0 is the ratio of the light intensity at the source to that at the extremity of the fiber. A comparison of light transmission with light loss in decibels through 1 km of fiber is as follows:

80% transmission per kilometer $\simeq$ loss of ~1 dB/km

10% transmission per kilometer $\simeq$ loss of ~10 dB/km

1% transmission per kilometer $\simeq$ loss of ~20 dB/km

Bandwidth. This is a rating of the information-carrying capacity of an optical fiber and is given either as pulse dispersion in nanoseconds per kilometer or as bandwidth length in megahertz-kilometers. Light pulses spread or broaden as they pass through a fiber, depending on the material used and its design. These factors limit the rate at which light carrier pulses can be transmitted and decoded without error at the terminal end of the optical fiber. In general, a large bandwidth and low losses favor optical fibers with a small core diameter and a low NA.

The longer the fiber, the more the dispersion. Thus modal dispersion limits the product of the pulse rate and distance. A step-index fiber can transmit a maximum of 20 Mbit of information per second for 1 km, and a graded-index fiber, more than 1000 Mbit. The process for making very low-loss fibers is essentially the same whether the fiber is step or graded index. Consequently nearly all multimode fibers presently used or contemplated for high-quality systems are of the higher-capacity graded-index type. Possibly for very high-capacity installations of the future, single-mode fibers may be attractive.

Cabling and Connections. Although optical fibers are very strong, having a tensile strength in excess of 500,000 psi (3450 MPa), a fiber with a diameter of 0.005 inch (0.1 mm), including the light-guide core and cladding, has a maximum tensile strength of only ~10 psi (0.07 MPa). Unlike metallic conductors, which serve as their own strength members, fiber cables must contain added strength members to withstand the required forces. Also, pulling forces on unprotected fibers may increase their losses, as the result of bending or being under tension. Sometimes this is called microbending loss.

Imaging Requirements. In imaging, both ends of the group of fibers must maintain the exact same orientation one to another so that a coherent image is transmitted from the source to the receiver. Flexible coherent bundles of optical fibers having only their terminal ends secured in coherent arrays are used primarily in endoscopes to examine the inside of cavities with limited access, such as stomach, bowel, and urinary tract.

Rigid fiber bundles fused together tightly along their entire length can be made to form a solid glass block of parallel fibers. Slices from the block with polished surfaces are sometimes used as fiber-optic faceplates to transmit an image from inside a vacuum to the atmosphere. A typical application is the cathode-ray tube (CRT) used for photorecording. The requirement for this type of application is for both image coherence and vacuum integrity, so that when the fiber-optic array is sealed to the tube, the vacuum required for the tube's operation is maintained. However, any image formed electronically by phosphor films on the inside surface of the fiber-optic face is clearly transmitted to the outside surface of the tube's face. High-resolution CRT images can easily be captured on photographic film through fiber-optic faceplates.

Light Sources and Detectors

LEDs produce a relatively broad range of wavelengths, and in the 0.8-μm-wavelength range, this limits present systems to $\sim$140 Mbit/s for a 1-km path. Semiconductor lasers emit light with a much narrower range of wavelengths. Chromatic dispersion is comparatively low, so $\sim$2500 Mbit/s may be transmitted for a 1-km path.

Another factor affecting system capacity is the response time of the sources and detectors. In general it is possible to build sources and detectors with sufficiently short response times that the fiber, rather than the devices, becomes the capacity limiting factor. With single-mode fibers, lasers, and high-speed detectors, transmission rates of more than 10^9 bit/s have been achieved experimentally. This corresponds to more than 15,000 digital voice channels. Although it is interesting to learn how fast a rate can be achieved, in practice the system designer must balance other technical, operational, and economic constraints in deciding how much capacity to require of an individual fiber.

Present semiconductor lasers can couple $\sim$1 mW of optical power into a fiber. On the decibel scale, this is expressed as 0 dBm, meaning 0 dB above a reference power of 1 mW. Although some increase in power is possible, the small size and the temperature sensitivity of these lasers make them inherently low-power devices. LEDs can be made that emit as much power as lasers, but since they project light over a wide angle, much of it is lost just coupling it into the fiber. This loss is typically $\sim$10 to 20 dB. Lasers are more complex and require more control circuitry than LEDs, but they are the light source of choice when repeaters must be far apart and the desired capacity is high.

Light-wave receivers contain photodiodes which convert incoming light to an electric current. The receivers used in telecommunications system are avalanche photodiodes (APDs) made of silicon. They are called avalanche devices because the electric current is amplified inside the diode. This results in a more sensitive receiver than photodiodes without internal amplification. Again, this improved performance is achieved at the expense of added complexity. APDs require high-voltage power supplies, but they are the detectors of choice when high performance is desired.

Even with APDs, light-wave receivers are less sensitive than the best electrical ones; they require a larger minimum received power. This is a consequence of the random fluctuations in optical signal intensity known as shot noise. Light-wave systems can compensate for this. They can carry a much wider bandwidth than electrical systems, and bandwidth can be used to offset noise.