Interface Electronic Circuits

6

"Engineers like to solve problems.
If there are no problems handily available,
they will create their own problems."

—Scott Adams

A system designer is rarely able to connect a sensor directly to processing, monitoring, or recording instruments, unless a sensor has a built-in electronic circuit with an appropriate output format. When a sensor generates an electric signal, that signal often is either too weak, or too noisy, or it contains undesirable components. Besides, the sensor output may be not compatible with the input parameters of a data acquisition system, that is, it may have a wrong output format. To mate a sensor and a processing device, they either must share a "common value" or use in-between some kind of a "mating" device. In other words, signals from a sensor usually have to be conditioned and modified before they are fed into a processing device (a load). Such a load usually requires voltage or current as its analog input signal, or a digital code. Nowadays, it is preferable if the sensor's output is preprocessed and presented at the output in a ready-to-use form. An example is an accelerometer that outputs a digital signal with an encoded number of measured g. Thus, a great majority of sensors that produce analog signals require the interface circuits.

Trend in modern sensor designs focuses on integration of sensing components with signal conditioning, converting, and communication circuits. Such a combination is called a *sensing module*. As an illustration, consider Fig. 6.1 that shows an integrated sensing module having two sensing elements that selectively respond to two input stimuli. For operation, a sensing element may require some supporting parts. For example, an RH (relative humidity) sensor may need a protective grid or

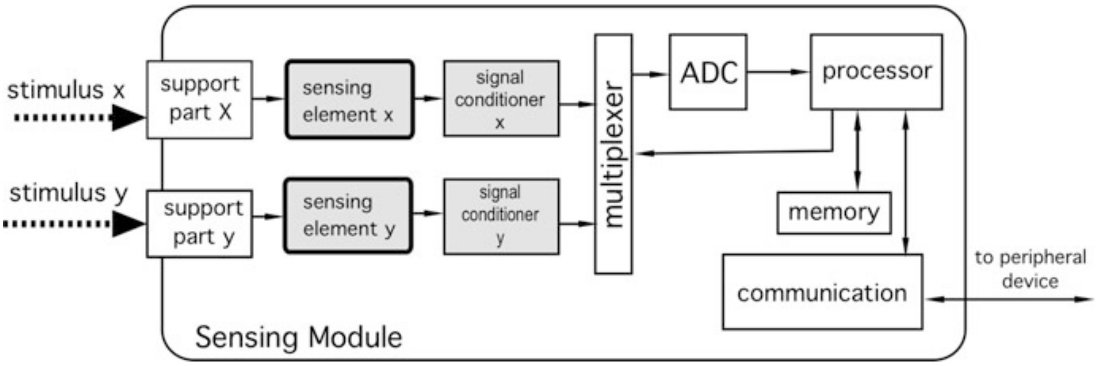


Fig. 6.1 Block diagram of sensing module

even blower for delivering sampled air to the sensing element. Another example is an imaging sensor that requires a focusing lens. Another type of a supporting part is required for an *active sensor*—a pilot signal generator (called excitation generator). For example, a hygristor (moisture sensitive resistor) for its operation requires a.c. current, thus the sensing element shall be appended with an a.c. current generator.

Since a typical sensing element produces low-level analog signals, its output signals need amplification, filtering, impedance matching, and perhaps a level shifting, before it can be digitized. All these functions are performed by signal conditioners. Since a module may comprise more than one channel, outputs of all signal conditioners should be converted into a common digital format. One option is to have one analog-to-digital converter (ADC) per channel, but in most cases it is more convenient and economical to have a single high-quality ADC common for all channels. Thus, the outputs of all signal conditioners shall be connected one at a time to the common ADC. This function is performed by a switching analog gate or *multiplexer* (MUX).

ADC generates a digital code that is fed into a processor for an on-board computation of the input stimuli. For example, if the stimulus is temperature, the processor should compute its value in a selected scale with acceptable resolution—for example, in degree Celsius with a resolution of 0.02 °C. Due to various imperfections, often accuracy cannot be achieved unless the entire system from the input to the processor is individually calibrated. A calibration process requires determination of certain unique parameters that are stored in the sensing module's memory. And finally, the processed information must be communicated in a selected format to the outside peripheral device. This is the function of a communication circuit. An example is an I^2C serial communication link that requires only two wires for transmitting multichannel digital information.

In this chapter we will discuss some electronic interface components. These may be parts of the sensing module, or used separately for supporting the nonintegrated sensors. For details of digital signal processors, memory, and communication devices, the reader is referred to specialized texts.

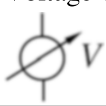
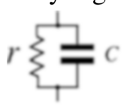
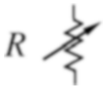
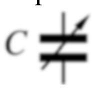
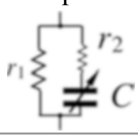
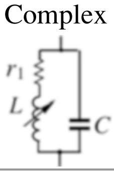
6.1 Signal Conditioners

A signal conditioning circuit has a specific purpose—to bring a signal from the sensing element up to the format that is compatible with the load device—typically an ADC. To do its job effectively, a signal conditioner must be a faithful slave of two masters: the sensing element (sensor) and the load device. The signal-conditioner input characteristics shall be compatible to the output characteristics of the sensor while its output should generate voltage for ease of interfacing with an ADC or other load. This book, however, is about sensors; therefore, below we discuss only typical front stages of the signal conditioning circuits that are coupled to various sensors.

The front end of a signal conditioner depends on type of the sensor's output electrical characteristics. Table 6.1 lists five basic types of the sensor output properties: voltage, current, resistive, capacitive, and inductive. Selecting the appropriate input stage of a signal conditioner is essential for the optimal data collection.

A difference between a voltage-generating and current-generating sensors should be clearly understood. The former has a relatively low-output impedance. Its output voltage is little dependent on the load but the current is function of the load. This sensor resembles a battery whose voltage is driven by a stimulus.

Table 6.1 Sensor types and corresponding inputs of signal conditioners

Sensor type	Sensor impedance	Signal conditioner front stage
Voltage out 	Very low resistive 	High input resistance amplifier ("voltmeter")
Current out 	Very high complex 	High input impedance amplifier or low input resistance circuit ("amperemeter")
Resistive 	Resistive 	Resistance-to-voltage converter ("Ohm-meter")
Capacitive 	Complex 	Capacitance-to-voltage converter ("capacitance meter")
Inductive 	Complex 	Inductance meter

The latter sensor has a large output impedance, typically much larger than the load, and thus produces current that is mainly independent of the load.

6.1.1 Input Characteristics

The input part of a signal conditioner may be specified through several standard parameters. These are useful for calculating how accurately the circuit can process the sensor's output signals and what would be the circuit's contribution to the total error budget?

The input impedance shows by how much the circuit loads the sensor at different frequencies. The impedance may be expressed in a complex form as:

$$\mathbf{Z} = \frac{\mathbf{V}}{\mathbf{I}}, \tag{6.1}$$

where **V** and **I** are the complex notations for the voltage and current across the input impedance. For example, if input of a signal conditioner is modeled as a parallel connection of the input resistance, R and input capacitance, C, (Fig. 6.2a), the complex input impedance may be represented as

$$\mathbf{Z} = \frac{R}{1 + j\omega RC}, \tag{6.2}$$

where ω is the circular frequency and $j = \sqrt{-1}$ is the imaginary unity.

Formula (6.2) suggests that the input impedance is function of the signal frequency. With increase in the signal rate of change, the input impedance becomes lower. On the other hand, at very low frequencies, a circuit, having a relatively low-input capacitance C and resistance R, has an input impedance that is almost equal to the input resistance: $\mathbf{Z} \approx R$. Relatively low, here means that the reactive part of the above equation becomes negligibly small, i.e., the following holds

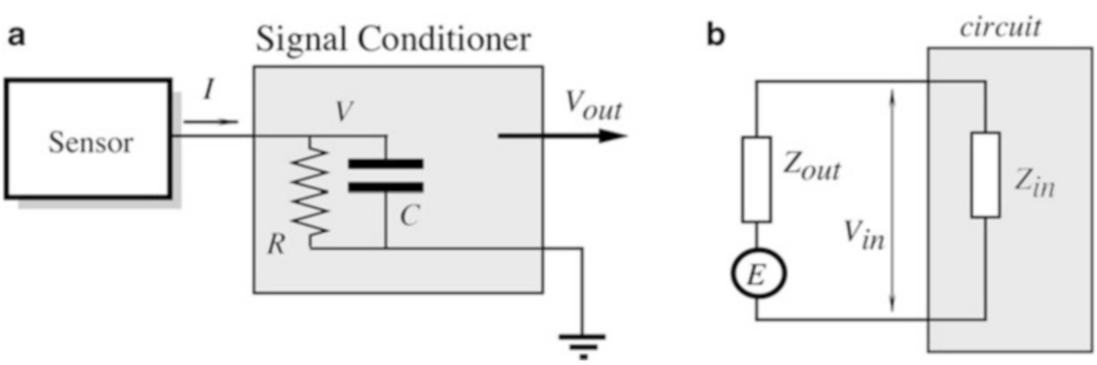


Fig. 6.2 Complex input impedance of interface circuit (**a**), and equivalent circuit of voltage-generating sensor (**b**)

$$RC\frac{1}{\omega}. \tag{6.3}$$

The product RC is called a time constant τ that is measured in seconds. Whenever the input impedance of a circuit is considered, the output impedance of the sensor must be taken into account for a very obvious reason—when the sensor is attached, the impedances are connected in parallel. For example, if the sensor is of a capacitive nature (Table 6.1), to define a frequency response of the input stage, sensor's capacitance with its resistive components r_1 and r_2 must be connected in parallel with the circuit's input capacitance.

Figure 6.2b shows an equivalent circuit for a voltage-generating sensor. The signal-conditioner's front end comprises the sensor output, Z_{out}, and the circuit input, Z_{in}, impedances (we use here scalar notations). Output signal from the sensor is represented by a voltage source, E, which is connected in series with the output impedance Z_{out}. Note that an ideal voltage source E by definition has a zero internal impedance. Its real impedance is modeled by Z_{out}. If the impedance is low, especially its resistive part over the entire frequency range, the sensor and signal-conditioner's reactive components (capacitive and/or inductive) can be safely ignored.

However, if the sensor's output impedance has a sizable value, it cannot be ignored. Let us analyze a voltage-generating sensor (see Table 6.1). By accounting for both impedances (the sensor's output and signal-conditioner's input), the signal-conditioner input voltage, V_{in} is represented by

$$V_{in} = E\frac{Z_{in}}{Z_{in} + Z_{out}}. \tag{6.4}$$

In any particular case, the output impedance of a sensor should be defined. This helps to analyze a frequency response and phase lag of the sensor-conditioner combination. For instance, a piezoelectric sensor that can be represented by a very high-output resistance (on the order of $10^{11}\,\Omega$) shunted by a capacitance (in the order of 10 pF) is modeled as a current-generating sensor.

To illustrate importance of the input impedance characteristics, let us consider a purely resistive sensor $Z_{out} = R_{out}$ connected to the input impedance as shown in Fig. 6.2a, b. A combined resistance at the circuit's input becomes:

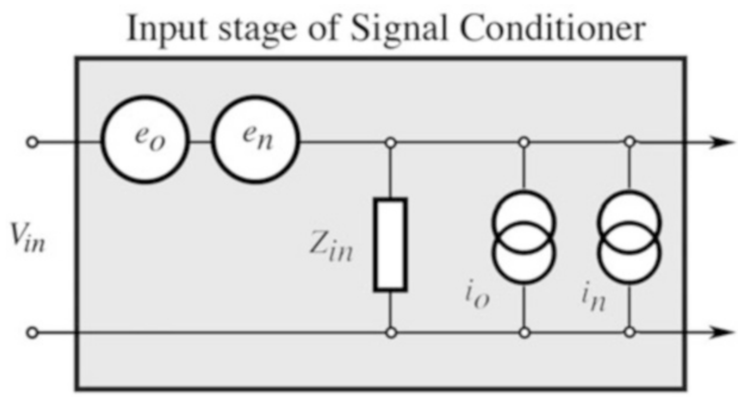
$$R_{in} = \frac{R_{out}R}{R_{out} + R} \tag{6.5}$$

The circuit's input voltage as function of frequency, f, can be expressed by a formula:

$$V_{in} = \frac{E}{\sqrt{1 + \left(\frac{f}{f_c}\right)^2}} \tag{6.6}$$

where $f_c = (2\pi R_{in}C)^{-1}$ is the corner frequency (i.e., the frequency where the voltage amplitude drops by 3 dB). If we assume a required accuracy in detection of the

Fig. 6.3 Equivalent circuit of electrical noise sources at input stage



amplitude as 1 %, we can calculate the maximum stimulus frequency, which can be processed by the circuit:

$$f_{\max} \approx 0.14 f_{\mathrm{c}}, \tag{6.7}$$

or $f_c \approx 7f_{max}$; that is, the impedance must be selected in such a way as to assure a sufficiently high-corner frequency. For example, if the stimulus' expected highest frequency is 1 kHz, the corner frequency of the signal conditioner must be selected at least at 7 kHz. In practice, f_c is selected even higher, because of additional frequency limitations in the subsequent circuits.

For a voltage-generating sensor $R_{out} \ll R$, then according to Eq. (6.5), $R_{in} \approx R_{out}$, thus f_c becomes very large. This lead to $V_{in} \approx E$, and as a result, the input stage of a signal conditioner does not distort the sensor's signal. Therefore, for the voltage-generating sensors, an input resistance of a signal conditioner should be as high as practical. The input capacitance makes no difference because it is shunted by a low-output resistance of the sensor.

Figure 6.3 is a more detailed equivalent circuit of the input properties of a signal conditioner, for instance, an amplifier. The circuit is modeled by the input impedance $\mathbf{Z}_{in}$ and several generators producing interfering signals. They represent voltages and currents that are spuriously generated by the circuit itself or picked up from various external sources, even if the sensor generates no signal. These undesirable voltages and currents may pose substantial problems if not handled properly. Besides, many such noise generators are temperature dependent.

Voltage e_o is nearly constant and called the input *offset voltage*. If the input terminals of the signal conditioner are shorted together (zero input signal), that voltage would simulate a presence of a virtual d.c. voltage input signal having value e_o. It should be noted that the offset voltage source is connected in series with the input and its resulting error is independent of the output impedance of the sensor or its signal, that is, the offset voltage has an *additive nature*.

The input *bias current* i_o is also generated internally by the front stage or introduced from some external interfering sources, such as circuit board leakage currents. Its value is quite high for many input bipolar transistors, much smaller for JFETs, and further much lower for the CMOS circuits. This current may present a serious problem when a sensor has a high-output impedance (a current-generating sensor). The bias current passes through the output resistance of the sensor, resulting in a spurious voltage drop. This voltage drop may be of a significant magnitude. For instance, if a piezoelectric sensor is connected to a circuit having a combined (Eq. 6.5) input resistance of 1 GΩ (10^9 Ω) and the input bias current is 100 pA (10^{-10} A), the voltage drop at the input becomes equal to 1 GΩ × 100 pA = 0.1 V—a very high erroneous value, indeed. Error resulting from a bias current is proportional to the combined resistance of the sensor and interface circuit. This error is negligibly small for sensors having low-output resistances (voltage generating) and thus can be ignored. For instance, an inductive detector is not sensitive to bias currents.

A circuit board *leakage current* may be a source of errors while working with the current-generating sensors. This spurious current may be resulted from a reduced surface resistance in a printed circuit board (PCB). Possible causes for that are a poor quality PCB material, surface contaminations by a solder flux residue (a poorly washed PCB), moisture condensation, and degraded conformal coating. Figure 6.4a shows that a power supply bus and board resistance, R_L, may cause leakage current, i_L, through the front stage combined impedance. If the sensor is capacitive, its output capacitance will be very quickly charged by the PCB leakage current. This will not only cause error, but may even lead to a sensor's destruction if the sensor uses chemical compounds (e.g., a humidity sensor).

There are several techniques known to minimize the board leakage current effect. One is a careful board layout to keep higher voltage conductors away from the high-impedance components. A leakage through the board thickness in the multilayer boards should not be overlooked. Another method is electrical guarding, which is an old trick. The so-called driven shield is also highly effective. Here, the input circuit is surrounded by a conductive trace that is connected to a low-impedance point at the same potential as the input. The guard absorbs leakage from other points on the board, drastically reducing spurious currents that may reach the input terminal. To be completely effective, there should be guard rings on

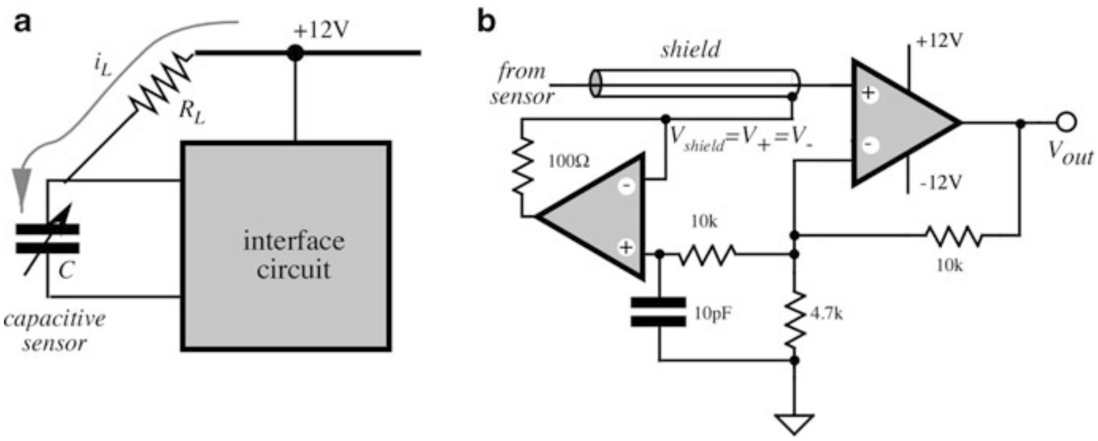


Fig. 6.4 Circuit board leakage affects input stage (**a**); driven shield of input stage (**b**)

both sides of the PCB. As an example, an amplifier is shown with a guard ring, driven by a relatively low impedance of the amplifier's inverting input.

It is highly advisable to position high-input impedance signal conditioners as close as possible to the sensors. However, sometimes the connecting lines cannot be avoided. The coaxial shielded cables with good isolation are recommended [1]. Polyethylene or virgin (not reconstructed) Teflon is best for critical applications. However, even short cable runs can reduce bandwidth unacceptably with high sensor resistances. These problems can be largely avoided by bootstrapping the cable's shield. Figure 6.4b shows a voltage follower connected to the inverting input of an amplifier. The follower drives the shield of the cable, thus reducing the cable capacitance, the leakage, and spurious voltages resulting from cable flexing. A small capacitance at the follower's noninverting input improves its stability.

Another problem that must be avoided is connecting to input of a signal conditioner any components, besides a sensor, that potentially may cause problems. An example of such a "troublemaker" is a ceramic capacitor. In a hope to filter out high frequency transmitted noise at the input, a designer quite frequently uses filter capacitors either at the input, or in the feedback circuit of an input stage. If for a cost saving or the space saving reason a ceramic capacitor is selected—she may get what she is not expecting. Many capacitors possess the so-called dielectric absorption which is manifested as a memory effect. If such a capacitor is subjected to a charge spike either from a sensor, or from a power supply, or just from any external noise source, the charge will alter the capacitor's dielectric properties in such a way that the capacitor starts behaving like a small battery. That spurious "battery" may take a long time to lose its charge—from few seconds to many hours. Voltage generated by that "battery" is added to the sensor's output signal and may cause significant errors. If a capacitor must be employed at the input stage, a film capacitor should be used instead of ceramic.

6.1.2 Amplifiers

Many sensing elements produce weak output signals. Magnitudes of these signals may be on the order of microvolts (μV) for the voltage-generating sensors or picoamperes (pA) for the current-generating sensors. On the other hand, standard electronic data processors, such as analog-to-digital converters (ADC), frequency modulators, data recorders, etc. require input signals of sizable magnitudes—in the order of volts (V). Therefore, amplification of the sensor output signals has to be made with a voltage gain up to 10,000 and a current gain up to one million. Amplification is part of a signal conditioning. There are several standard configurations of the amplifiers that might be useful for amplifying low-level signals. These amplifiers may be built of discrete components, such as semiconductors, resistors, capacitors, and inductors. Nowadays, amplifiers are frequently composed of standard building blocks, such as operational amplifiers (OP-AMPs) that are augmented with various discrete components.

The purpose of an amplifier is much broader than just increasing the signal magnitude. An amplifier also is an impedance matching device, an enhancer of a signal-to-noise ratio, a frequency filter, and an isolator between the sensor and rest of the circuit.

6.1.3 Operational Amplifiers

One of the principle building blocks for amplifiers is the so-called *operational amplifier* or OP-AMP, which is either an integrated (monolithic) or hybrid (a combination of monolithic and discrete parts) circuit. An integrated OP-AMP may contain hundreds of transistors, diodes, as well as resistors and capacitors. An analog circuit designer, by arranging around the OP-AMP discrete components (resistors, capacitors, inductors, etc.), may create endless number of useful circuits—not only amplifiers, but also many other circuits as well. OP-AMPs are also used as cells in custom-made integrated circuits—analog or mixed technology. A custom circuit is called *application-specific integrated circuit* or ASIC, for short. Below, we will describe some typical circuits with OP-AMP, which are often used as front ends of various signal-conditioning circuits.

As a building block, a good operational amplifier has the following properties (a symbol representation of OP-AMP is shown in Fig. 6.5a):

- Two inputs: one is inverting (−) and the other is noninverting (+).
- High-input resistance (on the order of GΩ).
- Low-output resistance (a fraction of Ω), mostly independent of a load.
- Ability to drive capacitive loads without becoming unstable.
- low input offset voltage e_0 (few mV or even few μV).
- low input bias current i_0 (few pA or even less).
- Very high *open-loop gain* A_{OL} (at least 10^4 and preferably over 10^6). That is, the OP-AMP must be able to magnify (amplify) a voltage difference V_{in}, between its two inputs by a factor of A_{OL}.

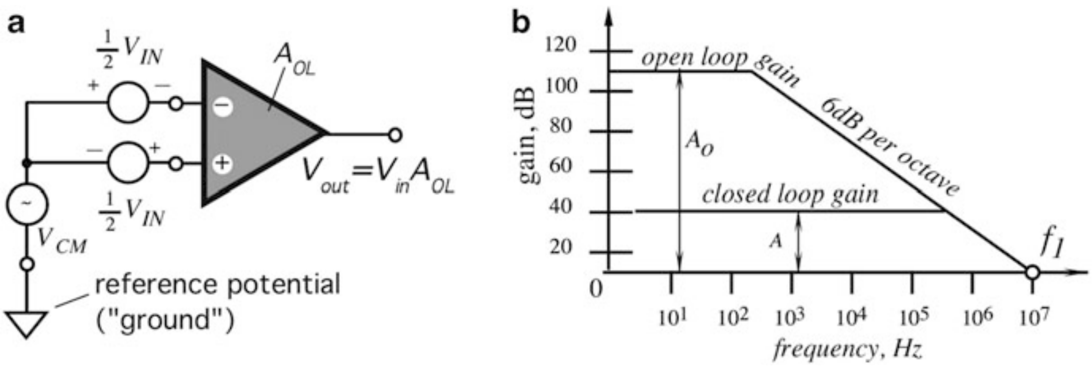


Fig. 6.5 General symbol of operational amplifier (**a**), and gain-frequency characteristic of OP-AMP (**b**)

– High common mode rejection ratio (CMRR). That is, the amplifier suppresses the in-phase equal magnitude input signals (common-mode signals) V_{CM} applied to its both inputs.
– Low intrinsic noise.
– Broad operating frequency range.
– Low sensitivity to variations in the power supply voltage.
– High environmental stability of its own characteristics.

For detailed information and application guidance the reader should refer to data sheets and catalogues published by the respective manufacturers and available online. Such catalogues usually contain selection guides for every important feature of an OP-AMP. For instance, OP-AMPs are grouped by such criteria as low offset voltages, low-bias currents, low noise, bandwidth, etc.

Figure 6.5a depicts an operational amplifier without feedback components. Therefore, it operates under the so-called *open-loop* conditions. An open-loop gain, A_{OL}, of an OP-AMP is always specified but is not a very stable parameter. Its frequency dependence may be approximated by a graph of Fig. 6.5b. The A_{OL} changes with the load resistance, temperature, and power supply fluctuations. Many amplifiers have an open-loop gain temperature coefficient in the order of 0.2–1 %/°C and the power supply gain sensitivity in the order of 1 %/%. An OP-AMP as a linear circuit is very rarely used with an open loop (without the feedback components) because the high open-loop gain may result in circuit instability, a strong temperature drift, noise, etc. For instance, if the open-loop gain is 10^5, the input voltage drift of 10 μV (ten microvolts) would cause the output drifts by about 1 V.

Ability of an OP-AMP to amplify small magnitude high-frequency signals is specified by the *gain-bandwidth product* (GBW) which is equal to the frequency f_1 where the amplifier's open-loop gain becomes equal to unity. In other words, above the f_1 frequency, the amplifier cannot amplify.

An example of a feedback with a resistive divider R_1 and R_2 is shown in Fig. 6.6a. The input voltage is applied to a noninverting (+) input of the OP-AMP. This input has a very high-input impedance. The feedback resistors convert the Op-AMP to a noninverting amplifier, where the resulting closed-loop gain:

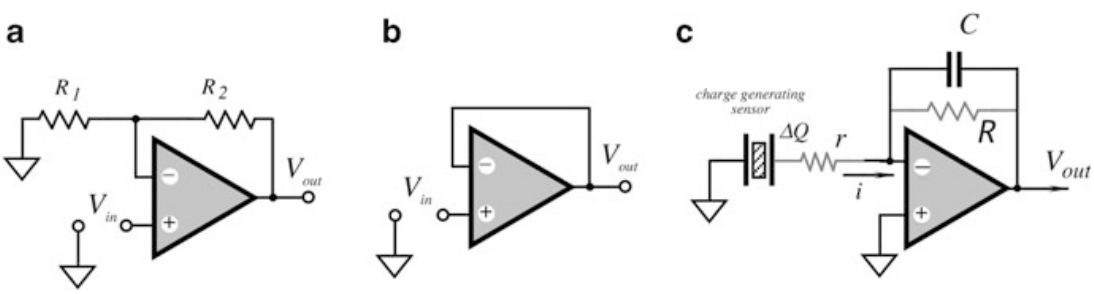


Fig. 6.6 Noninverting amplifier (**a**); voltage follower (**b**); charge-to-voltage converter (**c**)

$$A = 1 + \frac{R_2}{R_1} \tag{6.8}$$

Considering the A_{OL} being very large, the closed-loop gain A depends only on the feedback components and is nearly constant over a broad frequency range (see Fig. 6.5b). However, f_1 is still the frequency limiting factor, regardless of the feedback. Linearity, gain stability, and output impedance—all are improved by the amount of a feedback. The feedback may have various linear components, including resistors, capacitors, inductors, as well as nonlinear components, such as diodes. As a general rule for a moderate accuracy, the open-loop gain of an OP-AMP should be at least 100 times greater than the closed-loop gain at the highest frequency of interest. For even higher accuracy, the ratio of the open- and closed-loop gains should be 1000 or more.

6.1.4 Voltage Follower

A voltage follower shown in Fig. 6.6b is an electronic circuit that provides impedance conversion from a high to low level. It is a particular case of the amplifier shown in Fig. 6.6a where R_1 is removed ("infinite" value) and $R_2 = 0$. Then, according to Eq. (6.8) the closed-loop gain is unity. A typical follower has high-input impedance (very high-input resistance and low-input capacitance) and very low-output resistance (the output capacitance makes no difference). A good follower has a voltage gain being very close to unity (typically, 0.999 over a broad frequency range). The buffering properties—high-input and low-output impedances make it indispensable for interfacing between many sensors and signal processing devices.

When designing a follower, these tips might be useful:

- For the current-generating sensors, the input bias current of the follower must be at least 100 times smaller than the sensor's current.
- The input offset voltage must be either trimmable or smaller than the required LSB.
- A temperature coefficient of the bias current and the offset voltage should not result in errors of more than 1 LSB over an entire temperature range.

6.1.5 Charge- and Current-to-Voltage Converters

Charge-to-voltage converters (CVC) are employed to convert signals from the charge-generating sensors. Like a voltage follower, a CVC is a buffer between the charge sensor and ADC that requires voltage from a low-impedance source. A basic circuit of a CVC is shown in Fig. 6.6c. Capacitor, C, is connected into a negative feedback network of an OP-AMP. Its leakage resistance R must be substantially larger than impedance of the capacitor at the lowest operating frequency. A good film capacitor is usually recommended along with a good quality printed circuit board where the components are coated with conformal coating.

A transfer function of the converter is:

$$V_{\text{out}} = -\frac{\Delta Q}{C}. \tag{6.9}$$

If the sensor is of a capacitive nature (many charge-generating sensors are), connecting it to an inverting input of OP-AMP may cause instability at high frequencies. In other words, the amplifier may oscillate—a highly undesirable behavior. To prevent oscillations, a small resistor r should be added in series with the capacitive sensor.

Note, that when the OP-AMP operates, the feedback keeps voltage at its inverting input (−) very close to the noninverting voltage that in this circuit is zero (ground). That is why the noninverting input is called a *virtual ground* as compared to the real ground of the noninverting input. In may practical circuits, a provision should be made to periodically discharge the feedback capacitor, for example by use of a parallel analog switch.

Many sensors can be modeled as current generators. An example is a photodiode. A current-generating sensor is represented by a very large leakage resistance, r, connected in parallel with a current generator (a double-circle symbol) that by definition has an infinitely high internal resistance, Fig. 6.7a. To convert current to voltage, it shall be pushed through a load resistor R. The sensor current, i_0, has two ways to outflow: through the sensor's internal leakage resistance, r, as current i_r, and also through the load resistor as i_{out}. Current i_r is useless, thus to minimize error, leakage resistance r of the sensor must be much larger than the load resistor R.

According to Ohm's law, output voltage V is proportional to magnitude of the current and resistor R. Note that this voltage also appears across the sensor. This could be undesirable, since it may cause some errors, including nonlinearity and frequency limitations. To alleviate that problem, a special electronic circuit called current-to-voltage converter is employed. One of its functions is to keep voltage

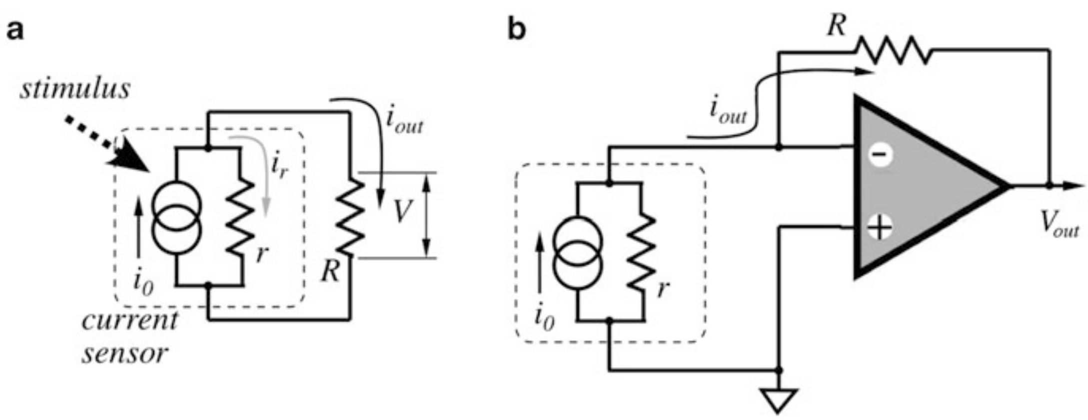


Fig. 6.7 Current-to-voltage converters

across the sensor on a constant level, often zero. Figure 6.7b shows a converter where the sensor is connected to a virtual ground (inverting input of the OP-AMP) and thus always is at a zero potential (the same as the grounded noninverting input). This is because of a very large A_{OL} that via the feedback keeps both inputs of the OP-AMP very close to one another. Another advantage of the virtual ground input is that the output voltage does not depend on the sensor's capacitance, no matter how large, and, as a result, has a much wider frequency response comparing with the basic circuit of Fig. 6.7a. The output voltage of the circuit:

$$V_{out} = -iR \tag{6.10}$$

Minus sign indicates that the output voltage is negative (below ground) for the positive (flowing in) currents.

6.1.6 Light-to-Voltage Converters

Light-to-voltage converters are required for converting output signals from photosensors to voltage. For detecting extremely low-intensity light—typically few photons—a photomultiplier is generally employed (Sect. 16.1); however, for less demanding applications, three types of photosensors are available: a photodiode, phototransistor, and photoresistor (Chap. 15). They all employ a photoeffect that was discovered by A. Einstein and won him the Nobel Prize. These photo sensors are called *quantum* detectors. The difference between a photodiode and a phototransistor is in construction of a semiconductor chip. A photodiode has one p-n junction, while a phototransistor has two junctions where the base of the transistor may be floating or may have a separate terminal. The base current is a photoinduced current that is multiplied by the transistor's β (current gain) to produce the collector current, that in turn can be converted to voltage as described above. Thus, a phototransistor is equivalent to a photodiode with a built-in current amplifier. The quantum detectors are the *current-generating* sensors having very large internal resistances.

From the electrical point of view, a photodiode can be represented by an equivalent circuit shown in Fig. 6.8a. It consists of a current generator (internal

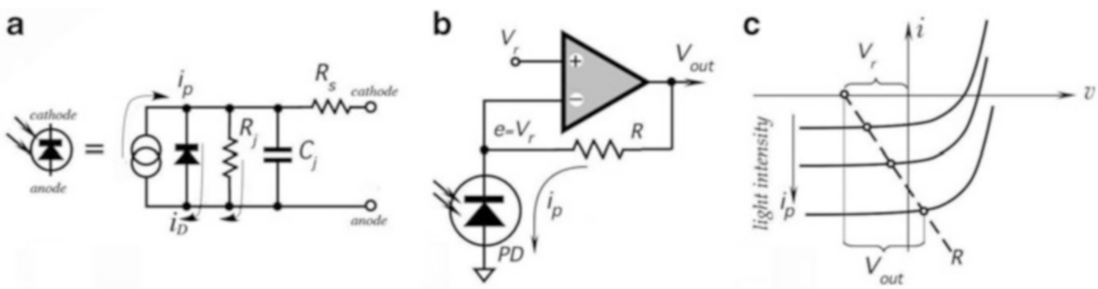


Fig. 6.8 Equivalent circuit of photodiode (**a**). Reverse-biased photodiode with current-to-voltage converter (**b**). Load diagram of circuit (**c**)

input impedance is infinitely large), a parallel regular diode (like a rectifier diode), resistance of the diode junction R_j, capacitance of the junction C_j, and a serial resistance R_s. The current generator produces current proportional to the absorbed photon flux. This current flows in the direction from the cathode to the anode of the photodiode, that is in the opposite direction to that where the diode would normally conduct. Note that for very strong illuminations, the portion i_D of the photocurrent i_p will start flowing through a nonlinear rectifier diode which will degrade the sensor's linearity.

A photodiode can be used in a voltaic or current modes. In the voltaic mode, a photodiode is connected to a very high resistor (10^7–10^9 Ω) and a good voltage amplifier. The diode will work like a battery where voltage is nearly proportional to the light intensity. This voltage is the result of a photocurrent i_p passing through the internal junction resistance R_j. In a current mode, the photodiode is either kept with a constant voltage across its terminals or virtually shorted (a voltage across the diode is held at zero) and current i_p is drawn to the current-to-voltage converter as described below. This mode is far more popular, especially for applications where a high-speed response is required.

A circuit with an operational amplifier is shown in Fig. 6.8b. Note that the reference voltage V_r creates a constant reverse bias across the photodiode. Figure 6.8c shows the operating points for a load feedback resistor R. Advantages of the circuits used with a reverse-biased photodiode are high-speed response and wide linear range of output. Therefore, this circuit is generally used. Frequently when illumination flux is rather small, the bias voltage is not applied, but rather the noninverting input of the OP-AMP is grounded.

The interface circuits for a phototransistor are similar, except that they have to provide a voltage across the collector-emitter terminals as shown in Fig. 6.9a. The transfer function of this circuit is shown in Fig. 6.9b. A phototransistor circuit is more sensitive to light but for the price of higher nonlinearity at stronger irradiances.

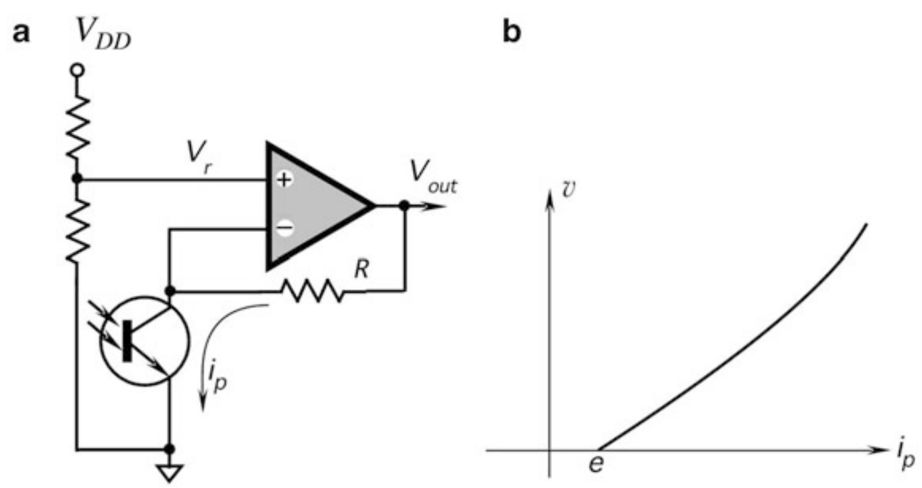


Fig. 6.9 Light-to-Voltage converted with phototransistor (**a**); transfer function (**b**)

6.1.7 Capacitance-to-Voltage Converters

Capacitive sensors are very popular. Nowadays, micromachining technologies allow fabrication of small monolithic capacitive sensors. Examples include a pressure transducer with a thin silicon diaphragm as a movable plate of the variable-gap capacitor. In a mechanical capacitive sensor, for example, an accelerometer or pressure sensor, a moving plate of a capacitor (diaphragm or proof mass), moves with respect to a stationary plate, thus modulating the capacitance that exists between the plates. This sensor is called a capacitive displacement sensor and presently is one of the widely produced sensors by employing MEMS technologies.

All capacitive sensors can be divided into asymmetrical and symmetrical structures. In a mechanical asymmetrical structure, Fig. 6.10a, the capacitance C change is measured only between the force-sensitive moving plate and a single stationary conductive plate (electrode). In a symmetrical structure, Fig. 6.10b, the capacitance measurement is performed between the moving plate and two conductive electrodes placed on both sides of the plate, rendering the differential capacitance ($C_1 - C_2$) measurement possible.

The principle problem with these tiny capacitors is a relatively low capacitance value per unit area of the plate (about 2 pF/mm^2) which may result is large die sizes. A typical capacitive pressure sensor offers a zero pressure capacitance on the order of few picofarads, so that a 10-bit resolution requires the detection of capacitive shifts on the order of 15 fF or less (1 femtofarad $= 10^{-15}$ F). This difficulty may be reduced by further narrowing a gap between the plates (down to a couple of micrometers) or even maintaining the gap on a nearly constant level by providing a force feedback, as described in Sect. 6.1.8. It is obvious that any external measurement circuit will be totally impractical, as a parasitic capacitance of connecting conductors at best can be on the order of 1 pF—too much with respect to the sensor capacitance. Therefore, the only way to make such a sensor practical is to build a signal conditioning and other interface circuits as an integral part of the sensor itself.

One quite effective way of designing such a capacitance-to-voltage (C/V) is to use a switched capacitor technique. The technique is based on a charge transfer from one capacitor to another by the solid-state analog switches.

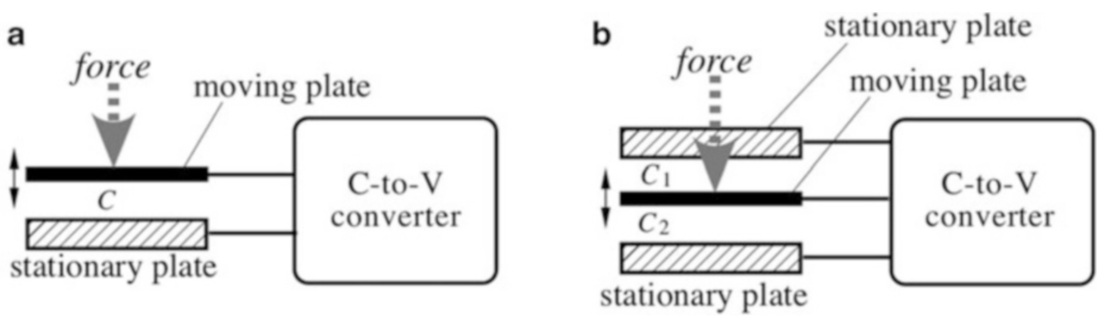


Fig. 6.10 Asymmetrical (**a**) and symmetrical (**b**) capacitive displacement sensor

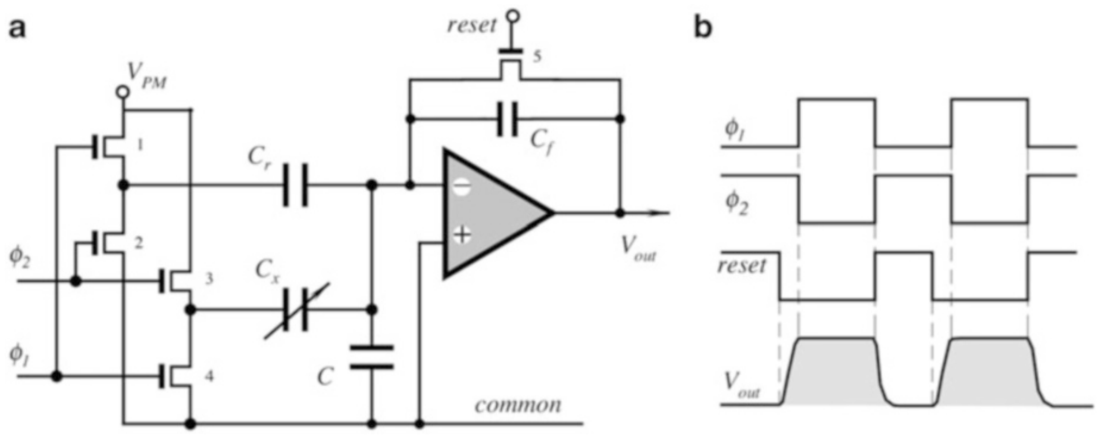


Fig. 6.11 Simplified schematic (**a**) and timing diagrams (**b**) of capacitance-to-voltage converter using switched-capacitors technique

Figure 6.11a shows a simplified circuit diagram of a differential switched-capacitor converter [2], where variable capacitance C_x and reference capacitance C_r are parts of a symmetrical silicon pressure sensor. The same circuit can be used for an asymmetrical sensor with C_r being a trimmed reference capacitor.

Monolithic MOS switches (1–4) are driven by the opposite phase clock pulses, ϕ_1 and ϕ_2. When the clocks switch, a charge appears at the common capacitance node—inverting input of OP-AMP. The charge is supplied by the constant voltage source, V_{PM}, and is proportional to $(C_x - C_r)$ and, therefore to the applied pressure in the sensor. This charge is fed to a charge-to-voltage converter which includes an operational amplifier, integrating capacitor C_f, and MOS discharge (reset) switch 5. The output signal is the variable-amplitude pulses shown in Fig. 6.11b. These pulses can be demodulated to produce a linear signal or can be directly converted into digital data. So long as the open-loop gain of the integrating OP-AMP is high, the output voltage is insensitive to the input capacitance C, offset voltage, and temperature drift. The minimum detectable signal (noise floor) is determined by the component noise and temperature drifts of the components.

When the MOS switch goes from the on-state to the off-state, the switching signal at the gate injects some charge from the gate to the inverting input of the OP-AMP. An injection charge results in the offset voltage at the amplifier output. This error can be compensated for by a charge-cancelling device [3], which can improve the signal-to-noise ratio by two orders of magnitude of the uncompensated charge.

For the capacitive sensors having much larger capacitances (10–1000 pF), simpler techniques can be employed. One is use of an *RC* or *LC* oscillator that converts value of a variable *C* into a variable frequency or duty cycle of an a.c. signal. Figure 6.12a illustrates an *LC* oscillator whose frequency depends on both the variable capacitor and fixed inductor:

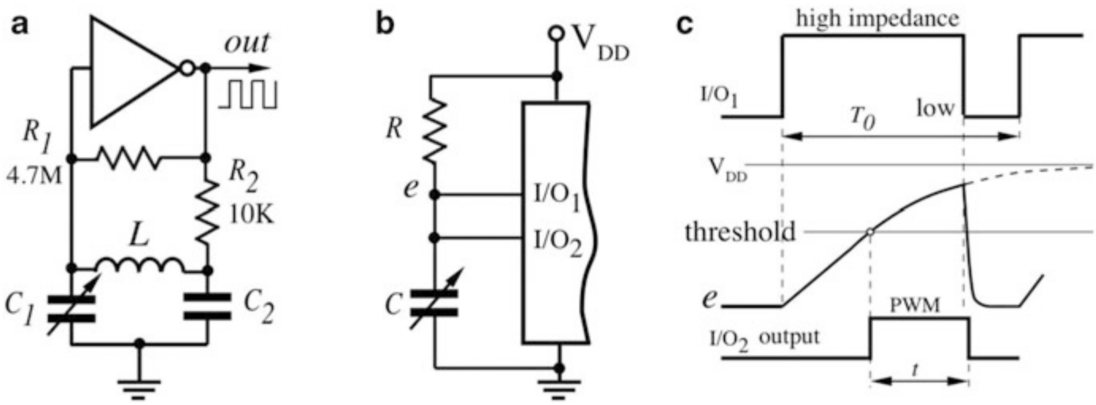


Fig. 6.12 LC oscillator (**a**); microcontroller converter of capacitance to PWM signal (**b**); timing diagram of PWM converter (**c**)

$$f = \frac{1}{2\pi\sqrt{L\frac{C_1+C_2}{C_1C_2}}} \tag{6.11}$$

Another circuit has a relaxation network of a fixed resistor R and variable sensor's capacitance C, Fig. 6.11b. The network is connected to a microcontroller having two I/O ports, where I/O_1 generates the tristate square pulses with a period T_0, alternating between a high impedance and ground low. The other I/O_2 is a digital input with a triggering threshold of about $0.5V_{DD}$. Preferably, it should have a Schmitt-trigger input. During the I/O_1 low state, capacitor C is discharged, see Fig. 6.11c. When the I/O_1 goes high impedance, the sensor's capacitor C charges through resistor R to V_{DD} voltage. At the moment of crossing the threshold, the I/O_2 registers the crossover and the microcontroller computes duty cycle M that is proportional to the sensor capacitance C:

$$M = \frac{0.693R}{T_0}C \tag{6.12}$$

6.1.8 Closed-Loop Capacitance-to-Voltage Converters

Closed feedback loop in a capacitance-to-voltage converter has the ability to extend a dynamic range, increase linearity, flatten frequency response, and improve cross-axis rejection in accelerometers. The idea behind the method is to generate a compensating force that would prevent the moving plate of a capacitive sensor from shifting from its balance position [4]. It is a practical implementation of the null-balanced bridge concept as described in Sect. 6.2.4. Figure 6.13a illustrates the closed-loop block diagram where the capacitive sensor is subjected to a difference force: the deflecting force caused by a mechanical stimulus (pressure, acceleration, sound, etc.) minus the feedback force from the voltage-to-force transducer (V-F). The difference force is sensed by a capacitive displacement sensor, converted into electric signal, amplified, and applied to the controller that modulates the V-F transducer. The mechanical feedback maintains the difference force applied to

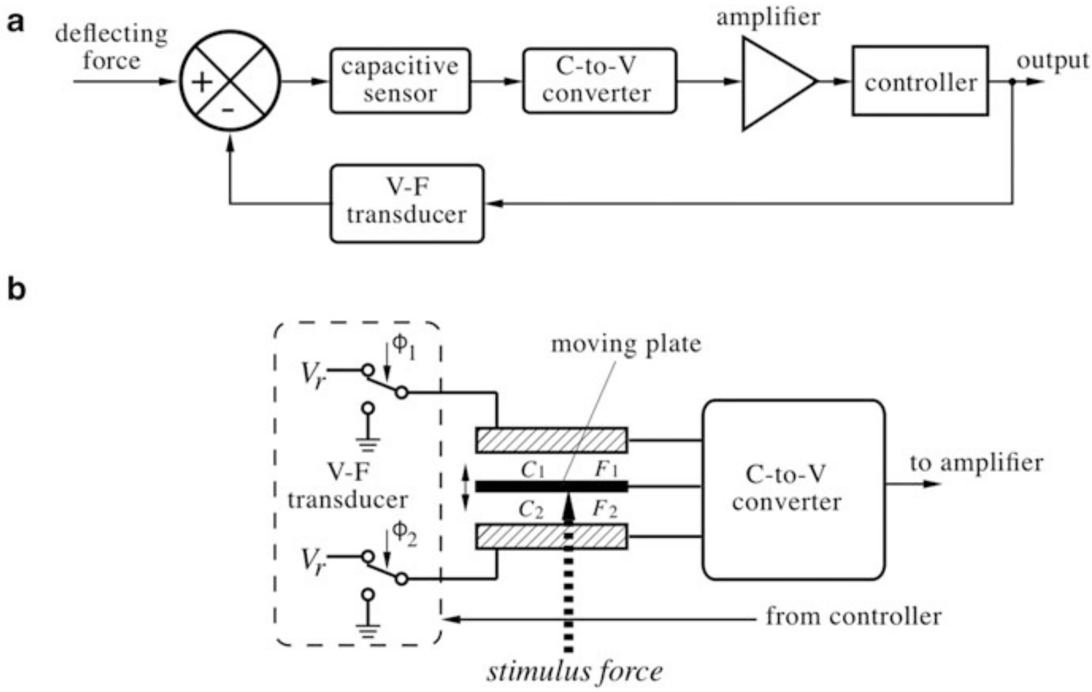


Fig. 6.13 Block diagram of closed-loop capacitive signal conditioner (**a**) and use of voltage-to-force transducer to generate electrostatic forces for balancing symmetrical capacitive sensor (**b**)

the moving plate of the sensor close to zero, thus the stimulus and compensating forces are nearly equal in magnitude. As a result, instead of measuring the deflecting force produced by the stimulus, the voltage that controls the V-F transducer is used as the output signal.

Design of a V-F transducer is not a trivial task. The most practical method for use with MEMS is to employ an electrostatic force that appears between the plates of a symmetrical capacitive sensor in response to a voltage gradient across the plates. This force F_e can be expressed through U—the potential difference between the capacitor plates (voltage), the distance between the plates d, $\varepsilon_0\kappa$—dielectric constant of the space between the plates, and the plate area A:

$$F_e = \frac{\varepsilon_0 \kappa}{2} \int_A \frac{U^2}{d^2} dA, \tag{6.13}$$

The electrostatic feedback force can be generated by a pulsed voltage. If the pulse rate of the applied voltage is essentially above the transducer dynamic response cutoff frequency (that is, the lowest natural frequency of the sensor), the moving plate is subjected to an average electrostatic force. Note that the C-to-V converter uses a.c. signal to measure capacitances C_1 and C_2 between the plates, however, frequency of this a.c. signal must be much higher then the electrostatic modulation rate. Thanks to the frequency difference, it is easy to separate them by employing the appropriate filters.

Figure 6.13b illustrates the V-F transducer where two switches alternate voltage V_r and ground for applying to the upper and lower electrodes, where the moving plate is at virtual ground (zero) potential. The stimulus force is applied to the moving plate whose movement is suppressed by the electrostatic forces F_1 and F_2 produced by the applied voltage V_r with the phase and pulse-width modulation (PWM) of the switching pulses. Phases of these pulses ϕ_1 and ϕ_2 are arranged to oppose the stimulus force. A differential capacitive sensor based on controlled force equilibrium excels by its linearity and low temperature dependence, since the central plate can be kept stationary by virtue of the applied feedback.

6.2 Sensor Connections

A sensor may be directly connected to a signal conditioner, but often it is desirable to reduce errors and noise caused by various interfering sources even before the signal is conditioned. Some errors can be reduced or even entirely eliminated by use of special sensor connections in front of the signal conditioner. Below are descriptions of the most popular techniques.

6.2.1 Ratiometric Circuits

A powerful method of improving accuracy of a sensor is a *ratiometric* technique. It should be emphasized, however, that the method is useful only if a source of error has a *multiplicative* nature but not additive. That is, the technique is useless for reduction, for instance, thermal noise. On the other hand, it is quite potent to solve problems as dependence of sensor's sensitivity to such factors as power supply instability, ambient temperature, humidity, pressure, effects of aging, etc. The technique essentially requires the use of two sensors where one is the acting sensor that responds to an external stimulus and the other is a compensating sensor, which is either shielded from that stimulus or is insensitive to it. Both sensors must be exposed to all other external effects, which may multiplicatively change their performance. The second sensor, which is often called *reference*, must be subjected to a reference stimulus, which is ultimately stable during the lifetime of the product. In many practical systems, the reference sensor is not necessarily exactly similar to the acting sensor, however its physical properties, which are subject to instabilities, should be the same.

A ratiometric technique essentially requires the use of a division. It can be performed by two standard methods: digital and analog. In a digital form, output signals from both the acting and the reference sensors are multiplexed and converted into binary codes in an analog-to-digital converter (ADC). Subsequently, a computer or a microprocessor performs the operation of a division. In an analog form, a divider may be part of a signal conditioner. A "divider", Fig. 6.14a, produces an output voltage or current proportional to a ratio of two input voltages or currents or numbers:

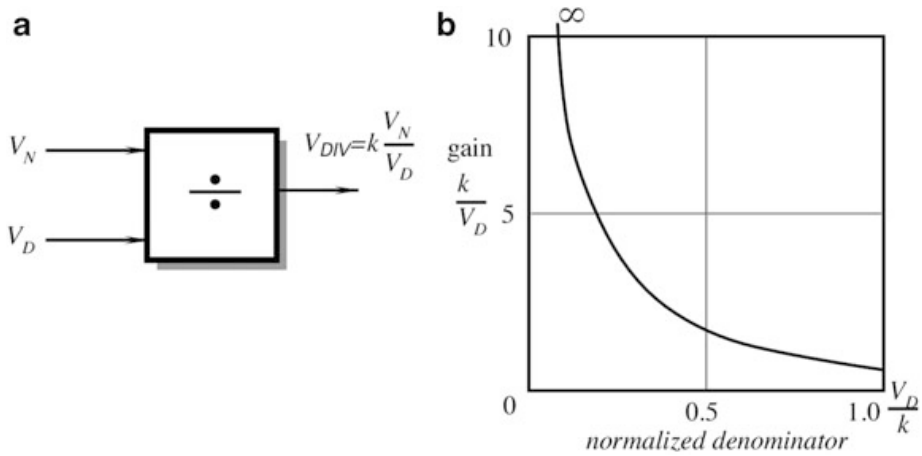


Fig. 6.14 Symbol of divider (**a**) and gain of divider as function of denominator (**b**)

$$V_{\mathrm{DIV}} = k\frac{V_{\mathrm{N}}}{V_{\mathrm{D}}}, \tag{6.14}$$

where the numerator is denoted as V_{N}, the denominator V_{D}, and k is equal to the output voltage, when $V_{\mathrm{N}} = V_{\mathrm{D}}$. The operating ranges of the variables (quadrants of operation) are defined by the polarity and magnitude ranges of the numerator and denominator inputs, and the output. For instance, if V_{N} and V_{D} are both either positive or negative, the divider is of a 1-quadrant type. If the numerator is bipolar, the divider is 2-quadrant. Generally, the denominator is restricted to a single polarity, since the transition from one polarity to another would require the denominator to pass through zero, which would call for an infinite output (unless the numerator is also zero). In practice, V_{D} is a signal from a reference sensor, which usually has a relatively constant value.

Division has long been the most difficult of the four arithmetic functions to implement. This difficulty stems primarily from the nature of division: the magnitude of a ratio becomes quite large, approaching infinity, for a denominator that is approaching zero (and a nonzero numerator). Thus, an ideal divider must have a potentially infinite gain and infinite dynamic range. For a real divider, both of these factors are limited by the magnification of drift and noise at low values of V_{D}. That is, the gain of a divider for a numerator is inversely dependent on the value of the denominator, Fig. 6.14b. Thus, the overall error is the net effect of several factors, such as gain dependence of denominator, numerator, and denominator input errors, like offsets, noise, and drift (which must be much smaller than the smallest values of input signals). Besides, the output of the divider must be constant for constant ratios of numerator and denominator, independent of their magnitudes. For example, $10/10 = 0.01/0.01 = 1$ and $1/10 = 0.001/0.01 = 0.1$. In practice, some simple division circuits are used quite extensively. An example is an amplifier of Fig. 6.15b whose output signal is function of the resistor ratio (note that the reference voltage V_{r} is negative):

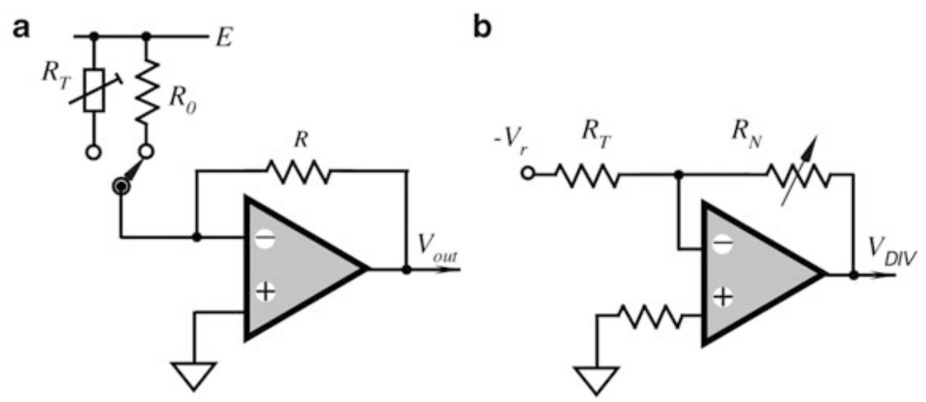


Fig. 6.15 Ratiometric temperature detector (**a**) and analog divider of resistive values (**b**)

$$V_{\mathrm{DIV}} = V_{\mathrm{r}}\frac{R_{\mathrm{N}}}{R_{\mathrm{T}}}, \tag{6.15}$$

The most popular and efficient ratiometric circuits are based on the Wheatstone bridge designs which are covered below.

To illustrate effects of a ratiometric technique, consider Fig. 6.14a that shows a simple temperature detector where the acting sensor is a negative temperature coefficient (NTC) thermistor R_T. A reference resistor R_0 has a value equal to resistance of the thermistor at some reference temperature, for instance at 25 °C. Both are connected via an analog multiplexer to the amplifier with a feedback resistor R. Let us assume that there is a some drift in the sensor value that can be described by a function of time $a(t)$ so that the sensor's resistance becomes $R_{\mathrm{T}}(t) = a(t)R_{\mathrm{T}}$. Property of the resistor R_0 is such that it also changes with the same function, so $R_0(t) = a(t)R_0$. The output signals of the amplifier produced by the sensor and the reference resistor respectively are:

$$V_{\mathrm{D}} = -\frac{ER}{a(t)R_{\mathrm{T}}} = -\frac{ER}{a(t)R_{\mathrm{T}}}, \tag{6.16}$$

$$V_{\mathrm{N}} = -\frac{ER}{a(t)R_0} = -\frac{ER}{a(t)R_0}. \tag{6.17}$$

It is seen that both voltages are functions of a power supply voltage E and the circuit gain, which is defined by resistor R. They also are functions of the drift $a(t)$. The multiplexing switch causes two voltages V_{N} and V_{D} to appear sequentially at the amplifier's output. If these voltages are fed into a divider circuit, the resulting signal is expressed as

$$N_{\mathrm{DIV}} = k\frac{V_{\mathrm{N}}}{V_{\mathrm{D}}} = k\frac{R_{\mathrm{T}}}{R_0}, \tag{6.18}$$

where k is the divider's factor (gain). Therefore, the divider's output signal does not depend on neither power supply voltage nor the amplifier gain. It also not a subject of the multiplicative drift $a(t)$. Thus, all these negative factors are rendered irrelevant. The divider's output depends only on the sensor and its reference resistor. This is true only if spurious variables, such as function $a(t)$, the power supply, or amplifier's gain, do not change rapidly. That is, they should not change appreciably during the multiplexing period. This requirement determines the rate of multiplexing.

6.2.2 Differential Circuits

Beside multiplicative interferences, the additive interferences are very common and pose serious problems for low-level output signals. Consider for example a pyroelectric sensor, Fig. 15.26a, where a heat flow sensitive ceramic plate is supported inside a metal can. Since a pyroelectric is also a piezoelectric, besides heat flow the sensor is susceptible to mechanical stress interferences. Even a slight vibration will generate a spurious piezoelectric signal that may be several orders of magnitude higher than a pyroelectric current. The solution is to fabricate the sensor with dual electrodes deposited on the same ceramic substrate as shown in Fig. 15.26b. This essentially creates two identical sensors on the same ceramic plate. Both sensors respond to all stimuli nearly identically. Since they are oppositely connected and assuming that V_{pyro} and V_{piezo} from one sensor are respectively equal to those of the other sensor, the resulting output voltage is essentially zero:

$$V_{out} = \left(V_{pyro1} + V_{piezo1}\right) - \left(V_{pyro2} + V_{piezo2}\right) = 0 \tag{6.19}$$

If one of the sensors is blocked from receiving thermal radiation ($V_{pyro2} = 0$), then $V_{out} = V_{pyro1}$. In other words, thanks to subtraction ($V_{piezo1} = V_{piezo2}$ are subtracted), the combined sensor becomes insensitive to piezoelectric spurious signals.

A differential method where a sensor is fabricated in a symmetrical form and connected to a symmetrical signal conditioning circuit (e.g., differential amplifier) so that one signal is subtracted from another, is a very powerful way of noise and drift reductions. Yet, this method is effective only if a dual sensor is fully symmetrical. An asymmetry will produce a proportional loss of noise cancellation. For example, if asymmetry is 5 %, interference will be cancelled by no more than 95 %.

6.2.3 Wheatstone Bridge

The Wheatstone bridge circuits are popular and very effective implementations of both the ratiometric technique (division) and differential techniques (subtraction) before a sensor is coupled to a signal conditioner. A general circuit of the bridge is shown in Fig. 6.16a. Impedances **Z** may be either active or reactive, that is they may be either simple resistances, like in the piezoresistive strain gauges, or capacitors, or inductors, or combinations of the above. For a pure resistor, the impedance is R, for an ideal capacitor, the magnitude of its impedance is equal to $1/(2\pi fC)$ and for an inductor, it is $2\pi fL$, where f is frequency of the current passing through the bridge arms, while at least one arm is the sensor. The bridge output voltage is represented by:

$$V_{\text{out}} = \left(\frac{\mathbf{Z_1}}{\mathbf{Z_1}+\mathbf{Z_2}} - \frac{\mathbf{Z_3}}{\mathbf{Z_3}+\mathbf{Z_4}}\right)V_{\text{ref}} = V_{\text{ref}}\left[\left(1+\frac{\mathbf{Z_2}}{\mathbf{Z_1}}\right)^{-1} - \left(1+\frac{\mathbf{Z_4}}{\mathbf{Z_3}}\right)^{-1}\right], \quad (6.20)$$

The bridge is considered to be in a balanced state when the following condition is met:

$$\frac{\mathbf{Z_1}}{\mathbf{Z_2}} = \frac{\mathbf{Z}_3}{\mathbf{Z}_4}. \quad (6.21)$$

Under the balanced condition, the output voltage is zero. When at least one impedance in the bridge changes, the bridge becomes imbalanced and the output voltage goes either in a positive or negative direction, depending on direction of the impedance change. To determine the bridge sensitivity with respect to each impedance, partial derivatives may be obtained from Eq. (6.20):

$$\begin{aligned}
\frac{\partial V_{\text{out}}}{\partial \mathbf{Z_1}} &= \frac{\mathbf{Z}_2}{(\mathbf{Z_1}+\mathbf{Z_2})^2}V_{\text{ref}} \\
\frac{\partial V_{\text{out}}}{\partial \mathbf{Z_2}} &= -\frac{\mathbf{Z}_1}{(\mathbf{Z_1}+\mathbf{Z_2})^2}V_{\text{ref}} \\
\frac{\partial V_{\text{out}}}{\partial \mathbf{Z_3}} &= -\frac{\mathbf{Z}_4}{(\mathbf{Z_3}+\mathbf{Z_4})^2}V_{\text{ref}} \\
\frac{\partial V_{\text{out}}}{\partial \mathbf{Z_4}} &= \frac{\mathbf{Z}_3}{(\mathbf{Z_3}+\mathbf{Z_4})^2}V_{\text{ref}}
\end{aligned} \quad (6.22)$$

By summing these equations, we obtain the bridge sensitivity:

$$\frac{\delta V_{\text{out}}}{V_{\text{ref}}} = \frac{\mathbf{Z}_2\delta\mathbf{Z_1} - \mathbf{Z_1}\delta Z_2}{(\mathbf{Z_1}+\mathbf{Z_2})^2} - \frac{\mathbf{Z}_4\delta\mathbf{Z_3} - \mathbf{Z_3}\delta\mathbf{Z_4}}{(\mathbf{Z_3}+\mathbf{Z_4})^2}, \quad (6.23)$$

It should noted that according to Eq. (6.20), the Wheatstone bridge possess both properties: ratiometric and differential. The ratios $\mathbf{Z_2}/\mathbf{Z_1}$ and $\mathbf{Z_4}/\mathbf{Z_3}$ are ratiometric, while the difference in the parenthesis represents a differential property, thus making the Wheatstone bridge a very useful circuit.

A closer examination of Eq. (6.23) shows that only the adjacent pairs of the impedances (i.e., $\mathbf{Z_1}$ and $\mathbf{Z_2}$, $\mathbf{Z_3}$ and $\mathbf{Z_4}$) have to be identical in order to achieve the ratiometric compensation (such as the temperature stability, drift, etc.). For the differential properties, impedances in the balanced bridge do not have to be equal, as long as the balance ratio of Eq. (6.21) is satisfied.

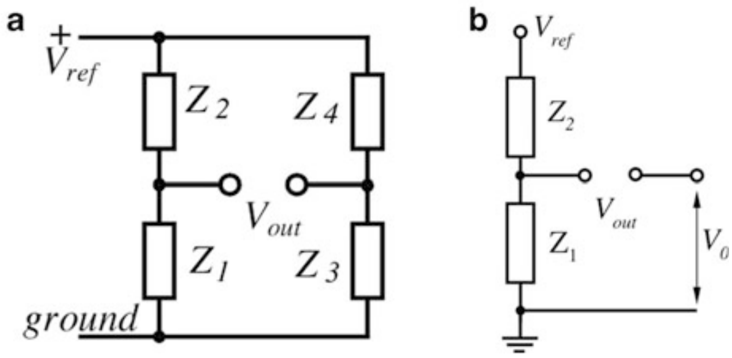


Fig. 6.16 General circuit of Wheatstone bridge (**a**) and half-bridge (**b**)

In many practical circuits, only one impedance is used as a sensor, thus for $\mathbf{Z_1}$ as a sensor, the bridge sensitivity becomes:

$$\frac{\delta V_{\text{out}}}{V_{\text{ref}}} = \frac{\delta \mathbf{Z_1}}{4\mathbf{Z_1}}. \tag{6.24}$$

A simplified version of the bridge is shown in Fig. 6.16b, where only two serial impedances are used as a voltage divider. The second divider is replaced by a fixed reference voltage V_0. As a result, the circuit that is called a *half-bridge* has no differential properties, but still possess the ratiometric properties because its output voltage is represented by:

$$V_{\text{out}} = V_{\text{ref}}\left(1 + \frac{\mathbf{Z_2}}{\mathbf{Z_1}}\right)^{-1} - V_0. \tag{6.25}$$

The resistive bridges are commonly used with strain gauges, piezoresistive pressure transducers, thermistor thermometers, hygristors, and many other sensors when immunity against environmental factors is required. Similar arrangements are used with the capacitive and magnetic sensors for measuring force, displacement, moisture, etc.

The basic Wheatstone bridge circuit (Fig. 6.16a) generally operates when the bridge is disbalanced. This is called the *deflection* method of measurement. It is based on detecting voltage V_{out} across the bridge diagonal. Let us consider a bridge with a sensor in place of impedance Z_1. When the sensor's impedance changes by the value Δ, the new impedance becomes $Z_v = Z_1(1+\Delta)$. The bridge output voltage is a nonlinear function of a disbalance Δ. However, for a small change ($\Delta < 0.05Z_1$), which often is the case, the bridge output may be considered quasilinear. The bridge maximum sensitivity is obtained when $Z_1 = Z_2$. When $Z_1 \gg Z_2$ or $Z_2 \gg Z_1$, the bridge output voltage decreases. Assuming that $k = Z_1/Z_2$, the bridge sensitivity may be expressed as:

$$\alpha = \frac{k}{(k+1)^2} \tag{6.26}$$

A normalized graph calculated from this equation is shown in Fig. 6.17. It indicates that the maximum sensitivity is achieved at $k = 1$, however, sensitivity

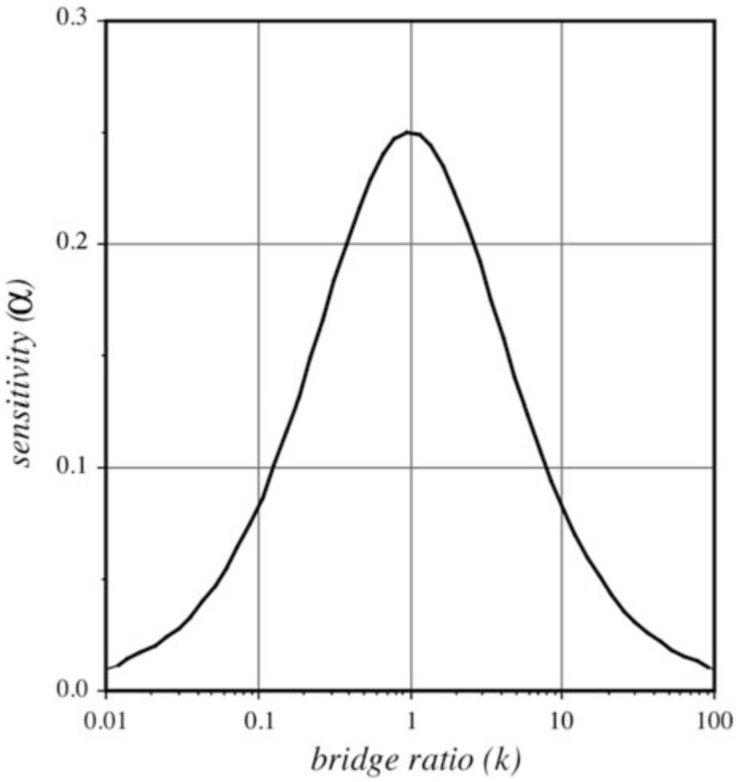


Fig. 6.17 Sensitivity of disbalanced bridge as function of impedance ratio

drops relatively little for the range where $0.5 < k < 2$. If the bridge is fed by a current source i_{ref}, rather by a voltage source V_{ref}, its output voltage for small Δ and a single variable component (sensor) is represented by:

$$V_{out} = i_{ref} \frac{k\Delta}{2(k+1)}, \tag{6.27}$$

6.2.4 Null-Balanced Bridge

Another method of using a bridge circuit is called *null-balanced*. The method overcomes the limitation of small changes (Δ) in the bridge arm to achieve a good linearity over a broad span of the input stimuli. The null-balance essentially requires that the bridge is *always* maintained near the balanced state. To satisfy the requirement for a bridge balance of Eq. (6.21), there are two methods to balance the bridge:

1. One of the impedances of the bridge, other than the sensor, is changing together with the sensor to keep the bridge balanced. Figure 6.18a illustrates this concept. The error amplifier magnifies the small bridge disbalances. The controller modifies value of Z_2 on command from the error amplifier. In effect, this is a closed-loop control circuit of the PID type (proportional-integral-differential). The output voltage is obtained from the control signal that balances the bridge

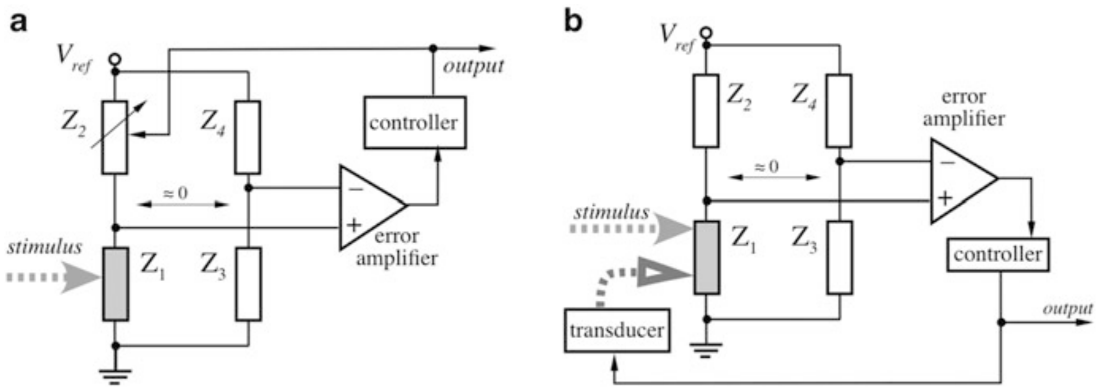


Fig. 6.18 Balancing of Wheatstone bridge by feedback to nonsensing impedance (**a**) and to sensor (**b**)

via Z_2. Consider the example: both Z_1 and Z_2 may be photoresistors—the light sensitive resistors where Z_1 is used for sensing an external light intensity. The Z_2 photoresistor could be optically coupled with a light emitting diode (LED) that is part of the controller. The controller modulates LED to modify Z_2 for balancing the bridge. When the bridge is balanced, the LED light is nearly equal to light sensed by Z_1. Therefore, electric current through the LED becomes a measure of the resistance Z_1, and, subsequently, of the light intensity detected by the sensor. Note that using impedance Z_2 for the feedback, rather than Z_3 or Z_4, is preferable as the bridge not only will stay balanced, but also the voltage across the sensor remains constant and the value of k will be close to unity, assuring the best linearity and sensitivity of the circuit.

2. The bridge balancing feedback is provided directly to the same sensor as measures the stimulus. Thus, the feedback shall be of the same nature and magnitude but of the opposite sign as the stimulus, Fig. 6.18b. When both the stimulus and "counter-stimulus" are equal to one another in magnitudes and opposite in phase, the bridge is balanced and the output of the controller becomes a measure of the stimulus. This approach was exemplified for mechanical forces in Sect. 6.1.8. It can be used when an appropriate transducer is available for converting voltage to the counter-stimulus. For example, a bidirectional force can be generated by an electrostatic effect and thus this method is useful for differential pressure sensors, accelerometers, and microphones, but there is no known way of generating, for example, "negative light" if one would like to use this option with photodetectors.

6.2.5 Bridge Amplifiers

The *bridge amplifiers* for resistive sensors are probably the most frequently used front stages of signal conditioners. They may be of various configurations, depending on the required bridge grounding and availability of either grounded or floating reference voltages. Below we review some basic circuits with operational

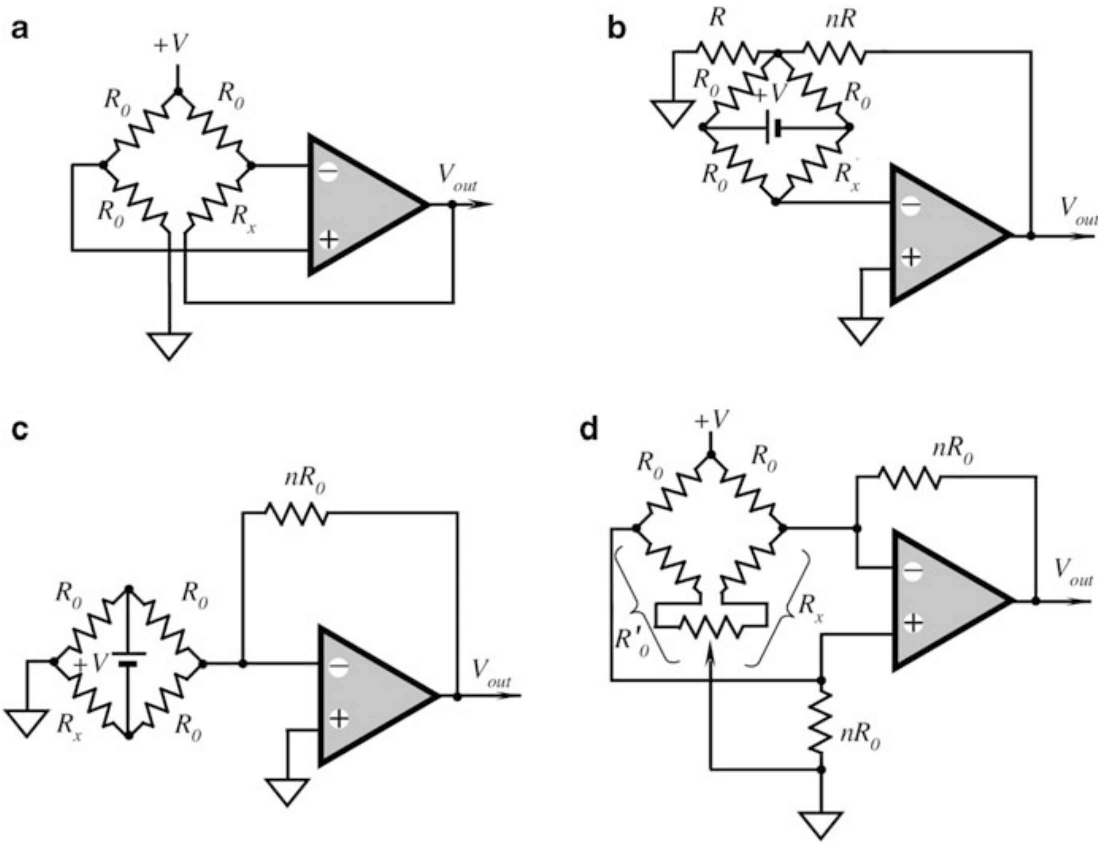


Fig. 6.19 Connection of operational amplifiers to resistive bridge circuits (deflection method)

amplifiers. Figure 6.19a shows the so-called active bridge, where a variable resistor (the sensor) is floating, i.e., isolated from ground, and is connected into a feedback of the OP-AMP. If a resistive sensor's transfer function can be approximated by a linear function:

$$R_x \approx R_o(1+\alpha), \tag{6.28}$$

where α is a small normalized input stimulus, then the output voltage of this circuit is:

$$V_{\text{out}} = \frac{V}{2}(1-\alpha). \tag{6.29}$$

A circuit with a floating bridge and floating reference voltage source V is shown in Fig. 6.19b. This circuit may provide gain that is determined by the feedback resistor whose value is nR_0, so the output voltage is:

$$V_{\text{out}} = V\left(\frac{1+\alpha}{2+\alpha} - \frac{1}{2}\right)(n+1). \tag{6.30}$$

Note that when floating, the sensor and reference voltage source shall have no connection whatsoever to ground neither directly nor through any other circuit.

All connecting wires to the amplifier must be very short, so the amplifier should be located close the sensor.

A bridge with the grounded sensor R_x but a floating reference voltage V is shown in Fig. 6.19c. The output voltage is:

$$V_{\text{out}} = V\left(\frac{1+\alpha}{2+\alpha} - \frac{1}{2}\right)n \tag{6.31}$$

Perhaps the most popular resistive bridge amplifier circuit is shown in Fig. 6.19d. It is for the grounded resistive sensor R_x with the amplifier having gain n. Its output voltage is:

$$V_{\text{out}} = Vn\left[\frac{1}{2n+1}\left(1+n\frac{2+\alpha}{1+\alpha}\right) - 1\right]. \tag{6.32}$$

Note that the circuit may contain a balancing potentiometer whose resistance sectors should be included into the corresponding arms of the bridge. The potentiometer is used to adjust the bridge component tolerances or offset the bridge balance by some fixed bias. When the bridge is perfectly balanced, its output voltage V_{out} is equal to zero. To better utilize the operational amplifier open-loop gain, the value of n should not exceed 100.

Note that circuits Fig. 6.19b–d are nonlinear with respect to α, even if the sensor is linear, while circuit Fig. 6.19a is linear. The reason for this is that in circuit of Fig. 6.19a, the sensor is supplied with a constant current being independent of the sensor's resistance, as it will be explained in the next section. In other circuits, the sensor's current varies when the sensor resistance changes, thus producing a nonlinear output, albeit for small α, this nonlinearity is usually very small.

6.3 Excitation Circuits

External signals are required for operation of the so-called *active* sensors. Examples are the absolute temperature sensors (thermistors and RTDs), pressure sensors (piezoresistive and capacitive), and displacement sensors (electromagnetic, capacitive and resistive). Different active sensors need different types of external signals for their operation. Depending on the sensor, these may be constant voltage, constant current, sinusoidal, or pulsing currents. It may even be light, magnetic field, or ionizing radiation. The names for that external signal is the *excitation signal* or *pilot signal*. In many cases, stability and precision of excitation signals directly relate to the sensor's accuracy and stability. Hence, it is imperative to generate the signal with such accuracy that the overall performance of the sensing system is not degraded. Below, we review several basic circuits, which feed sensors with the appropriate electric excitation signals.

When selecting the excitation circuit, one should think not only about the sensor but also what kind if signal processing is expected. Excitation signal is generally

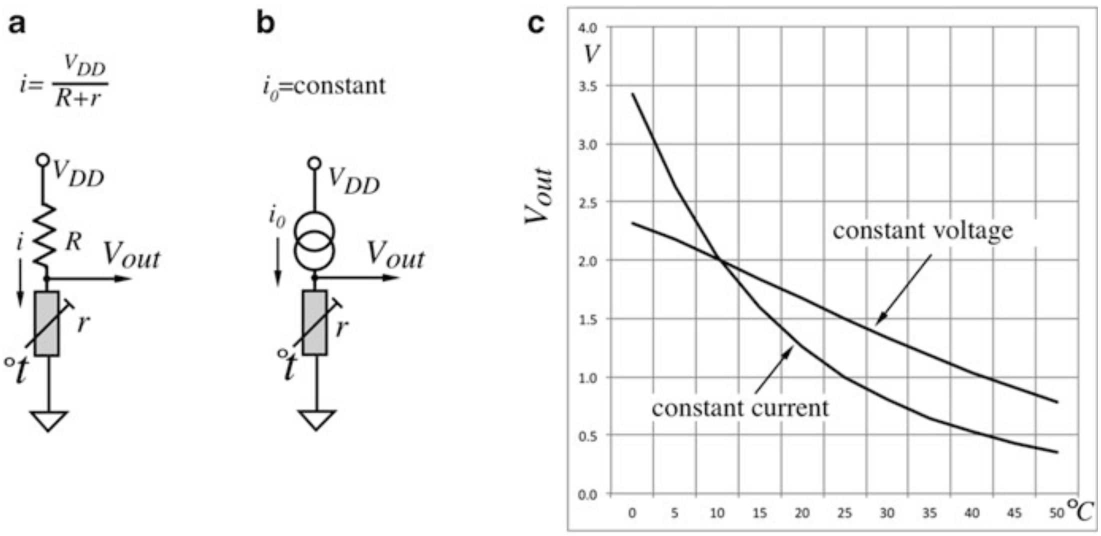


Fig. 6.20 Generating of excitation currents by a constant voltage source V_{DD} (**a**) and by constant current source i_0 (**b**); voltages across thermistor r at two different excitation currents as functions of temperature (**c**)

multiplicative to the sensor's transfer function and thus directly defines the shape and behavior of the output. To illustrate this, consider two possible circuits shown in Fig. 6.20a, b. These circuits are intended to pump electric current through a thermistor (temperature-dependent resistor) having resistance r. Each resistance r corresponds to a unique temperature. Thus, to measure temperature, one needs to measure resistance. However, to measure resistance it is necessary to force electric current i through that resistance. Then according to Ohm's law, the voltage V_{out} across the thermistor is function of current i, resistance r, and subsequently—temperature:

$$V_{\mathrm{out}} = ir \tag{6.33}$$

Figure 6.20a shows that current through r is also function of a pull-up fixed resistor R connected to a constant voltage source V_{DD}. Since r has a highly nonlinear transfer function with a negative temperature coefficient (NTC), resistance r and the current vary dramatically over a temperature range. Luckily, a combination of a variable resistor r and variable current i cause a linearization of the output voltage V_{out} in a narrow temperature range—see Fig. 6.20c, line "constant voltage". The linearization works only in a narrow temperature range, yet for many nondemanding applications it may be what is needed. On the other hand, circuit in Fig. 6.20b shows that instead of resistor R, a constant current generator is used. It produces fixed current i_0 that is independent of the thermistor resistance and power supply voltage V_{DD} and, as a result, according to Eq. (6.33), V_{out} is directly proportional to r. Figure 6.20c shows the line "constant current" that is highly nonlinear with temperature. Both circuits are useful, but which one to employ—depends on how V_{out} is going to be processed.

6.3.1 Current Generators

Current generators are often used as excitation circuits to feed sensors with predetermined currents that, within limits, are independent of the sensor properties, stimulus value, or environmental factors. In general terms, a current generator (current pump or current source or current sink) is a device that produces electric current independent of the load impedance. That is, within the capabilities of the generator, the amplitude of its output current remains substantially independent of any changes in the load and power supply voltage. It is said that an ideal current generator has infinitely high output resistance, so any series load will not change the output current. When generating a fixed current for a variable load, according to Ohm's Law, the corresponding voltage across the load changes in synch with the load.

Usefulness of the current generators for the sensor interfaces is in their ability to produce excitation currents of precisely controlled magnitude and shape. Hence, a current generator should not only produce current which is load-independent, but it also must be controllable from an external signal source (e.g., a wave-form generator), which in most cases has a voltage output.

There are two main characteristics of a current generator: the output resistance and voltage compliance. The output resistance should be as high as practical. A voltage compliance is the highest voltage which can be developed across the load without affecting the output current. For a high-resistive load, according to Ohm's law, Eq. (6.33), a higher voltage is required for a given current. For instance, if the required excitation current is $i = 10$ mA and the highest load impedance at any given frequency is $Z_L = 10$ kΩ, a voltage compliance of at least $iZ_L = 100$ V would be needed. Below, we show some useful circuits with the increased voltage compliance where the output currents can be controlled by external signals.

A unipolar current generator is called either a current *source* (generates the out-flowing current), or a current *sink* (generates the in-flowing currents). Here, unipolar means that it can produce currents of any shape flowing in one direction only, usually toward the ground. Many of such generators utilize current-to-voltage characteristics of transistors. A voltage controlled current source or sink may include an operational amplifier, Fig. 6.21a. In such a circuit, a precision and stable resistor R_1 defines the output current, i_L, that flows through the load impedance Z_L. The circuit contains a feedback loop through the OP-AMP that keeps voltage across resistor R_1 constant and thus assuring a constant current. To maximize the voltage compliance, a voltage drop across the sensing resistor R_1 should be as small as possible. For a better performance, current through the base of the output transistor should be as small as possible, hence, a field effect rather than bipolar transistor is used as an output current delivering device. Note that in that circuit the load is not grounded.

For many sensors, *bipolar* current generators may be required. Such a generator feeds a sensor with the excitation current that may flow in both directions (in- and out-flowing). In cases where the sensor must be grounded, a useful current pump is the circuit invented by Brad Howland at MIT [5]. One of its implementations is shown in Fig. 6.21b. The pump operation is based on utilizing both the negative and

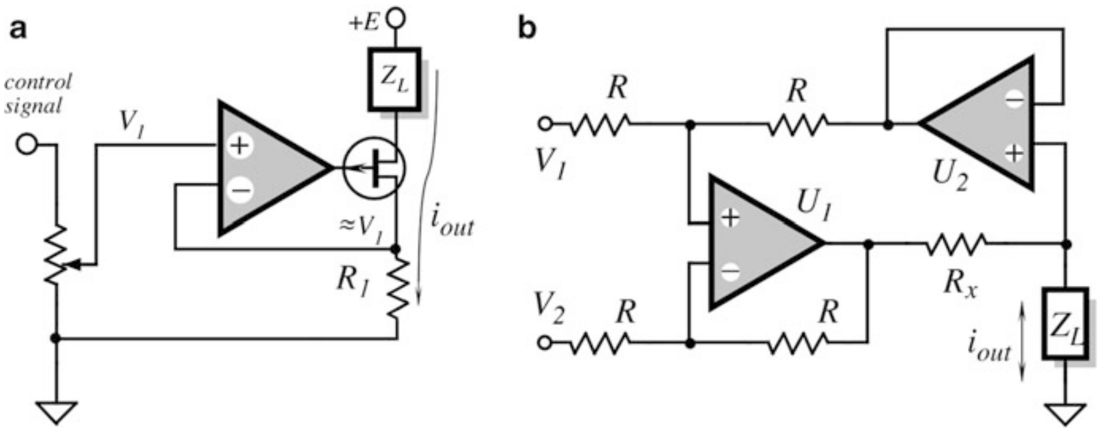


Fig. 6.21 Current sink with JFET transistor (**a**) and Howland current pump (**b**)

positive feedbacks around the operational amplifier. The load is connected to the positive loop. In that circuit all resistors should be nearly equal with high tolerances. Resistor R_x should be of a relatively low value for a sufficient output current magnitude. The circuit is stable for most of the resistive loads, however, to insure stability, a few picofarad capacitor may be added in a negative feedback or/and from the positive input of the operational amplifier U_1 to ground. When the load is inductive, an infinitely large compliance voltage would be required to deliver the set current when a fast transient control signal is applied. Therefore, the real current pump will produce a limited rising slope of the output current. The flowing current will generate an inductive spike across the output terminal, which may be fatal to the operational amplifier. It is advisable, for large inductive loads, to clamp the load with diodes to the power supply buses. The output current of the Howland pump is defined by the equation

$$i_{\mathrm{L}} = \frac{(V_1 - V_2)}{R_{\mathrm{s}}}. \tag{6.34}$$

The advantage of this circuit is that resistors R may be selected with a relatively high value and housed in the same thermally homogeneous packaging for better thermal tracking.

When the sensor is floating (not connected to ground), a simpler current source can be used. Figure 6.22 shows a noninverting (b) and inverting (a) circuits with operational amplifiers where the load (sensor) is connected as a feedback. Current through the load Z_{L} is equal to V_1/R_1 and is load independent. The load current follows variations in V_1 within the operating limits of the amplifier. An obvious limitation of the circuit is that the load is "floating", i.e., it is not connected to the ground bus or any other reference potential. For some applications, this is quite all right, however, many sensors need to be grounded or otherwise referenced. A circuit shown in Fig. 6.22b keeps one side of the load impedance near the ground potential, because a noninverting input of the OP-AMP is a virtual ground. Nevertheless, even in this circuit, the load is still fully isolated from the ground.

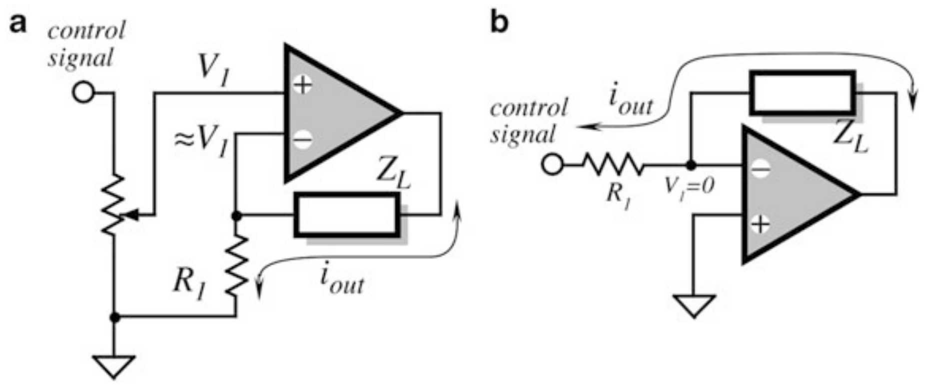


Fig. 6.22 Bidirectional current sources with floating loads (sensors)

6.3.2 Voltage Generators

As opposed to current generators, voltage generators (voltage sources or voltage drivers) must produce output voltages which over broad ranges of the loads and operating frequencies are independent of the load impedances and thus of the output currents. Sometimes, the voltage generators are called *hard voltage sources*. Usually, when a sensor that has to be driven by hard voltage is purely resistive, a driver can be a simple output stage which can deliver a sufficient current magnitude. However, when the load contains capacitances or inductances, that it, the load is reactive, the output stage of a voltage generator becomes a more complex device.

In many instances, even when the load is purely resistive, there still can be some capacitance associated with it. This may happen when the load is connected through lengthy wires or coaxial cables. A coaxial cable behaves as a capacitor connected from its central conductor to its shield if the length of the cable is less than 1/4 of the wavelength in the cable at the frequency of interest f. For a coaxial cable, this maximum length is given by

$$l_{\max} \leq 0.0165\frac{c}{f}, \tag{6.35}$$

where c is the velocity of light in a coaxial cable dielectric.

For instance, if $f = 100$ kHz, $l_{\max} \leq 0.0165\,\frac{3\times10^8}{10^5} = 49.5$, that is, a cable less than 49.5 m (162.4 ft) long will behave as a capacitor connected in parallel with the load, Fig. 6.23a. For example, cable R6-58A/U has the capacitance of 95 pF/m. This capacitance must be considered for two reasons: for the speed and stability of the circuits having a feedback coefficient β. The instability results from the phase shift produced by the output resistance of the voltage driver R_o and the loading capacitance C_L:

$$\varphi = \text{arctang}(2\pi f R_0 C_L). \tag{6.36}$$

For example, if $R_0 = 100\,\Omega$ and $C_L = 1000$ pF, at $f = 1$ MHz, the phase shift $\varphi \approx 32°$. This shift significantly reduces the phase margin in a feedback network that may cause a substantial degradation of the response and a reduced ability to

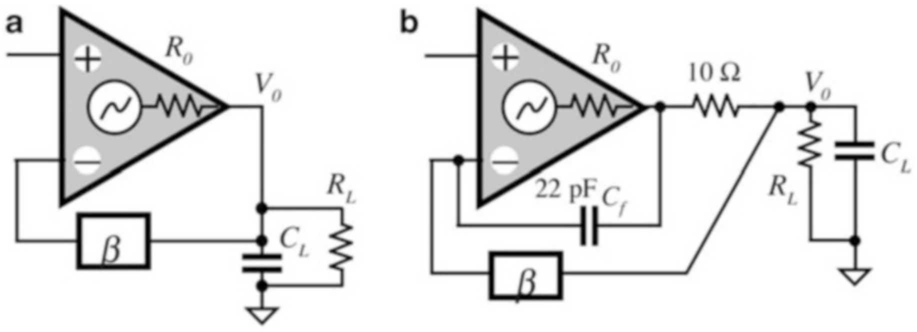


Fig. 6.23 Driving capacitive load. The load capacitor is coupled to driver's input through feedback (**a**); decoupling of capacitive load (**b**)

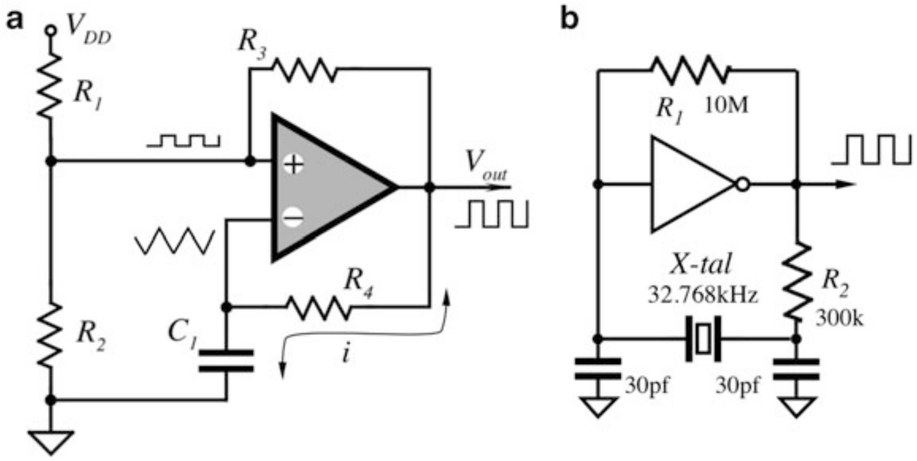


Fig. 6.24 Square-wave oscillator with OP-AMP (**a**) and crystal oscillator using a digital inverter (**b**)

drive the capacitive loads. The instability may be either overall, when an entire system oscillates, or localized when the driver alone becomes unstable. The local instabilities often can be cured by large by-pass capacitors (on the order of 10 μF) across the power supply or the so-called Q-spoilers consisting of a serial connection of 3–10 Ω resistor and a disc ceramic capacitor connected from the power supply pins of the driver chip to ground.

To make a driver stage more tolerant to capacitive loads, it can be isolated by a small serial resistor as it is shown in Fig. 6.24b. A small capacitive feedback (C_f) to the inverting input of the amplifier, and a 10 Ω resistor may allow to drive loads as large as 0.5 μF. However, in any particular case it is recommended to find the best values for the resistor and capacitor experimentally.

6.3.3 Voltage References

A voltage reference is an electronic device for generating precisely known constant voltage that is affected very little by variations in power supply, temperature, load, aging, and other factors. Many voltage references are available in a monolithic form for a large variety of the output voltages. Most of them operate with the so-called internal band-gap circuits. A good voltage reference should be a good voltage source, that is, it shall possess two critical features: have a very high stability of the output voltage and low-output resistance.

6.3.4 Oscillators

Oscillators are generators of variable electrical signals. Some of them are for generating a single signal wave (called “one-shot”), while others are free running. In many applications, free-standing oscillators may be replaced with digital outputs of a microprocessor or microcontroller, where the square-wave single or free-running pulses may be generated at one of the I/O ports.

Any oscillator essentially is comprised of a circuit with a gain stage, nonlinearity, and a certain amount of a positive feedback. By definition, an oscillator is an unstable circuit (as opposed to an amplifier that better be stable!) whose timing characteristics are either steady or changeable according to a predetermined functional dependence. The latter is called a *modulation*. Generally, there are three types of the electronic oscillators classified according to their time-keeping components: the RC, LC, and crystal oscillators (mechanical).

An RC-oscillators is called a *relaxation oscillator* because its functionality is based on a capacitor discharge (relaxation of a charge). The operating frequency is defined by a capacitor (C) and resistor (R)

An LC oscillator contains capacitive (C) and inductive (L) components that define the operating frequency.

In crystal oscillators, operating frequency is defined by a mechanical resonant in the specific cuts of piezoelectric crystals, usually quartz or ceramic. There is a great variety of the oscillation circuits, coverage of which is beyond the scope of this book. Just as an introduction, below we briefly describe a couple of practical circuits.

Many free-running oscillators (multivibrators) can be built with logic circuits, for instance—with NOR, NAND gates, or binary inverters. These circuits possess input nonlinearities, such as thresholds, that upon crossing, produce sharp transients at the outputs. Also, many multivibrators can be designed with comparators or operational amplifiers having a high open-loop gain. In all these oscillators, a combination of a capacitor and a resistor, or a crystal is a time-keeping combination.

In an RC-multivibrator, a voltage across a charging or discharging capacitor is compared with either constant or changing thresholds. When the capacitor charges, the moment of a threshold crossing is detected and causing generation of the output pulse transient. The transient is fed back to the RC-network (positive feedback) to cause the capacitor discharging that goes on until the next moment of comparison and generation of another pulse transient. Then, the cycle repeats.

This basic principle essentially requires the following minimum components: a capacitor, a charging circuit, and a threshold device (a comparator which is a nonlinear circuit). Several monolithic relaxation oscillators are available from many manufacturers, for instance a very popular timer, type 555, that can operate in either monostable (one-shot), or actable (free-running) modes. A great variety of oscillating circuits the reader can find in many books on operational amplifiers and digital systems, for instance [6].

A very popular free-running square-wave oscillator, Fig. 6.24a, can be built with one OP-AMP or a voltage comparator.[1] The amplifier is surrounded by two feedback loops—one is negative (to the inverting input) and the other is positive (to the noninverting input). A positive feedback (via R_3) controls the threshold level, while the negative loop charges and discharges timing capacitor C_1, through resistor R_4. Frequency of this oscillator can be determined from

$$f = \frac{1}{R_4 C_1}\left[\ln\left(1 + \frac{R_1 \| R_2}{R_3}\right)\right]^{-1}, \tag{6.37}$$

where $R_1 \| R_2$ is an equivalent resistance of parallel-connected R_1 and R_2.

A crystal oscillator is shown in Fig. 6.24b. It utilizes a digital voltage inverter that can be described as an inverting amplifier having very high gain in a very narrow linear range, that is near its threshold value (about 50 % of the power supply voltage). To bias the input close to that linear range, a feedback resistor R_1 is used as a negative feedback. The inverter amplifies the input signal so much that the output voltage is saturated either to ground or the positive power supply rail. It also flips its phase by 180°. The crystal inverts the output by another 180°, thus providing a positive feedback to the input, casing continuous oscillations.

6.4 Analog-to-Digital Converters

The conversion of an analog signal to a digital format involves quantization of the input, so it necessarily introduces a small amount of error. The converter periodically samples the analog signal and, at specific moments, performs conversions. The result is a sequence of digital values that have been converted from a continuous-time and continuously variable analog signal to a discrete-time and discrete-value digital signal.

[1] A voltage comparator differs from an operational amplifier by its faster transient response and special output circuit, which is easier interfaceable with digital circuits. A comparator may have a built-in hysteresis input circuit and it's called Schmitt trigger. Schmitt trigger is a digital comparator having two thresholds: upper and lower. When the input voltage transient goes upward and crosses the upper threshold, the trigger output switches high. When voltage moves downward and crosses the lower threshold, the output switches low. It was invented in 1934 by Otto H. Schmitt.

6.4.1 Basic Concepts

The analog-to-digital converters (abbreviated as A/D, or ADC, or A2D, or A-to-D) range from discrete circuits, to monolithic ICs (integrated circuits), to high-performance hybrid circuits, modules, and even boxes. Also, the converters are available as standard cells for custom and semicustom application-specific integrated circuits (ASIC). The ADCs transform continuous analog data—usually voltage—into an equivalent discrete digital format, compatible with digital data processing devices. Key characteristics of ADC include absolute and relative accuracy, linearity, no-missing codes, resolution, conversion speed, stability, and price. When price is of a major concern, a monolithic ADC as an embedded part of a microcontroller is the most efficient.

The most popular ADC converters are based on a successive-approximation technique because of an inherently good compromise between speed and accuracy. However, other popular techniques are used in a large variety of applications, especially when no high-conversion speed is required and only a small number of channels is needed. These include dual-ramp, quad-slope, pulse-width modulators (PWM), voltage-to-frequency (V/F) converters, and resistance-to-frequency (R/F) converters. The art of ADC is well developed. Here, we briefly review some popular architectures of ADCs, however, for detailed descriptions the reader should refer to specialized texts, such as [7].

The best-known digital code is *binary* (base 2). Binary codes are most familiar in representing integers, i.e., in a natural binary integer code having n bits, the LSB (least significant bit) has a weight of 2^0 (i.e., 1), the next bit has a weight of 2^1 (i.e., 2), and so on up to MSB (most significant bit), which has a weight of 2^{n-1} (i.e., $2^n/2$). The value of a binary number is obtained by adding up the weights of all nonzero bits. When the weighted bits are added up, they form a unique number having any value from 0 to $2^n - 1$. Each additional trailing zero-bit, if present, essentially doubles the size of the number.

When converting signals from analog sensors, because full scale is independent of the number of bits of resolution, a more useful coding is *fractional* binary which is always normalized to a full scale. Integer binary can be interpreted as fractional binary if all integer values are divided by 2^n. For example, the MSB has a weight of 1/2 (i.e., $2^{n-1}/2^n = 2^{-1}$), the next bit has a weight of 1/4 (i.e., 2^{-2}), and so forth down to the LSB, which has a weight of $1/2^n$ (i.e., 2^{-n}). When the weighted bits are added up, they form a number with any of 2^n values, from 0 to $(1 - 2^{-n})$ of full scale. Additional bits simply provide more fine structure without affecting the full-scale range. To illustrate these relationships, Table 6.2 lists 16 permutations of 5-bit's worth of 1's and 0's, with their binary weights, and the equivalent numbers expressed as both decimal and binary integers and fractions.

When all bits are "1" in natural binary, the fractional number value is $1 - 2^{-n}$, or normalized full-scale less 1 LSB ($1 - 1/16 = 15/16$ in the example). Strictly speaking, the number that is represented, written with an "integer point", is 0.1111 ($=1 - 0.0001$). However, it is almost a universal practice to write the

Table 6.2 Integer and fractional binary codes

Decimal fraction	Binary fraction	MSB ×1/2	Bit2 ×1/4	Bit3 ×1/6	Bit4 ×1/16	Binary integer	Decimal integer
0	0.0000	0	0	0	0	0000	0
1/16 (LSB)	0.0001	0	0	0	1	0001	1
2/16 = 1/8	0.0010	0	0	1	0	0010	2
3/16 = 1/8 + 1/16	0.0011	0	0	1	1	0011	3
4/16 = 1/4	0.0100	0	1	0	0	0100	4
5/16 = 1/4 + 1/16	0.0101	0	1	0	1	0101	5
6/16 = 1/4 + 1/8	0.0110	0	1	1	0	0110	6
7/16 = 1/4 + 1/8 + 1/16	0.0111	0	1	1	1	0111	7
8/16 = 1/2 (MSB)	0.1000	1	0	0	0	1000	8
9/16 = 1/2 + 1/16	0.1001	1	0	0	1	1001	9
10/16 = 1/2 + 1/8	0.1010	1	0	1	0	1010	10
11/16 = 1/2 + 1/8 + 1/16	0.1011	1	0	1	1	1011	11
12/16 = 1/2 + 1/4	0.1100	1	1	0	0	1100	12
13/16 = 1/2 + 1/4 + 1/16	0.1101	1	1	0	1	1101	13
14/16 = 1/2 + 1/4 + 1/8	0.1110	1	1	1	0	1110	14
15/16 = 1/2 + 1/4 + 1/8 + 1/16	0.1111	1	1	1	1	1111	15

code simply as the integer 1111 (i.e., "15") with the fractional nature of the corresponding number understood: "1111" → 1111/(1111 + 1), or 15/16.

For convenience, Table 6.3 lists bit weights in binary for numbers having up to 20 bits. However, the practical range for the vast majority of sensors rarely exceeds 16 bits.

The weight assigned to the LSB is the resolution of numbers having n bits. The dB column represents the logarithm (base 10) of the ratio of the LSB value to unity (full scale), multiplied by 20. Each successive power of 2 represents a change of 6.02 dB [i.e., $20 \log_{10}(2)$] or "6 dB/octave".

6.4.2 V/F Converters

A voltage-to-frequency (V/F), as the name implies, converts voltage to a variable frequency of pulses, in other words—input voltage modulates frequency. This is called *frequency modulation* or FM. V/F can provide a high-resolution conversion, that is also is useful for some additional sensor features, such as a voltage isolation, communication, and storage of data. The converter accepts analog output from a sensor, which can be either voltage or current (in latter case, of course, it should be called a current-to-voltage converter). Here we will discuss a conversion of voltage to frequency, or, in other words, to a *number of square pulses per unit of time*. A frequency is a digital format because pulses can be gated (selected for a given interval of time) and then counted, resulting in a binary

Table 6.3 Binary bit weights and resolutions

BIT	2^{-n}	$1/2^n$ fraction	dB	$1/2^n$ decimal	%	ppm
FS	2^0	1	0	1.0	100	1,000,000
MSB	2^{-1}	1/2	−6	0.5	50	500,000
2	2^{-2}	1/4	−12	0.25	25	250,000
3	2^{-3}	1/8	−18.1	0.125	12.5	125,000
4	2^{-4}	1/16	−24.1	0.0625	6.2	62,500
5	2^{-5}	1/32	−30.1	0.03125	3.1	31,250
6	2^{-6}	1/64	−36.1	0.015625	1.6	15,625
7	2^{-7}	1/128	−42.1	0.007812	0.8	7812
8	2^{-8}	1/256	−48.2	0.003906	0.4	3906
9	2^{-9}	1/512	−54.2	0.001953	0.2	1953
10	2^{-10}	1/1024	−60.2	0.0009766	0.1	977
11	2^{-11}	1/2048	−66.2	0.00048828	0.05	488
12	2^{-12}	1/4096	−72.2	0.00024414	0.024	244
13	2^{-13}	1/8192	−78.3	0.00012207	0.012	122
14	2^{-14}	1/16,384	−84.3	0.000061035	0.006	61
15	2^{-15}	1/32,768	−90.3	0.0000305176	0.003	31
16	2^{-16}	1/65,536	−96.3	0.0000152588	0.0015	15
17	2^{-17}	1/131,072	−102.3	0.00000762939	0.0008	7.6
18	2^{-18}	1/262,144	−108.4	0.000003814697	0.0004	3.8
19	2^{-19}	1/524,288	−114.4	0.000001907349	0.0002	1.9
20	2^{-20}	1/1,048,576	−120.4	0.0000009536743	0.0001	0.95

number. All V/F converters are of the *integrating* type because the number of pulses per second, or *frequency*, is proportional to the *average* value of the input voltage.

By using a V/F converter, a conversion to a digital format can be performed in the most simple and economical manner. The time required to convert an analog voltage into a digital number relates to a full-scale frequency of the V/F converter and the required resolution. Generally, the V/F converters are relatively slow, as compared with successive-approximation devices, however, they are quite appropriate for many sensor applications. When acting as an ADC, the V/F converter is coupled to a counter which is clocked with the required sampling rate. For instance, if a full-scale frequency of the converter is 32 kHz, and the counter is clocked eight times per second, the highest number of pulses which can be accumulated every counting cycle is 4000 which approximately corresponds to a resolution of 12 bit (see Table 6.3). By using the V/F converter and counter, an integrator can be build for the applications, where the stimulus needs to be integrated over a certain time. The counter accumulates pulses over the gated interval rather than as an average number of pulses per counting cycle.

Another useful feature of a V/F converter is that output pulses can be easily transmitted through communication lines. The pulsed signal is much less susceptible to noisy environment than a high-resolution analog signal. In the ideal

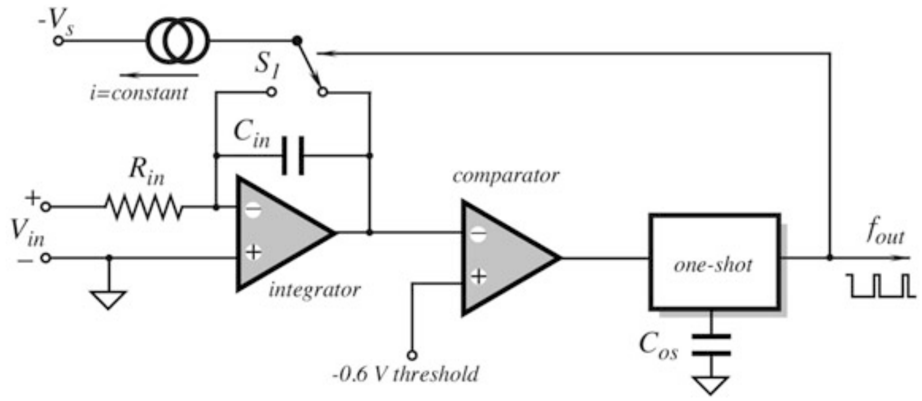


Fig. 6.25 Charge-balance V/F converter

case, the output frequency f_{out} of the converter is proportional to the input voltage V_{in}:

$$\frac{f_{\mathrm{out}}}{f_{\mathrm{FS}}} = \frac{V_{\mathrm{in}}}{V_{\mathrm{FS}}}, \tag{6.38}$$

where f_{FS} and V_{FS} are the full-scale frequency and input voltage, respectively. For a given linear converter, ratio $f_{\mathrm{FS}}/V_{\mathrm{FS}} = G$ is constant and is called a conversion factor:

$$f_{\mathrm{out}} = GV_{\mathrm{in}}. \tag{6.39}$$

There are several known types of V/F converters. The most popular of them are the multivibrator and the charge-balance circuit.

A *multivibrator* V/F converter employs a free-running square-wave oscillator where charge-discharge currents of a timing capacitor are controlled by the input signal. However, a more accurate type is the *charge-balance* type of converter that employs an analog integrator and a voltage comparator as shown in Fig. 6.25. This circuit has such advantages as high speed, high linearity, and good noise rejection. The circuit is available in an integral form from several manufacturers, for instance, AD652 from Analog Devices and LM331 from Texas Instruments.

The converter operates as follows. Input voltage V_{in} is applied to an integrator through the input resistor R_{in}. The integrating capacitor is connected as a negative feedback loop to the operational amplifier whose output voltage is compared with a small negative threshold of -0.6 V. The integrator generates a saw-tooth voltage (Fig. 6.26) that at the moment of comparison with the threshold results in a transient at the comparator's output. That transient enables a one-shot generator to produce a square pulse of a fixed duration t_{os}. A precision current source generates constant current i which is alternatively applied either to the summing node of the integrator, or to its output. The switch S_1 is controlled by the one-shot pulses. When the current source is connected to the summing node, it delivers a precisely defined packet of charge $\Delta Q = it_{\mathrm{os}}$ to the integrating capacitor. The same summing node also receives

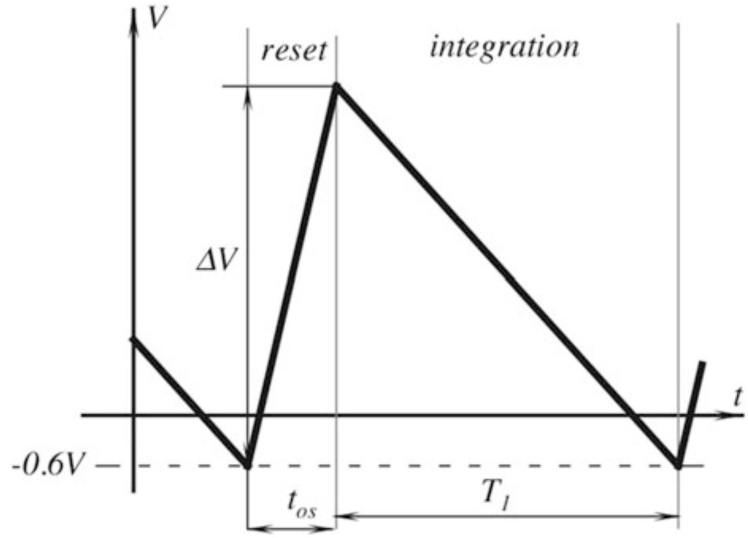


Fig. 6.26 Integrator output in charge-balanced V/F converter

an input charge through the resistor R_{in}, thus the net charge is accumulated on the integrating capacitor C_{in}.

When the threshold is reached, the one-shot generator is triggered and the switch S_1 changes its state to high, thus initiating a reset period. During the reset period, the current source delivers its current to the summing node of the integrator. The input current charges the integrating capacitor upward. The total voltage between the threshold value and the end of the deintegration is determined by the duration of a one-shot pulse:

$$\Delta V = t_{os}\frac{\mathrm{d}V}{\mathrm{d}t} = t_{os}\frac{i - I_{in}}{C_{in}}. \tag{6.40}$$

When the output signal of the one-shot circuit goes low, switch S_1 diverts current i to the output terminal of an integrator, which makes no effect on the state of the integrating capacitor C_{in}. That is, the current source sinks a portion of the output current from the operational amplifier. This time is called the integration period. During the integration, the positive input voltage delivers current $I_{in} = V_{in}/R_{in}$ to the capacitor C_{in}. This causes the integrator to ramp down from its positive voltage with the rate proportional to V_{in}. The amount of time required to reach the comparator's threshold is:

$$T_1 = \frac{\Delta V}{\mathrm{d}V/\mathrm{d}t} = t_{os}\frac{i - I_{in}}{C_{in}}\frac{1}{I_{in}/C_{in}} = t_{os}\frac{i - I_{in}}{I_{in}}. \tag{6.41}$$

It is seen that the capacitor value does not affect duration of the integration period.

The output frequency is determined by:

$$f_{out} = \frac{1}{t_{os} + T_1} = \frac{I_{in}}{t_{os}i} = \frac{V_{in}}{R_{in}}\frac{1}{t_{os}i}. \tag{6.42}$$

Therefore, the frequency of one-shot pulses is proportional to the input voltage. It depends also on quality of the integrating resistor, stability of the current generator, and a one-shot circuit. With a careful design, this type of a V/F converter may reach nonlinearity error below 100 ppm and can generate frequencies from 1 Hz to 1 MHz.

A major advantage of the integrating-type converters, such as a charge-balanced V/ F converter, is the ability to reject large amounts of additive noise. By integrating of the measurement, noise is reduced or even totally eliminated. Pulses from the converter are accumulated for a gated period T in a counter. Then, the counter behaves like a filter having a transfer function in the form

$$H(f) = \frac{\sin \pi f T}{\pi f T}, \tag{6.43}$$

where f is the frequency of pulses. For low frequencies, value of this transfer function $H(f)$ is close to unity, meaning that the converter and the counter make correct measurements. However, for a frequency $1/T$ the transfer function $H(1/T)$ is zero, meaning that these frequencies are completely rejected. For example, if gating time $T = 16.67$ ms which corresponds to a frequency of 60 Hz (the power line frequency which is a source of substantial noise in many sensors) the 60 Hz noise will be rejected. Moreover, the multiple frequencies (120 Hz, 180 Hz, 240 Hz, and so on) will also be rejected.

6.4.3 PWM Converters

The *pulse-width modulation* (PWM) in many respects is similar to FM. The main difference is that in PWM, period T_0 of the square pulses remains constant (therefore frequency of pulses is also constant), while the pulse duration t_{PWM} is proportional to the input voltage. In other words, a duty cycle D is proportional to voltage:

$$D = \frac{t_{\text{PWM}}}{T_0} = kV_{\text{in}}, \tag{6.44}$$

where k is the conversion constant. Theoretically, duty cycle varies from 0 to 1, however, practically, it has a bit narrower range, typically from 0.05 to 0.95, thus utilizing the period T_0 with about 0.9 or 90 % efficiency.

To make conversion from PWM signal to a binary code, a PWM pulse can be used as a gating function for a high-frequency pulse train and the subsequent counting of the gated pulses. For example, if period T_0 is 10 ms (the PWM conversion is with frequency $F_0 = 1/T_0 = 100$ Hz) and the pulse train is of 1 MHz (10^6 Hz), then with a PWM efficiency 0.9, each PWM pulse can gate maximum $10^6/10^2 \times 0.9 = 9000$ high-frequency pulses. This is approximately equivalent to a 13-bit resolution (see Table 6.3).

A PWM modulator can be implemented with a saw-tooth generator as shown in Fig. 6.27. The reset pulses (voltage V_1) of a fixed period T_0 are generated by the microcontroller at its IO_1. Each V_1 pulse starts generation by the saw-tooth generator of a positive ramp V_{saw}. Input voltage V_{in} and V_{saw} are fed to the

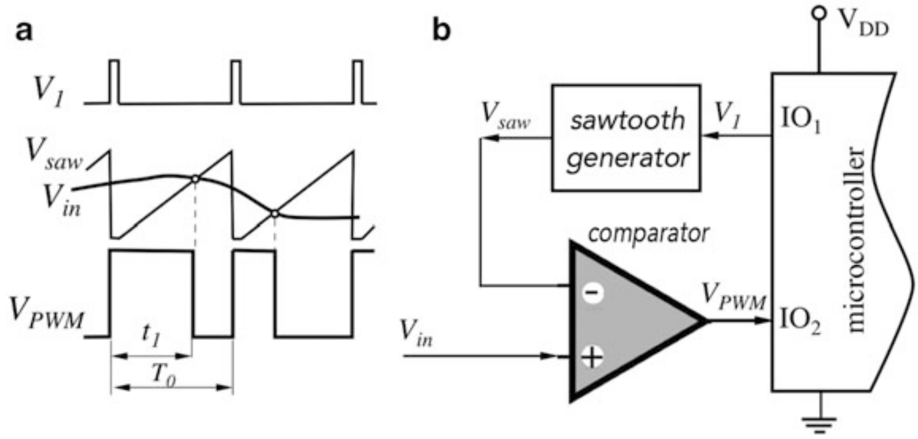


Fig. 6.27 Voltages (**a**) and block diagram of PWM converter with a microcontroller (**b**)

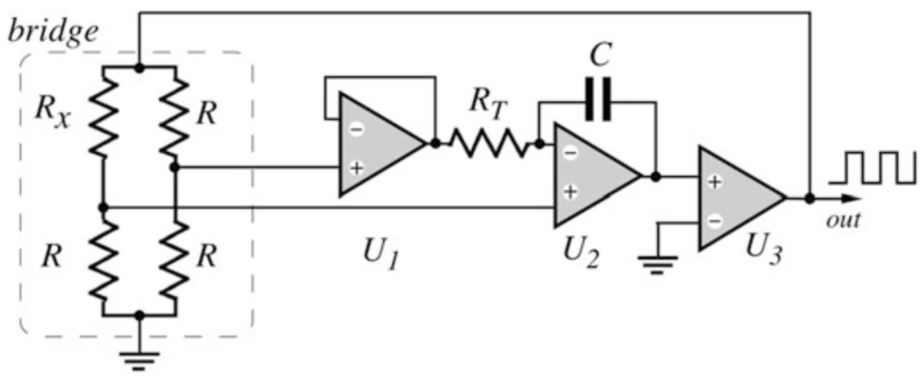


Fig. 6.28 Schematic of R/F converter for Wheatstone bridge

respective noninverting and inverting inputs of the analog comparator, whose output is a PWM pulse V_{PWM}. The next reset pulse clears and restarts the saw-tooth generator and a new cycle starts. The PWM pulses may be provided to the IO_2 of the microcontroller whose firmware may control a further conversion of PWM to a binary code.

6.4.4 R/F Converters

For a resistive sensor, a conversion to a digital format can be performed without an intermediate conversion of resistance to voltage. In a direct conversion, the sensor is used as a component in a pulse modulator, usually as a frequency modulator of an oscillator.

As a first example of this approach we discuss an R/F converter for a resistive Whetstone bridge, where resistance of the sensor $R_x = R + \Delta R$. Figure 6.28 illustrates a simplified schematic [8] of a bridge being part of a free-running relaxation-type oscillator. The oscillator comprises a timing capacitor C and timing resistor R_{T} that define the base frequency f_0 when the bridge is perfectly balanced.

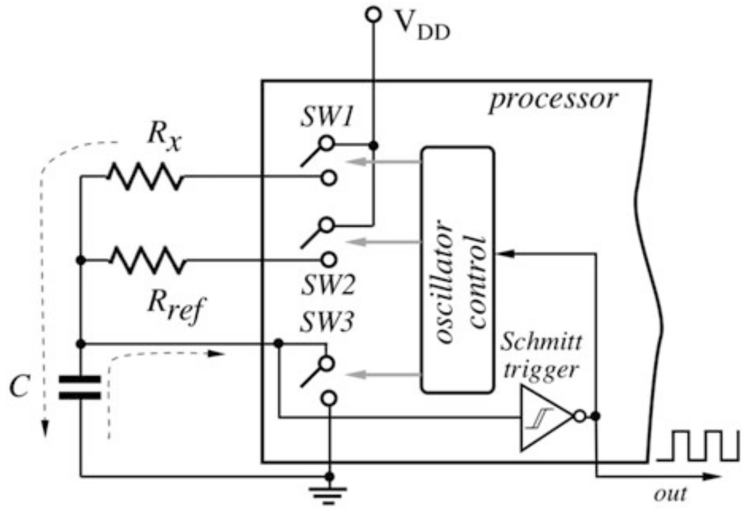


Fig. 6.29 Ratiometric resistance-to-frequency converter

The circuit also contains a voltage follower U_1, integrator U_2, and comparator U_3. A positive feedback from U_3 to the resistive bridge results in a continuous generation of square pulses whose frequency deviation Δf is a linear function of the bridge disbalance ΔR:

$$\Delta f = \frac{\Delta R}{2R} f_0 \tag{6.45}$$

Another R/F approach utilizes a microprocessor as illustrated in Fig. 6.29. A resistive sensor R_x, for example—a thermistor or resistive humidity sensor, and a reference resistor R_{ref} are connected to a capacitor C. Under command of the oscillator control circuit, the capacitor can be charged from the power supply voltage V_{DD} via one of these resistors and discharged to ground through the solid-state switch SW3. Initially, the switch SW2 stays open, while the sensing resistor R_x is connected to the capacitor via the charging SW1 that is alternatively turned on and off out of phase with the discharging SW3. The capacitor C develops a saw-tooth voltage that is fed to a Schmitt trigger producing pulses whose frequency f_x is function of the sensor resistor R_x and capacitor C. The processor for some fixed time accumulates and counts these pulses to measure their frequency related to the stimulus.

The next phase is generating the reference pulses by the same circuit, except that the switch SW1 stays open and SW2 alternates with SW3 to use R_{ref} for charging C. The new frequency is f_{ref} and again for the same fixed time the processor accumulates and counts these reference pulses. After both frequencies are measured, the output digital number representing the stimulus that modulates R_x is computed as a ratio:

$$x = \frac{f_{\mathrm{ref}}}{f_x} = \frac{R_x}{R_{\mathrm{ref}}} \tag{6.46}$$

Thanks to the ratiometric technique, output (Eq. 6.46) depends only on resistors R_{ref} and R_x, while all other factors, such as capacitance C, power supply voltage, thermal effects, circuit characteristics, and other interfering factors are cancelled out. This method was utilized in the integrated circuit S1C6F666 from Epson.

6.4.5 Successive-Approximation Converter

These converters are widely used in a monolithic form thanks to their high speed (up to 1 MHz sampling rates) and high resolution (16 bit and higher). Conversion time is fixed and independent of the input signal. Each conversion is unique, as the internal logic and registers are cleared after each conversion, thus making these ADC converters suitable for a multichannel multiplexing. The converter (Fig. 6.30) consists of a precision voltage comparator, a module comprising shift registers and a control logic, and a digital-to-analog converter (DAC) that serves as a feedback from the digital outputs to the input analog comparator.

The conversion technique consists of comparing the unknown input, V_{in}, against a precise voltage, V_a, or current generated by the DAC. The conversion technique is similar to a weighing process using a balance, with a set of n binary weights (for instance, 1/2 kg, 1/4 kg, 1/8 kg, 1/16 kg, etc. up to total of 1 kg). Before the conversion cycles, all the registers must be cleared and the comparator's output is HIGH. The DAC has MSB (1/2 scale) at its inputs and generates an appropriate analog voltage, V_a, equal to 1/2 of a full-scale input signal. If the input is still greater than the DAC voltage (Fig. 6.31), the comparator remains HIGH, causing "1" at the register's output. Then, the next bit ($2/8 = 1/4$ of FS) is tried. If the second bit does not add enough weight to exceed the input, the comparator remains HIGH ("1" at the output), and the third bit is tried. However, if the second bit tips the scale too far, the comparator goes LOW resulting in "0" in the register, and the third bit is tried. The process continues in order of descending bit weight until the last bit has been tried. After the completion, the status line indicates the end of conversion and data can be read from the register as a valid number corresponding to the input signal.

To make the conversion valid, the input signal V_{in} must not change until all the bits are tried, otherwise, the digital reading may be erroneous. To avoid any problems with the changing input, a successive-approximation converter usually is supplied with a sample-and-hold (S&H) circuit. This circuit is a short-time analog memory that samples the input signal and stores it as a dc voltage during an entire conversion cycle.

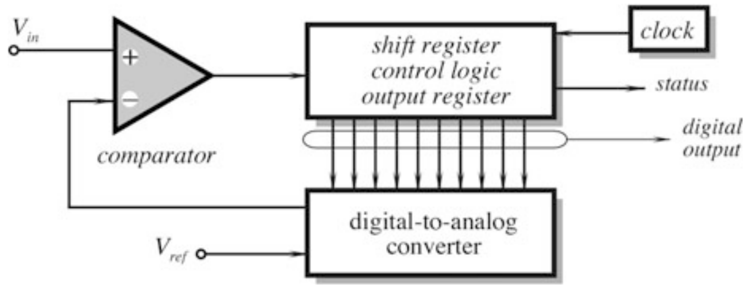


Fig. 6.30 Block Diagram of successive-approximation DAC

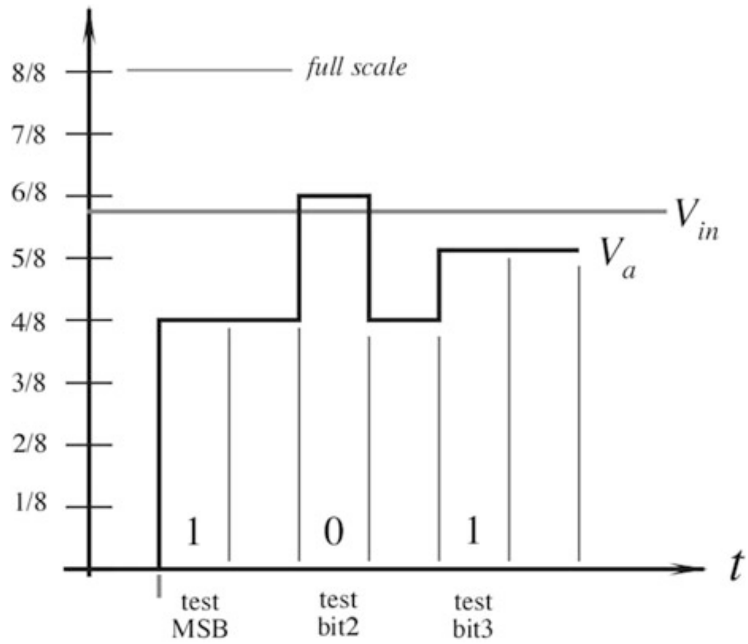


Fig. 6.31 3-Bit weighing in successive-approximation ADC

6.4.6 Resolution Extension

In a typical data acquisition system, many low-cost monolithic microcontrollers contain analog-to-digital converters, whose maximum resolutions are limited to 10 or 12 bits. When the resolution of a built-in converter is higher, or an external converter of a high resolution is used, the cost may become prohibitively high. In many applications, 12 bits may be not sufficient for a correct representation of a minimum increment of a stimulus (the required input resolution R_o). There are several ways of resolving this problem. One is to use an analog amplifier in front of the ADC. For example, an amplifier of gain 4 will effectively increase the input resolution R_o by two bits, say from 12 to 14. Of course, the price to pay is an uncertainty in the amplifier's characteristics. Another method of achieving higher resolution is to use a dual-slope ADC converter whose resolution limited only by the available counter rate and the speed response of a comparator.[2] And another method is to use a 12-bit ADC converter (for instance, of a successive-approximation type) with a resolution extension circuit. Such a circuit can boost the resolution by several bits, for instance from 12 to 15. A block diagram of the resolution extension circuit is shown in Fig. 6.32a. In addition to a conventional 12-bit ADC converter, it includes a digital-to-analog converter (DAC), a subtraction circuit, and an amplifier having gain A. In the ASIC or discrete circuits, a DAC may be shared with the ADC part (see Fig. 6.30).

[2] A resolution should not be confused with accuracy.

The input signal V_m has a full-scale value E, thus for instance, if we have a 12-bit ADC, the initial resolution will be:

$$R_o = E/\left(2^{12}-1\right) = E/4095, \tag{6.47}$$

which is expressed in volts per bit. For instance, for $E = 5$ V, the 12-bit resolution is 1.22 mV/bit. Initially, the multiplexer (MUX) connects the input signal to the ADC which produces the output digital value, M, which is expressed in bits. Then, the microprocessor outputs that value to the DAC that produces the output analog voltage V_c, which is an approximation of the input signal. This voltage is subtracted from the input signal and the difference is amplified by the amplifier to value

$$V_D = (V_m - V_c) \cdot A \tag{6.48}$$

The voltage V_D is an amplified error between the actual and digitally represented input signals. For a full-scale input signal, the maximum error $(V_m - V_c)$ is equal to a resolution of an ADC converter, therefore, for an 12-bit conversion $V_D = 1.22 \cdot A$ mV. The multiplexer connects that voltage to the ADC converter which converts V_D to a digital value C:

$$C = \frac{V_D}{R_0} = (V_m - V_c)\frac{A}{R_0}. \tag{6.49}$$

As a result, the microprocessor combines two digital values: M and C, where C represents the high-resolution bits. If $A = 255$, then for the 5-V full-scale, LSB $\approx$ 4.8 μV which corresponds to a total resolution of 20 bit. In practice, it is hard to achieve such a high resolution because of the errors originated in the DAC, reference voltage, amplifier's drift, noise, etc. Nevertheless, the method is quite efficient when a modest resolution of 14 or 15 bit is deemed to be sufficient.

Another powerful method of a resolution extension is based on the so-called *oversampling* [9]. The idea works only if the input analog signal is changing between the sampling points. For example, if the ADC conversion steps are at 50 mV, 70 mV, 90 mV, etc., while the input signal is steady 62 mV, the digital number will indicate 70 mV, thus producing a digitization error of 8 mV and no oversampling would make any difference. However, if the input signal changes with the maximum spectral frequency f_m, according to the Nyquist-Shannon-Kotelnikov theorem,[3] the sampling frequency $f_s > 2f_m$. The oversampling requires a much higher sampling frequency than defined by the Nyquist. Specifically, it is based on the formula

[3] A fundamental theorem of the information theory. It states that the minimum sampling must be twice as fast as the highest frequency of the signal.

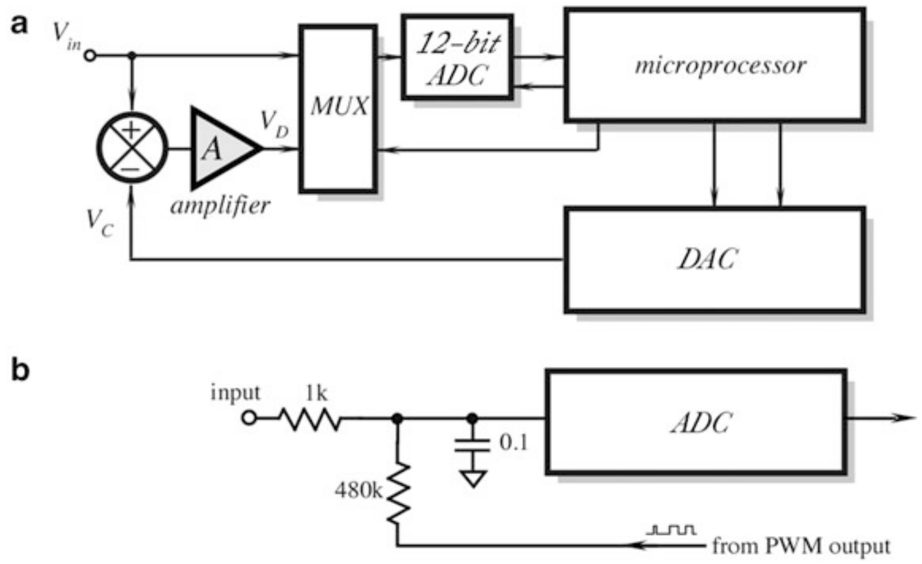


Fig. 6.32 Resolution enhancement circuit with DAC (**a**); adding artificial noise to input signal for oversampling (**b**)

$$f_{os} > 2^{2+n} f_m, \tag{6.50}$$

where n is a number of the extension bits. For example, if we have a 10-bit ADC and would like to generate with it a number of 12-bits ($n = 2$), the sampling rate must be at least 16 times higher than f_m. The oversampling allows trading a resolution of an ADC conversion for the maximum converted frequency. Thus, this method is useful for converting relatively slow changing signals as compared with the maximum sampling rate of an ADC.

As it was said above, the method requires the signal to change between samplings. If the analog signal does not include natural variations or inherent noise, an artificial noise can be added to the input signal or the ADC reference voltage to jitter signal between the samples. A practical method of adding artificial noise is shown in Fig. 6.32b. The microcontroller generates PWM (Pulse-Width Modulated) random-width pulses that are smoothed by a capacitor and added to the analog input signal. The magnitude of jittering must correspond to at least 0.5 LSB of the original resolution but preferably should be about 2 LSB. After sampling, to get an increased resolution, 2^{2+n} samples from the ADC are added and the result is right-shifted n times. For the above example, 16 sequential 10-bit numbers are added and then right-shifted two times, resulting in a 12-bit output number.

6.4.7 ADC Interface

When a sensor or sensing circuit, such as the Wheatstone bridge, is connected to an ADC, it is important to assure that the coupling between the two does not introduce an unexpected error. As we indicated above, a sensor can be powered either by a voltage source (Sect. 6.3.2) or by a current source (Sect. 6.3.1). Further, a voltage

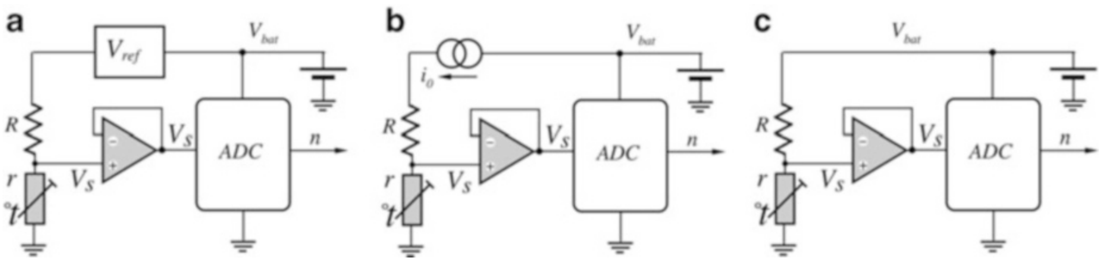


Fig. 6.33 Powering the sensor circuit with constant voltage (**a**), constant current (**b**), and ratiometric voltage (**c**)

source can be connected by two options: constant voltage and ratiometric. On the other hand, ADC also requires a voltage reference, that may be either internal or external. Thus, to avoid a mismatch between the sensor and ADC, a coupling between the sensor and ADC should consider how the sensor and ADC are referenced? Figure 6.33 illustrates three possibilities for powering (referencing) a resistive sensor—in this example a thermistor r with a pull-up resistor R.

A power supply that feeds the sensor may be not very stabile, have some ripples, noise, and even may drift substantially during operation. An example is a battery producing voltage V_{bat} that drops during discharge. The sensor that requires an excitation power may be fed by a constant voltage as shown in Fig. 6.33a. A constant voltage is produced by a precision voltage reference V_{ref} whose value is independent of the battery voltage V_{bat} (see Sect. 6.3.3). Thus, the sensor's voltage divider outputs V_{r} that depends only on the sensor, pull-up resistor, and reference voltage. This method is used when ADC has its own precision voltage reference, thus conversion to a digital format is not affected by the battery.

Some sensor applications require a constant current i_0 excitation (Fig. 6.33b). In that case, the divider output V_{r} depends on two factors: the constant current and sensor resistance r. As in the previous case, to be independent of the battery, the ADC must have its own voltage reference.

An efficient way of powering a sensor is to use a ratiometric technique, where both the sensor and ADC are powered from the same voltage source (battery) that does not have to be regulated (Fig. 6.33c). It is important that ADC uses V_{bat} as its reference. Thus, the output ADC counts are proportional to the battery voltage. Here is how it works.

Voltage from the sensor at the ADC input depends on the battery:

$$V_{\text{r}} = \frac{r}{r+R} V_{\text{bat}}, \tag{6.51}$$

while ADC transfer function is also battery dependent. Its output digital count n is:

$$n = \frac{V_{\text{r}}}{V_{\text{bat}}} N_{\text{FS}}, \tag{6.52}$$

where N_{FS} is the maximum ADC count or MSB (most significant bit) corresponding to the maximum (full-scale) input voltage. Substituting Eq. (6.51) into Eq. (6.52) we arrive at the output count as function of the sensor resistance:

$$n = \frac{r}{r + R} N_{FS} \tag{6.53}$$

We see that the digital output is independent of the power supply voltage and relates only to the sensor's and pull-up resistances.

6.5 Integrated Interfaces

In the past, an application engineer had to design her own interface circuit and often that was a challenge. The modern trend in signal conditioning is to integrate in a single chip the amplifiers, multiplexers, ADC, memory, and other circuits (see Sect. 3.1.2). This frees the application engineers from designing the interface and signal-conditioning circuits—few engineers might have an experience in designing such systems. Thus, standardized interfaces are the reliable and efficient solutions. Below are two examples of many highly useful commercially available integrations.

6.5.1 Voltage Processor

Figure 6.34 illustrates an integrated signal conditioning circuit on a single chip ZSC31050 from the German company ZMDI (www.zmdi.com). It is optimized for several low-voltage and low-power multiple sensors, including a resistive bridge. The differential voltage from the bridge sensor is preamplified by the programmable gain amplifier and, along with several other sensor signals, fed to the multiplexer. The multiplexer transmits the signals, including one from the internal temperature sensor, to the ADC in a specific sequence. Next, the ADC converts

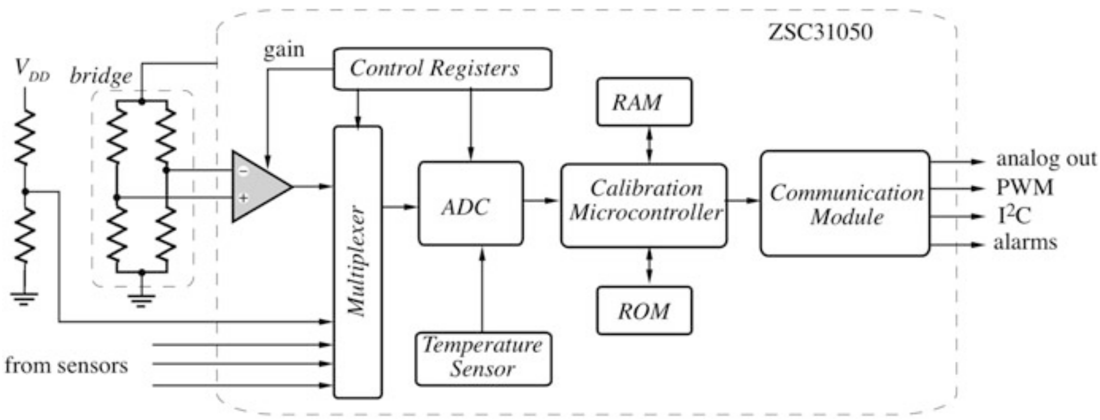


Fig. 6.34 Integrated ZMDI Signal Conditioner

these signals into 15-bit digital values. The digital signal correction takes place in the calibration microcontroller. The correction is based on sensor-special formulas residing in the ROM and the sensor-specific calibrating coefficients (stored in the EEPROM during calibration). Depending on the programmed output configuration, the corrected sensor signal goes to the output through the communication module as analog voltage, PWM signal, or in various communication formats, includes a serial link I^2C. The configuration data and the correction parameters can be programmed into the EEPROM via the digital interfaces. The modular circuit concept used in the design of the ZSC31050 allows fast customization of the IC for high-volume applications.

6.5.2 Inductance Processor

Magnetic sensors operating on the inductive principle are highly popular for detecting proximity, presence of magnetic and conductive objects, measuring electrical impedance, etc. Many of these sensors are described throughout this book. The chip LDC1000 from TI is an integrated inductance-to-digital converter (LDC) that monitors a combined impedance of an external resonant tank consisting of an inductive coil (L) and capacitor (C). Loss of power in the tank is function of its inductive coupling with the outside. By monitoring the loss through measuring the LC-resonator's parallel loss resistivity it is possible to monitor the external conductive objects with high degree of accuracy.

Consider Fig. 6.35 where inductance L of the sensing coil causes the circular eddy currents in a conductive material positioned at some distance d from the coil. An a.c. current flowing through a coil generates an a.c. magnetic field. If a conductive material, such as a target made of metal, is brought into the vicinity of the coil, the alternate magnetic field will induce circulating eddy currents inside the target. These eddy currents are a function of the distance, size, and

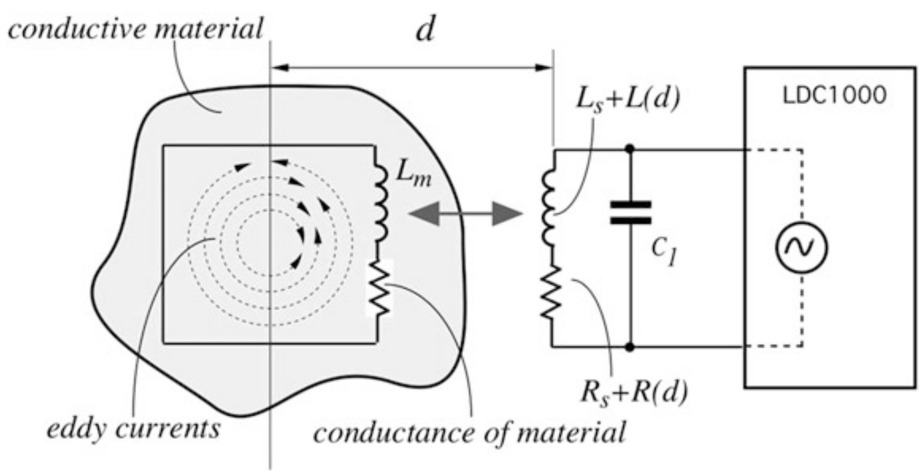


Fig. 6.35 Generation of eddy currents in conductive material

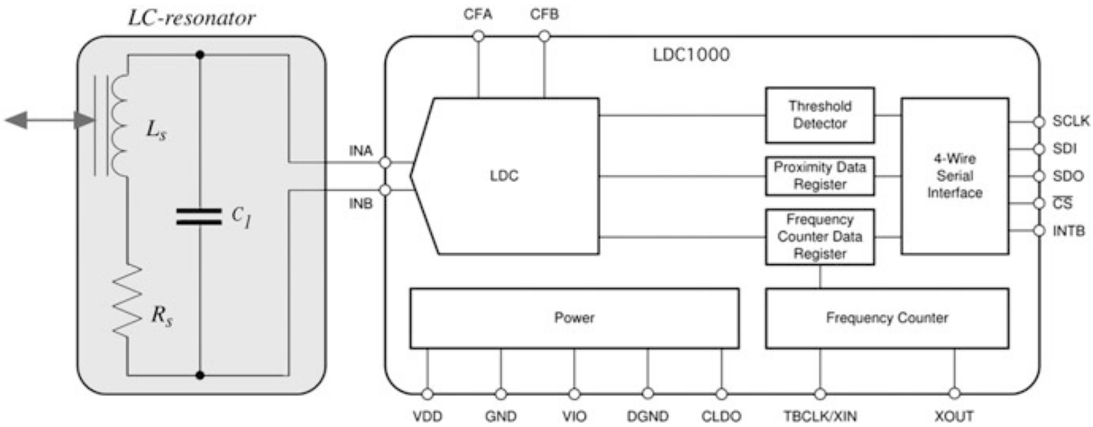


Fig. 6.36 Block diagram of integrated inductance-to-digital converter (adapted from TI data sheet of LDC1000)

composition of the target. The eddy currents then generate their own magnetic field, that according to Lenz law, opposes the original field generated by the coil. This mechanism is best compared to a transformer, where the coil is the primary core and the eddy current is the secondary core. The inductive coupling between both cores depends on distance and shape. Hence the resistance and inductance of the secondary core shows up as a distant-dependent loss resistive $R(d)$ and inductive $L(d)$ components on the primary side (coil). The conductive material shifts the resonant frequency of the LC-resonator and increases the loss resistance.

When the resonant shifts, a value of the power loss is the measure of the external object properties (distance, composition, size, etc.). The integrated circuit (Fig. 6.36) LDC1000 does not measure the series loss resistance directly; instead it measures the equivalent parallel resonance impedance of the LC circuit that operates in the range from 5 kHz to 5 MHz. The measured impedance is digitized with a 16-bit resolution and processed by the integrated circuit.

6.6 Data Transmission

Signal from a sensor may be transmitted to a receiving end of the system either in a digital format or analog. In most cases, a digital format essentially requires use of an analog-to-digital converter at the sensor's site. Transmission in a digital format has several advantages, the most important of which is a noise immunity. Since transmission of digital information is beyond the scope of this book we will not discuss it further. In many cases, however, digital transmission cannot be done for some reasons. Then, the sensor's output signal is transmitted to the receiving site in an analog form. Depending on connection, transmission methods can be divided into a two, four, and six-wire methods.

6.6.1 Two-Wire Transmission

The two-wire analog transmitters are used to couple sensors to control and monitoring devices in the process industry [10]. For example, when a temperature measurement is taken within a process, a two-wire transmitter relays that measurement to the control room or interfaces the analog signal directly to a process controller. Two wires can be used to transmit either voltage or current, however, current was accepted as an industry standard. The current carried by a two-wire loop varies in the range from 4 to 20 mA which represents the entire span of input stimuli. The minimum stimulus corresponds to 4 mA while the maximum to 20 mA.

Two wires form a loop (Fig. 6.37) of serially connected *two-wire transmitter*, conductors, power supply, and the load resistor R_{load}. The transmitter may be a voltage-to-current converter. That is, it converts the sensor signal into a variable current in a 4–20 mA range. The load resistor R_{load} develops output voltage representing the stimulus. When the sensor signal varies, the transmitter's output current varies accordingly and so the voltage across R_{load}. For example, if $R_{\text{load}} = 250\,\Omega$, the output voltage varies from 1 to 5 V.

The same current that carries information about the stimulus may also be used by the transmitter side to harvest its operating power. The lowest sensor signal produces a 4 mA current that often is sufficient for powering the transmitting side of the loop. Thus, the same two-wire loop is used for both the information transmission and delivering power to the sensor and transmitter.

The advantage of a two-wire current loop is that the loop current is independent of R_{load} and the connecting wires resistance, obviously within the limits. Since the current is produced by a current generator having very large output impedance (see Sect. 6.3.1), it remains independent on many disturbing factors, including voltages induced to the loop by external noise sources.

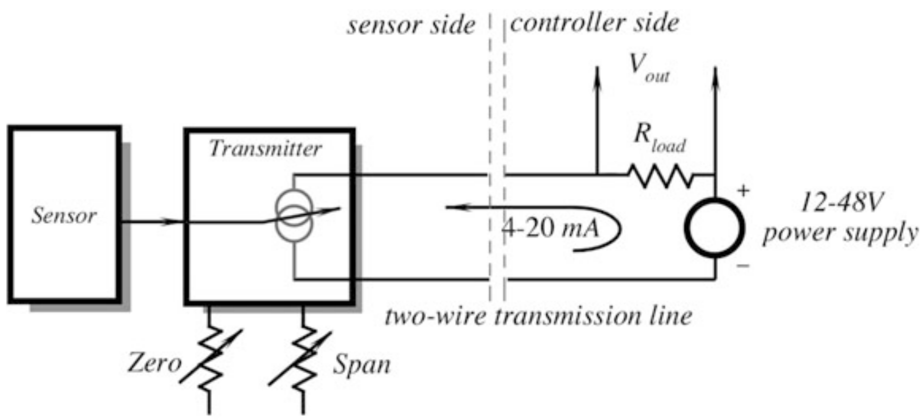


Fig. 6.37 Two-wire 4–20 mA analog data transmission

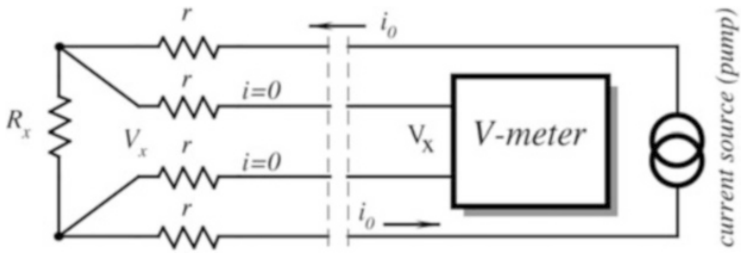


Fig. 6.38 Remote measurements of resistive sensor by 4-wire transmission

6.6.2 Four-Wire Transmission

Sometimes, it is desirable to connect a resistive sensor to a remotely located interface circuit. When such a sensor has a relatively low resistance (for instance, it is normal for piezoresistors or RTDs to have resistances on the order of 100 Ω), the connecting wire resistances pose a serious problem since they alter the excitation voltage across the sensor and add up to the sensor' resistance. The problem can be solved by using the so-called *four-wire* method shown in Fig. 6.38. The method allows measuring resistance of a remote sensor without accounting for the connecting conductors. A sensor is connected to the interface circuit through four wires. Two pairs of wires form two loops: a current loop (excitation) and voltage loop. The excitation loop includes two wires that are connected to a current source-generating excitation current i_0. The voltage loop wires are attached to a voltmeter or amplifier. A constant current source (current pump) has a very high-output resistance (see Sect. 6.3.1), therefore current i_0 in the current loop is independent of the sensor R_x and any resistance r of the wire loop. Thus, effect of wires is eliminated.

The input impedance of a voltmeter or amplifier is very high in comparison with any resistances in the voltage loop, hence no current is diverted from the current loop to the voltmeter, therefore wire resistances r that are in the voltage loop also can be ignored. A voltage drop across the sensor resistor R_x is

$$V_x = R_x i_0, \tag{6.54}$$

which is independent of any resistances r of the connecting wires. The four-wire method is a very powerful means of measuring resistances of remote detectors and is used in industry and science quite extensively.

6.7 Noise in Sensors and Circuits

Noise in sensors and circuits may present a substantial source of errors and should be seriously considered. Like diseases, noise is never eliminated, just prevented, cured, or endured, depending on its nature, seriousness, and the cost/difficulty of treating. There are two basic types of noise for a given circuit: they are *inherent* noise, which is noise arising within the circuit, and *interference* (*transmitted*) noise, which is noise picked up from outside the circuit.

Any sensor, no matter how well it was designed, never produces electric signal, which is an ideal representation of the input stimulus. Often, it is a matter of judgment to define the goodness of the signal. The criteria for this are based on the specific requirements to accuracy and reliability. Distortions of the output signal can be either *systematic* or *stochastic*. The former are related to the sensor's transfer function, its linearity, dynamic characteristics, etc. They all are the results of the sensor's design, manufacturing tolerances, material quality, and calibration. During a reasonably short time, these factors either do not change or may drift relatively slowly. They can be well defined, characterized, and specified (see Chap. 3). In many applications, such a determination may be used as a factor in the error budget and can be accounted for. Stochastic disturbances, on the other hand, often are irregular, unpredictable to some degree and may change rapidly. Generally, they are termed *noise*, regardless of their nature and statistical properties. It should be noted that word *noise*, in association with audio equipment noise, is often mistaken for an irregular, somewhat fast changing signal. We use this word in a much broader sense for all disturbances, either in stimuli, environment, or in components of the sensors and circuits from dc to the upper operating frequencies.

6.7.1 Inherent Noise

Signal which is amplified and converted from a sensor into a digital format should be regarded not just by its magnitude and spectral characteristics, but also in terms of a digital resolution. When a conversion system employs an increased digital resolution, the value of the least-significant bit (LSB) decreases. For example, the LSB of a 10-bit system with a 5 V full scale is about 5 mV, the LSB of 16 bits is 77 μV. This by itself poses a significant problem. It makes no sense to employ, say a 16-bit resolution system, if a combined noise is, for example, 300 μV. In a real world, the situation is usually much worse. There are just few sensors that are capable of producing a 5 V full-scale output signals. Most of them require amplification. For instance, if a sensor produces a full-scale output of 5 mV, at a 16-bit conversion it would correspond to a LSB of 77 nV—an extremely small signal which makes amplification an enormous task by itself. Whenever a high resolution of a conversion is required, all sources of noise must be seriously considered. In electrical circuits, noise can arise within the monolithic amplifiers and other components which are required for the feedback, biasing, filtering, etc.

Input offset voltages and bias currents may drift. In d.c. circuits, they are indistinguishable from the low magnitude signals produced by a sensor. These drifts are usually slow (within a bandwidth of tenths and hundredths of a Hz), therefore they are often called ultralow frequency noise. They are equivalent to random or predictable (e.g., temperature dependent) changing voltage and current offsets and biases. To distinguish them from the higher frequency noise, the equivalent circuit (Fig. 6.3) contains two additional generators. One is a *voltage offset* generator e_0 and the other is a *current bias* generator i_0. The noise signals (voltage and current) result from physical mechanisms within the resistors and semiconductors that are used to fabricate the circuits. There are several sources of noise whose combined effect is represented by the noise voltage and current generators.

One cause for noise is a discrete nature of electric current because current flow is made up of moving charges, and each charge carrier transports a definite value of charge (charge of an electron is 1.6×10^{-19} C). At the atomic level, current flow is very erratic. The motion of the current carriers resembles popcorn popping. This term was chosen as a good analogy for current flow and has nothing to do with the "popcorn noise" which we will discuss below. Just as popcorn, the electron movement may be described in statistical terms. Therefore, one never can be sure about very minute details of current flow. The movement of carriers are temperature related and noise power, in turn, is also temperature related. In a resistor, these thermal motions cause *Johnson noise* [11]. The mean-square value of noise voltage (which is representative of noise power) can be calculated from

$$\overline{e_n{}^2} = 4kTR\Delta f \quad \left[\mathrm{V}^2/\mathrm{Hz}\right], \tag{6.55}$$

where $k = 1.38 \times 10^{-23}$ J/K (Boltzmann constant), T is temperature in K, R is the resistance in Ω, and Δf is the bandwidth over which the measurement is made, in Hz.

For practical purposes, noise generated by a resistor at room temperature may be estimated from a simplified formula $\overline{e_n} \approx 0.13\sqrt{R \cdot \Delta f}$ in nV. For example, if noise bandwidth is 100 Hz and the resistance of concern is 10 MΩ ($10^7\,\Omega$), the average Johnson noise voltage at room temperature is estimated as $\overline{e_n} \approx 0.13\sqrt{10^7}\sqrt{100} \approx 4\,\mu\mathrm{V}$.

Even a simple resistor is a source of noise, behaving as a perpetual generator of random electric signal. Naturally, relatively small resistors generate extremely small noise, however, in some sensors Johnson noise must be taken into account. For instance, a typical pyroelectric detector uses a bias resistor on the order of 50 GΩ. If a sensor is used at room temperature within a bandwidth of 10 Hz, one may expect the average noise voltage across the resistor to be on the order of 0.1 mV—a pretty high value. To keep noise at bay, bandwidths of the interface circuits must be maintained small, just wide enough to pass the minimum required signal. It should be noted that noise voltage is proportional to square root of the bandwidth. It implies that if we reduce the bandwidth 100 times, average noise voltage will be reduced by a factor of 10. Johnson noise magnitude is constant over a broad range of frequencies. Hence, it is often called *white noise* because of the similarity to white light, which is composed of all the frequencies in the visible spectrum.

Another type of noise results from dc current flows in semiconductors. It is called *shot noise*—the name which was suggested by Schottky not in association with his own name but rather because this noise sounded like "a hail of shot striking the target" (nevertheless, shot noise is often called *Schottky noise*). Shot noise is also white noise. Its value becomes higher with increase in the bias current. This is the reason why in FET and CMOS semiconductors the current noise is quite small.

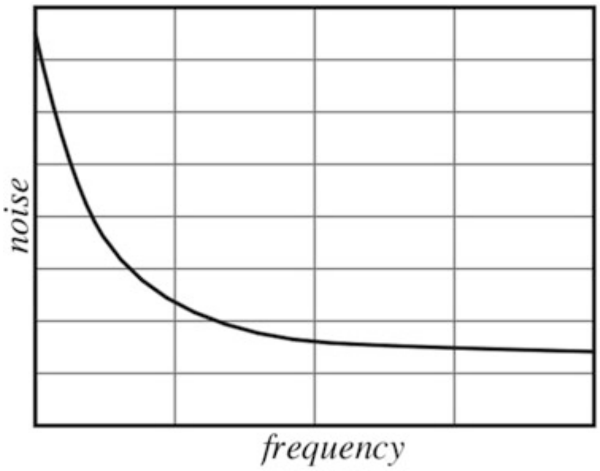


Fig. 6.39 Spectral distribution of 1/*f* "pink" noise

For a bias current of 50 pA, it is equal to about $4\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$—an extremely small current that is equivalent to movement of about 6000 electrons per second. A convenient equation for shot noise is

$$i_{sn} = 5.7 \times 10^{-4}\sqrt{I\Delta f}, \tag{6.56}$$

where I is a semiconductor junction current in pA and Δf is a bandwidth of interest in Hz.

Another a.c. noise mechanism exists at low frequencies (Fig. 6.39). Both the noise voltage and noise current sources have a spectral density roughly proportional to 1/*f*, which is called the *pink noise*, because of the higher noise contents at lower frequencies (lower frequencies are also at red side of the visible spectrum and red mixing with white makes pink). This 1/*f* noise occurs in all conductive materials, therefore it is also associated with resistors. At extremely low frequencies, it is impossible to separate the 1/*f* noise from dc drift effects. The 1/*f* noise is sometimes called a *flicker* noise. Mostly it is pronounced at frequencies below 100 Hz, where many sensors operate. It may dominate Johnson and Schottky noise and becomes a chief source of errors at these frequencies. The magnitude of pink noise depends on current passing through the resistive or semiconductive material. Nowadays progress in semiconductor technology resulted in significant reduction of 1/*f* noise in semiconductors, however, when designing a circuit, it is a good engineering practice to use metal film or wirewound resistors in sensors and the front stages of interface circuits wherever significant currents flow through the resistor and low noise at low frequencies is a definite requirement.

Combined noise from all voltage and current sources is given by the sum of squares of individual noise voltages:

$$e_E = \sqrt{e_{n1}^2 + e_{n2}^2 + \cdots + (R_1 i_{n1})^2 + (R_1 i_{n2})^2 + \cdots}. \tag{6.57}$$

A combined random noise may be presented by its *root mean square* (r.m.s.) value, that is

Table 6.4 Peak-to-peak value vs. r.m.s. (for Gaussian distribution)

Nominal p-p voltage	% of time that noise will exceed nominal p-p value
$2\times$ r.m.s.	32.0
$3\times$ r.m.s.	13.0
$4\times$ r.m.s.	4.6
$5\times$ r.m.s.	1.2
$6\times$ r.m.s.	0.27
$7\times$ r.m.s.	0.046
$8\times$ r.m.s.	0.006

$$E_{\mathrm{rms}} = \sqrt{\frac{1}{T}\int_0^T e^2 \mathrm{d}t}, \tag{6.58}$$

where T is total time of observation, e is noise voltage, and t is time.

Also, noise may be characterized in terms of the peak values which are the differences between the largest positive and negative peak excursions observed during an arbitrary interval. For some applications, in which peak-to-peak (p-p) noise may limit the overall performance (in a threshold-type devices), p-p measurement may be essential. Yet, due to a generally Gaussian distribution of noise signal, p-p magnitude is very difficult to measure in practice. Because r.m.s. values are so much easier to measure repeatedly, and they are the most usual form for presenting noise data noncontroversially, the Table 6.4 should be useful for estimating the probabilities of exceeding various peak values given by the r.m.s. values. The casually-observed p-p noise varies between $3\times$ r.m.s. and $8\times$ r.m.s., depending on patience of the observer and amount of data available.

6.7.2 Transmitted Noise

A large portion of the environmental stability is attributed to immunity of the sensor and interface circuit to noise which is originated in external sources. Figure 6.40 shows a diagram of the transmitted noise propagation. Noise comes from a source which often cannot be identified. Examples of the transmitted noise sources are voltage surges in power lines, lightnings, changes in ambient temperature, sun activity, etc. These interferences propagate toward the sensors and interface circuits, and to present a problem eventually must appear at the outputs. However, before that, they somehow affect the sensing element inside the sensor, its output terminals, or the electronic components in the circuit. Both the sensor and circuit act as receivers of the interferences.

There can be several classifications of transmitted noise, depending on how it affects the output signal, how it enters the sensor or circuit, etc.

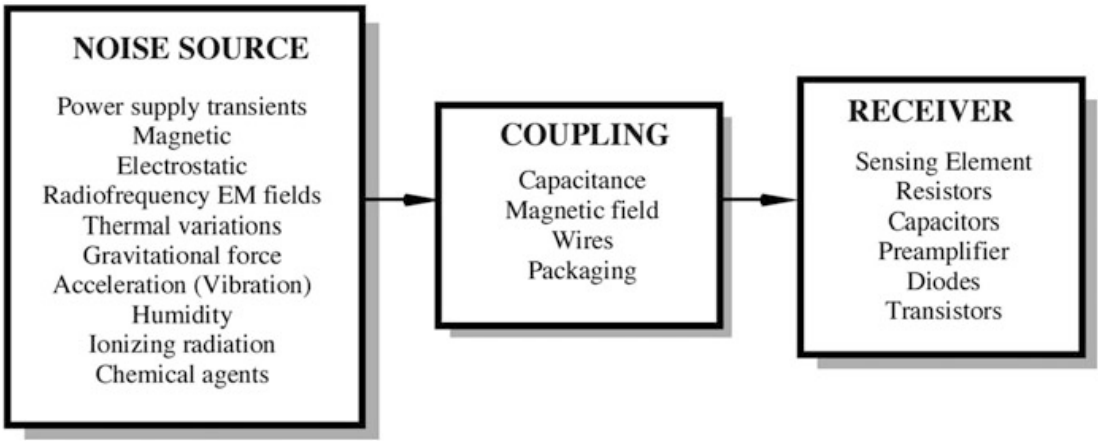


Fig. 6.40 Sources and coupling of transmitted noise

With respect to its relation to the output signals, noise can be either *additive* or *multiplicative*.

6.7.2.1 Additive Noise

Additive noise e_n is added to the useful signal V_s and mixed with it as a fully independent voltage (or current)

$$V_{out} = V_s + e_n. \tag{6.59}$$

An example of such a disturbance is depicted in Fig. 6.41b. It can be seen, that the noise magnitude does not change when the actual (useful) signal changes. As long as the sensor and interface electronics can be considered linear, the additive noise magnitude is totally independent of the signal magnitude and, if the signal is equal to zero, the output noise still will be present.

To improve noise stability against transmitted additive noise, quite often sensors are combined in pairs, that is, they are fabricated in a dual form whose output signals are subtracted from one another (Fig. 6.42). This method is called a *differential* technique (see Sect. 6.2.2). One sensor of the pair (it is called the main sensor) is subjected to the stimulus of interest s_1, while the other (reference) is shielded from the stimulus perception. Since additive noise is specific for the linear or quasilinear sensors and circuits, the reference sensor does not have to be subjected to any particular stimulus. Often, it may be equal to zero. It is anticipated that both sensors are subjected to identical transmitted noise (nose generated inside the sensor cannot be cancelled by a differential technique), which is a *common-mode* noise. This means that the noisy effects at both sensors are in-phase and have the same magnitude. If both sensors are identically influenced by the common-mode spurious stimuli, the subtraction removes the noise component. Such a sensor is often called either a dual or a *differential* sensor. Quality of the noise rejection is measured by a number which is called the *common-mode rejection ratio* (CMRR):

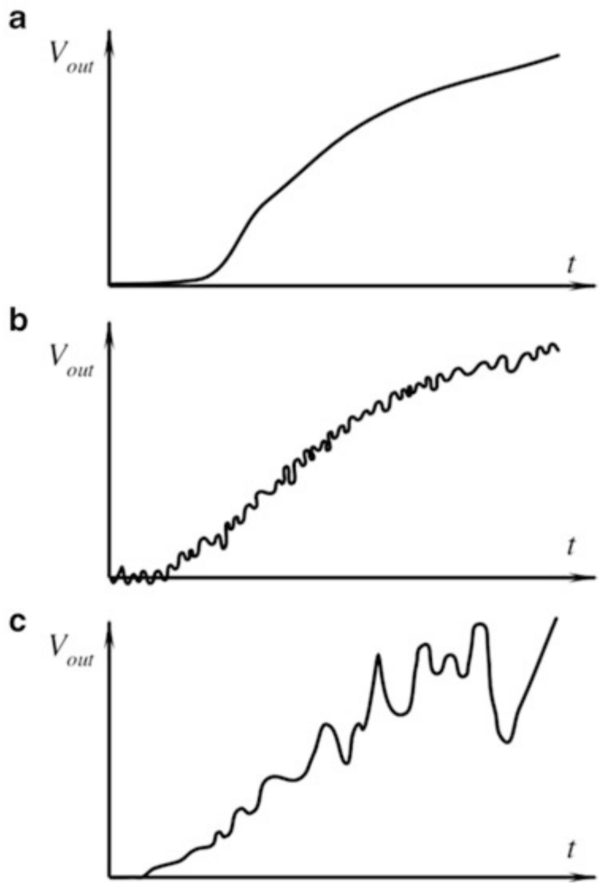


Fig. 6.41 Types of noise: noise-free signal (**a**); additive noise (**b**); multiplicative noise (**c**)

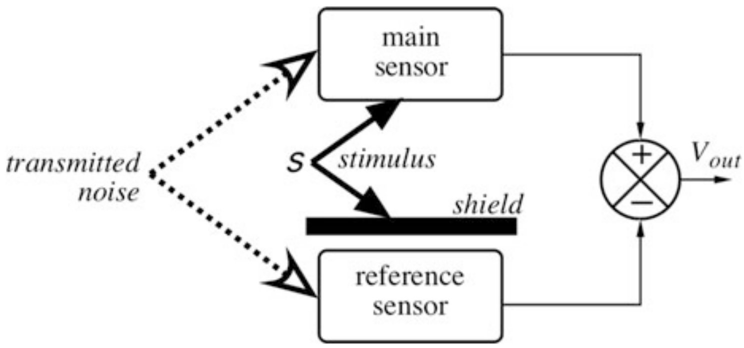


Fig. 6.42 Differential technique. Shield prevents stimulus *s* from reaching reference sensor

$$CMMR = 0.5\frac{S_1 + S_0}{S_1 - S_0}, \tag{6.60}$$

where S_1 and S_0 are output signals from the main and reference sensors, respectively. CMRR may depend on magnitude of stimuli and usually becomes smaller at greater input signals. The ratio shows how many times stronger the actual stimulus will be represented at the output, with respect to a common mode noise having the same magnitude. The value of the CMRR is a measure of the sensor's symmetry. To be an effective means of noise reduction, both sensors must be positioned as close as possible to each other, they must be very identical and subjected to the same environmental conditions. Also, it is very important that the reference sensor is reliably shielded from the actual stimulus, otherwise the combined differential response will be diminished.

6.7.2.2 Multiplicative Noise

Multiplicative noise affects the sensor's transfer function or the circuit's nonlinear components in such a manner as V_s signal's value becomes altered or *modulated* by the noise:

$$V_{out} = [1 + N(t)]V_s\ , \tag{6.61}$$

where $N(t)$ is a dimensionless function of noise. An example of such noise is shown in Fig. 6.41c. Multiplicative noise at the output disappears or becomes small (it also may become additive) when the signal magnitude nears zero. Multiplicative noise grows together with the signal's V_s magnitude. As its name implies, multiplicative noise is a result of multiplication (which essentially is a nonlinear operation) of two values where one is a useful signal and the other is a noise-dependent spurious signal.

To reduce transmitted multiplicative noise, a ratiometric technique should be used instead of a differential (see Sect. 6.2.1). Its principle is quite simple. Like for a differential technique, the sensor is fabricated in a dual form where one part is subjected to the stimulus of interest and both parts are subjected to the same environmental conditions that may cause transmitted multiplicative noise. The first, main sensor is responsive to a stimulus s_1 and is affected by a transmitted noise. The second sensor is called *reference* because a constant fixed reference stimulus s_o is applied to its input. For example, let us consider a transmitted noise being an ambient temperature that affects both the main and reference sensor identically. The output voltage of a main sensor may be approximated by

$$V_1 \approx F(T)f(s_1)\ , \tag{6.62}$$

where F(T) is a temperature-dependent function affecting the sensor's transfer function and T is temperature. Note that $f(s_1)$ is a noise-free sensor's transfer function. The reference sensor whose fixed reference input is s_0 generates voltage

$$V_o \approx \mathrm{F}(T)f(s_0). \tag{6.63}$$

Taking ratio of the above equations we arrive at

$$\frac{V_1}{V_0} = \frac{1}{f(s_0)}f(s_1). \tag{6.64}$$

Since $f(s_0)$ is constant, the ratio is not temperature dependent, and thus effect of temperature as a transmitted noise is eliminated. It should be emphasized however, that the ratiometric technique is useful only when the anticipated noise has a multiplicative nature, while a differential technique works only for additive transmitted noise. Neither technique is useful for inherent noise that is generated internally in sensors and circuits. Also, the reference sensor output may not be zero, nor too small, otherwise Eq. (6.4) will increase enormously. Value of the reference stimulus should be selected near the center of the input stimulus span, as long as the output $f(s_0)$ is far away from zero.

While inherent noise is mostly Gaussian, the transmitted noise is usually less suitable for conventional statistical description. Transmitted noise may be monotonic and systematic (like temperature effects), periodic, irregularly recurring, or essentially random, and it ordinarily may be reduced substantially by a careful sensor design and taking precautions to minimize electrostatic and electromagnetic pickup from power sources at line frequencies and their harmonics, radio broadcast stations, arcing of mechanical switches, and current and voltage spikes resulting from switching in reactive (having inductance and capacitance) circuits. Temperature effects can be reduced by placing a sensor in a thermostat. The electrical precautions may include filtering, decoupling, shielding of leads and components, use of guarding potentials, elimination of ground loops, physical reorientation of leads, components and wires, use of damping diodes across relay coils and electric motors, choice of low impedances where possible, and choice of power supply and references having low noise. Transmitted noise from vibration may be reduced by proper mechanical design. A list outlining some of the sources of transmitted noise, their typical magnitudes, and some ways of dealing with them is given in Table 6.5.

The most frequent channel for coupling electrical noise is a "parasitic" capacitance. Such a coupling exists everywhere. Any object is capacitively coupled to another object. For instance, a human standing on isolated earth develops a

Table 6.5 Typical sources of transmitted electric noise (adapted from [12])

External source	Typical magnitude	Typical cure
60/50 Hz power	100 pA	Shielding; attention to ground loops; isolated power supply
120/100 Hz supply ripple	3 μV	Supply filtering
180/150 Hz magnetic pickup from saturated 60/50 Hz transformers	0.5 μV	Reorientation of components
Radio broadcast stations	1 mV	Shielding
Switch-arcing	1 mV	Filtering of 5–100 MHz components; attention to ground loops and shielding
Vibration	10 pA (10–100 Hz)	Proper attention to mechanical coupling; elimination of leads with large voltages near input terminals and sensors
Cable vibration	100 pA	Use a low noise (carbon coated dielectric) cable
Circuit boards	0.01–10 pA/$\sqrt{\text{Hz}}$ below 10 Hz	Clean board thoroughly; use Teflon insulation where needed and guard well

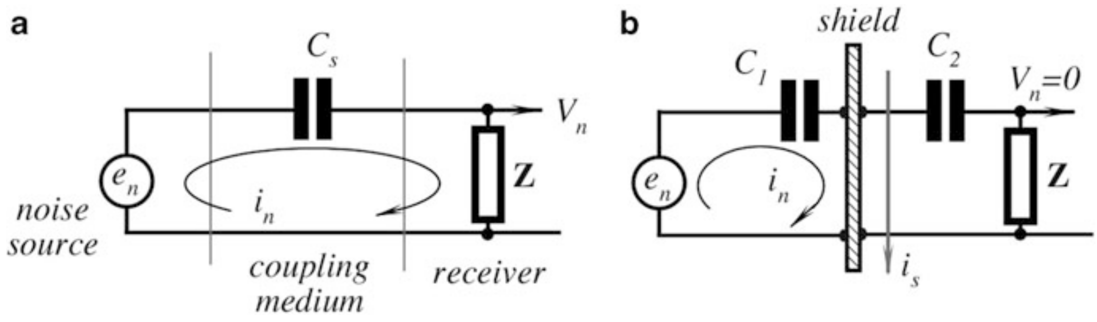


Fig. 6.43 Capacitive coupling (**a**) and electric shield (**b**)

capacitance to ground on the order of 700 pF, electrical connectors have a pin-to-pin capacitance of about 2 pF, and an optoisolator has an emitter-detector capacitance of about 2 pF. Figure 6.43a shows that an electrical noise source is connected to the sensor's internal impedance **Z** through a coupling capacitance C_s. That impedance may be a simple resistance or a combination of resistors, capacitors, inductors, and nonlinear elements, like diodes and transistors. Voltage across the impedance is a direct result of the noise signal, the value of coupling capacitance C_s and impedance **Z**. For instance, a pyroelectric detector may have an internal impedance which is equivalent to a parallel connection of a 30 pF capacitor and a 50 GΩ resistor. Let us assume that the impedance is coupled to a moving person through just 1 pF, while the person on her body carries a surface electrostatic charge of 1000 V. If we limit the main frequency of human movement to 1 Hz, the sensor would pickup the electrostatic interference of about 30 V! This is over five orders of magnitude higher than the sensor would normally produce in response to thermal radiation received from the human body.

Since many sensors and virtually all electronic circuits have some nonlinearities, the high-frequency interference signals may be rectified and appear at the output as a d.c. or slow changing noise voltage.

6.7.3 Electric Shielding

Interferences attributed to electric fields can be significantly reduced by appropriate shielding of the sensor and circuit, especially high impedance and nonlinear components. Each shielding problem must be analyzed separately and carefully. It is very important to identify the noise source and how it is coupled to the circuit. Improper shielding and guarding may only make matters worse or create a new problem.

A shielding serves two purposes [13]. First, it confines noise to a small region. This will prevent noise from getting into nearby circuits. However, the problem with such shields is that the noise captured by the shield can still cause problems if a return path that the noise is not carefully planned and implemented by an understanding of the ground system and making the connections correctly.

Second, if noise source is present in the circuit, shields can be placed around critical parts to prevent noise from getting into sensitive portions of the detectors and circuits. These shields may consist of metal boxes around circuit regions or cables with shields around the center conductors.

As it was shown in Sect. 4.1, noise that resulted from electric fields can be well controlled by metal enclosures because electric charge cannot exist on the interior of a closed conductive surface. Coupling by a mutual, or stray, capacitance can be modeled by a circuit shown in Fig. 6.43a. The parasitic capacitance C_s is the stray capacitance (having impedance Z_s at a particular frequency) between the noise source and the circuit impedance Z, which acts as a receiver of the noise. Voltage V_n is a result of the capacitive coupling. A noise current is defined as

$$i_n = \frac{e_n}{Z + Z_s}, \tag{6.65}$$

and actually produces noise voltage

$$V_n = \frac{e_n}{\left(1 + \frac{Z_s}{Z}\right)}. \tag{6.66}$$

For example, if $C_s = 2.5$ pF, $Z = 10$ kΩ (resistor) and $e_n = 100$ mV, at 1.3 MHz, the output noise will be 67 mV.

One might think that 1.3 MHz noise is relatively easy to filter out from low-frequency signals produced by a sensor. In reality, it cannot be done, because many sensors and, especially the front stages of the amplifiers, contain nonlinear components (semiconductor junctions) which act as rectifiers. As a result, after passing a nonlinear component, the spectrum of high-frequency noise shifts into a low-frequency region making the noise signal similar to voltages produced by a sensor.

When a shield is added, the coupling changes as shown in Fig. 6.43b. With the assumption that the shield has zero impedance, the noise current at the left side will be $i_n = e_n/Z_c$. On the other side of the shield, noise current will be essentially zero since there is no driving source at the right side of the circuit. Subsequently, the noise voltage over the receiving impedance will also be zero and the sensitive circuit becomes effectively shielded from the noise source.

There are several practical rules that should be observed when applying electrostatic shields:

- An electrostatic shield, to be effective, should be connected to the reference potential of any circuitry contained within the shield. If the reference is connected to ground (chassis of the frame or to earth), the shield must be connected to that ground. Grounding of shield is *useless* if current from the reference is not returned to the ground.
- If a shielding cable is used, its internal shield must be connected to the signal referenced node at the signal source side, Fig. 6.44a.

- If the shield is split into sections, as might occur if connectors are used, the shield for each segment must be tied to those for the adjoining segments, and ultimately connected only to the signal referenced node, Fig. 6.44b.
- The number of separate shields required in a data acquisition system is equal to the number of independent signals that are being measured. Each signal should have its own shield, with no connection to other shields in the system, unless they share a common reference potential (signal "ground"). In that case, all connections must be made by separate jumping wires connected to each shield at a *single point*.
- A shield must be grounded only at one point—preferably next to the sensor. A shielded cable must never be grounded at both ends (Fig. 6.45). The potential difference (V_n) between two "grounds" will cause shield current i_s to flow which may induce a noise voltage into the center conductor via magnetic coupling.
- If a sensor is enclosed into a shield box and data are transmitted via a shielded cable, Fig. 6.44c, the cable shield must be connected to the box. It's a good practice to use a separate conductor for the reference potential ("ground") inside the shield, and not to use the shield for any other purposes except shielding: *do not allow shield current to exist*.

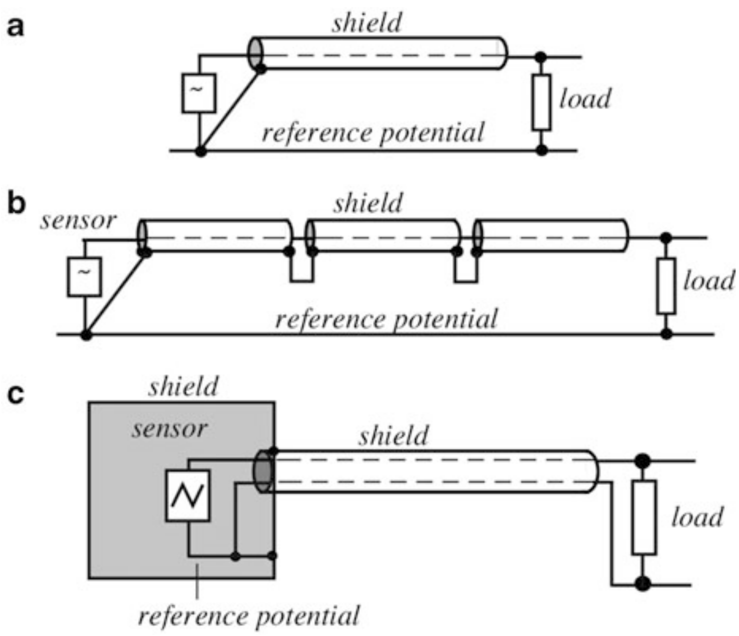


Fig. 6.44 Connections of an input cable to reference potential

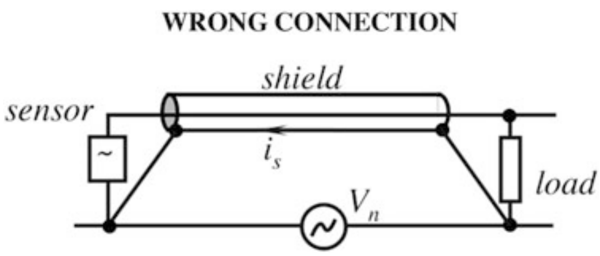


Fig. 6.45 Cable shield is erroneously grounded at both ends

- Never allow the shield to be at any potential with respect to the reference potential (except in case of driven shields as shown in Fig. 6.4b). The shield voltage couples to the center conductor (or conductors) via a cable capacitance.
- Connect shields to a ground via short wires to minimize inductance.

6.7.4 Bypass Capacitors

The bypass capacitors are used to maintain a low power supply impedance at the point of a load. Parasitic resistance and inductance in supply lines mean that the power supply impedance can be quite high. As the frequency goes up, the parasitic inductive becomes troublesome and may result in circuit oscillation or ringing effects. Even if the circuit operates at lower frequencies, the bypass capacitors are still important as high-frequency noise may be transmitted to the circuit and power supply conductors from external sources, for instance radio stations. At high frequencies, no power supply or regulator has zero output impedance. What type of capacitor to use is determined by the application, frequency range of the circuit, cost, board space, and some other considerations. To select a bypass capacitor one must remember that a practical capacitor at high frequencies may be far away from the idealized capacitor which is described in textbooks.

A generalized equivalent circuit of a capacitor is shown in Fig. 6.46. It is comprised of a nominal capacitance C, leakage resistance r_l, lead inductances L, and resistances R. Further, it includes dielectric absorption terms r and c_a, which are manifested in capacitor's "memory". In many interface circuits, especially amplifiers, analog integrators and current (charge)-to-voltage converters, dielectric absorption is a major cause for errors. In such circuits, film capacitors should be used whenever possible.

In bypass applications, r_l and dielectric absorption are second order terms but series R and L are of importance. They limit the capacitor's ability to damp transients and maintain a low power supply output impedance. Often, bypass capacitors must be of large values (10 μF or more) so they can absorb longer transients, thus electrolytic capacitors are often employed. Unfortunately, these capacitors have large series R and L. Usually, tantalum capacitors offer better results, however, a combination of aluminum electrolytic with nonpolarized (ceramic or film) capacitors may offer even further improvement. Nowadays, high-volume ceramic capacitors are available for low price.

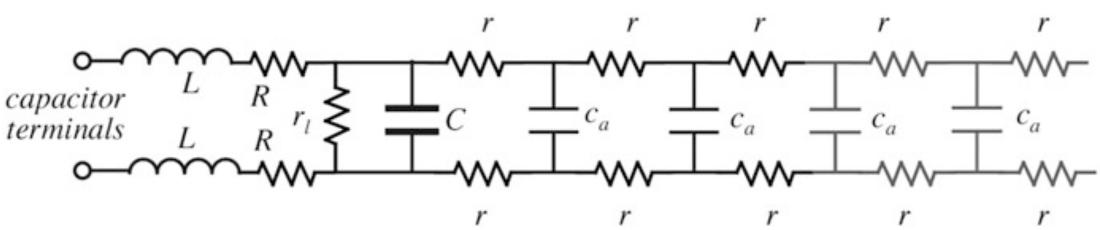


Fig. 6.46 Equivalent circuit of a capacitor

A combination of wrong types of bypass capacitors may lead to ringing, oscillation, and crosstalk between data communication channels. The best way to specify a correct combination of bypass capacitors is to first try them on a breadboard.

6.7.5 Magnetic Shielding

Proper shielding may dramatically reduce noise resulting from electrostatic and electrical fields. Unfortunately, it is much more difficult to shield against magnetic fields because they penetrate conducting materials. A typical shield placed around a conductor and grounded at one end has little if any effect on the magnetically induced voltage in that conductor. When magnetic field B_o penetrates the shield, its amplitude drops exponentially as shown in Fig. 6.47b. The skin depth δ of the shield is the depth required for the field attenuation to 37 % of that in the air. Table 6.6 lists typical values of δ for several materials at different frequencies. At high frequencies, any material from the list may be used for effective magnetic shielding, however at a lower range, steel yields a much better performance. The high-frequency magnetic shielding by an electrically conductive material arises thanks to the induced eddy currents. These circular currents according to Lenz law (Sect. 4.4.1) generate their own magnetic fields that oppose the originating field and thus provide shielding. At lower frequencies however, eddy currents have much lower efficiencies.

For improving low-frequency magnetic field shielding, a shield consisting of a high-permeability magnetic material (e.g., mumetal) should be considered. However, the mumetal effectiveness drops at higher frequencies and strong magnetic fields. An effective magnetic shielding can be accomplished with thick steel shields at higher frequencies. Since magnetic shielding is very difficult, the most effective approach at low frequencies is to minimize the strength of magnetic fields, minimize the magnetic loop area at the receiving end, and select the optimal geometry of conductors. Some useful practical guidelines are as follows:

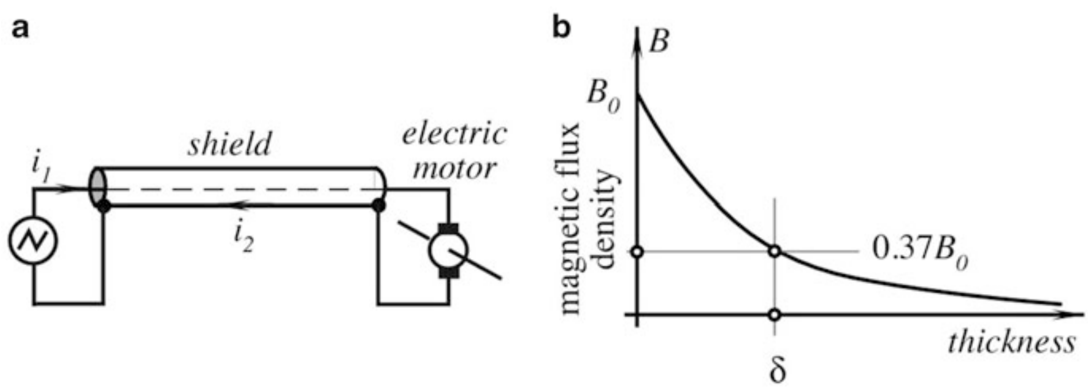


Fig. 6.47 Reduction of transmitted magnetic noise. Powering load device through coaxial cable (**a**); magnetic shielding improves with thickness of shield (**b**)

Table 6.6 Skin depth δ (mm) versus frequency

Frequency	Copper	Aluminum	Steel
60 Hz	8.5	10.9	0.86
100 Hz	6.6	8.5	0.66
1 kHz	2.1	2.7	0.20
10 kHz	0.66	0.84	0.08
100 kHz	0.2	0.3	0.02
1 MHz	0.08	0.08	0.008

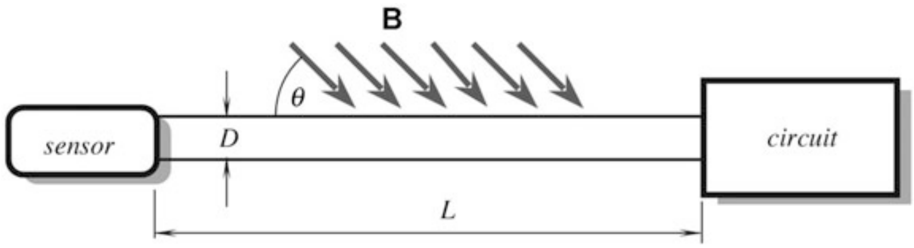


Fig. 6.48 Receiver's loop formed by long conductors

- Locate the receiving circuit as far as possible from the source of the magnetic field.
- Avoid running wires parallel to the magnetic field; instead, cross the magnetic field at right angles.
- Shield the magnetic field with an appropriate material for the frequency and strength.
- Use a twisted pair of wires for conductors carrying the high-level current that is the source of the magnetic field. If the currents in the two wires are equal and opposite, the net field in any direction over each cycle of twist will be zero. For this arrangement to work, none of the current can be shared with another conductor, for example, a ground plane, which may result in ground loops.
- Use a shielded cable with the high-level source circuit's return current carried by the shield, Fig. 6.47a. If the shield current i_2 is equal and opposite to that of the center conductor i_1, the center conductor field and the shield field will cancel, producing a zero net field. This case seems a violation of a rule "*no shield currents*" for the receiver's circuit, however, the shielded cable here is not used to electrostatically shield the center conductor. Instead, the geometry produces a cancellation of the magnetic field which is generated by current supplied to a "current-hungry" device (an electric motor in this example).
- Since magnetically induced noise depends on area of the receiver loop, the induced voltage due to magnetic coupling can be reduced by making the loop's area smaller.

What is the receiver's loop? Figure 6.48 shows a sensor which is connected to the load circuit via two conductors having length L and separated by distance D. The rectangular circuit forms a loop area $a = L \times D$. The voltage induced in series with the loop is proportional to magnetic field $\boldsymbol{B}$, the area and cosine of its angle to the field. Thus, to minimize noise, the loop should be oriented at right angles to the field, and its area should be minimized.

The area can be decreased by reducing the length of the conductors and/or decreasing the distance between the conductors. This is easily accomplished with a twisted pair, or at least with a tightly cabled pair of conductors. It is a good practice to pair the conductors so that the circuit wire and its return path will always be together. This requirement shall not be overlooked. For instance, if wires are correctly positioned by a designer, a service technician may reposition them during the repair work. A new wire location may create a disastrous noise level. Hence, a general rule is—know the area and orientation of the wires and permanently secure the wiring.

6.7.6 Mechanical Noise

Vibration and *acceleration effects* are also sources of transmitted noise in sensors which otherwise should be immune to them. These effects may alter transfer characteristics (multiplicative noise), or the sensor may generate spurious signals (additive noise). If a sensor incorporates certain mechanical elements, vibration along some axes with a given frequency and amplitude may cause resonant effects. For some sensors, acceleration is a source of noise. For instance, pyroelectric detectors possess piezoelectric properties. The main function of a pyroelectric detector is to respond to thermal radiation. However, such environmental mechanical factors as fast changing air pressure, strong wind, or structural vibrations cause the sensor to respond with output signals which often are indistinguishable from responses to normal stimuli. If this is the case, a differential noise cancellation may be quite efficient (see Sect. 6.7.2).

6.7.7 Ground Planes

For many years ground planes have been known to electronic engineers and printed circuit designers as a "mystical and ill-defined" cure for spurious circuit operation. Ground planes are primarily useful for minimizing circuit inductance. They do this by utilizing the basic magnetic theory. Current flowing in a wire produces an associated magnetic field (Sect. 4.3). The field's strength is proportional to the current i and inversely related to the distance r from the conductor:

$$B = \frac{\mu_0 i}{2\pi r}. \tag{6.67}$$

Thus, we can imagine a current carrying wire surrounded by a magnetic field. Wire inductance is defined as energy stored in the field setup by the wire's current. To compute the wire's inductance requires integrating the field over the wire's length and the total area of the field. This implies integrating on the radius from the wire surface to infinity. However, if two wires carrying the same current in opposite directions are in close proximity, their magnetic fields are cancelled. In this case, the virtual wire inductance is much smaller. An opposite flowing current is called *return current*. This is the underlying reason for ground planes. A ground plane provides a return path directly under the signal carrying conductor through which return current can flow. Return current has a direct path to ground, regardless of the number of branches associated with the conductor. Currents will always flow through the return path of the lowest impedance. In a properly designed ground plane this path is directly under the signal conductor. In practical circuits, a ground plane is one side of the board and the signal conductors are on the other. In the multilayer boards, a ground plane is usually sandwiched between two or more conductor planes. Aside from minimizing parasitic inductance, ground planes have additional benefits. Their flat surface minimizes resistive losses due to “skin effect” (a.c. current travel along a conductor’s surface). Additionally, they aid the circuit’s high-frequency stability by referring stray capacitance to the ground. Even though ground planes are very beneficial for digital circuits, using them for current return of analog sensor signals are dangerous—likely digital currents in a ground will create strong interferences in the analog part of the circuit.

Some practical suggestions:

- Make ground planes of as much area as possible on the components side (or inside for the multilayer boards). Maximize the area especially under traces that operate with high frequency or digital signals.
- Mount components that conduct fast transient currents (terminal resistors, ICs, transistors, decoupling capacitors, etc.) as close to the board as possible.
- Wherever a common ground reference potential is required, use separate conductors for the reference potential and connect them all to the ground plane at a common point to avoid voltage drops due to ground currents.
- Use separate nonoverlapping ground planes for digital and analog sections of the circuit board and connect them at one point only at the power supply terminals.
- Keep the trace length short. Inductance varies directly with length and no ground plane will achieve perfect cancellation.

6.7.8 Ground Loops and Ground Isolation

When a circuit is used for low-level input signals, a circuit itself may generate enough noise and interferences to present a substantial problem for accuracy. Sometimes, when a circuit is correctly designed on paper and a bench breadboard shows quite a satisfactory performance, when a production prototype with the printed circuit board is tested, the accuracy requirement is not met. A difference between a breadboard and PC-board prototypes may be in the physical layout of conductors. Usually, conductors between electronic components are quite

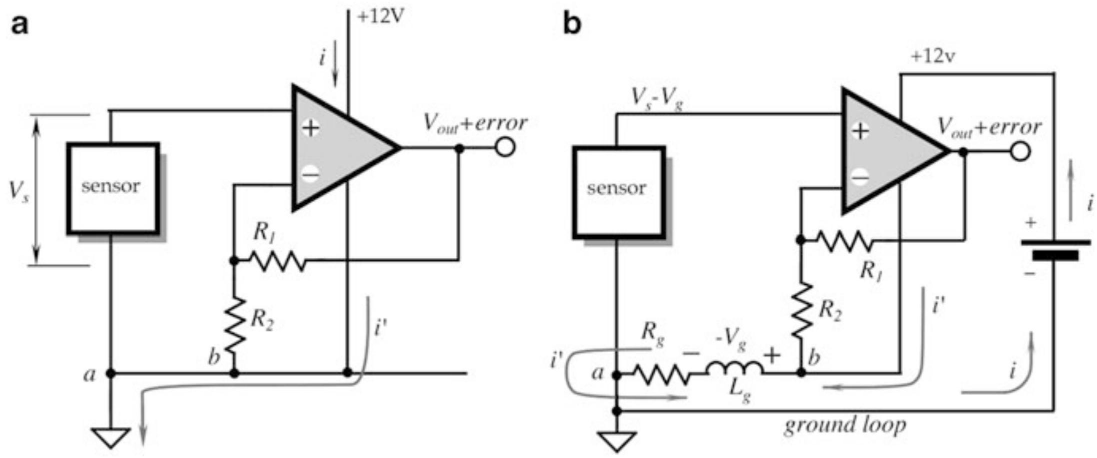


Fig. 6.49 Wrong connection of ground terminal to circuit (**a**); path of supply current through ground conductors (**b**)

specific—they may connect a capacitor to a resistor, a gate of a JFET transistor to the output of an operational amplifier, etc. However, there are at least two conductors, which, in most cases are common for the majority of the electronic circuit. These are the power supply bus and the ground bus. Both of them may carry undesirable signals from one part of the circuit to another, specifically, they may couple strong output signals to the sensors and input stages.

A power supply bus carries supply currents to all stages. A ground bus also carries supply currents, but, in addition, it is often used to establish a reference base for an electrical signal. For any measurement circuit electrical cleanliness of a reference is essential. Interaction of the two functions (power supply and reference) may lead to a problem which is known as *ground loop*. We illustrate it in Fig. 6.49a where a sensor is connected to a noninverting input of an amplifier which may have a substantial gain. The amplifier is connected to the power supply and draws current i which is returned to the ground bus as i'. The sensor generates voltage V_s which is fed to the input of the amplifier. A ground wire is connected to the circuit at point a—right next to the sensor's terminal. A circuit has no visible error sources, nevertheless, the output voltage contains substantial error. A noise source is developed in a wrong connection of ground wires. Figure 6.49b shows that the ground conductor is not ideal. It may have some finite resistance R_g and inductance L_g. In this example, supply current, while returning to the battery from the amplifier, passes through the ground bus between points b and a resulting in voltage drop V_g. This drop, however small, may be comparable with signals produced by the sensor. It should be noted that voltage V_g is serially connected with the sensor and directly applied to the amplifier's input. In other words, the sensor is not referenced to a clean ground. Ground currents may also contain high-frequency components, than the bus inductance will produce quite strong spurious high-frequency signals which not only add noise to the sensor, but may cause circuit instability as well.

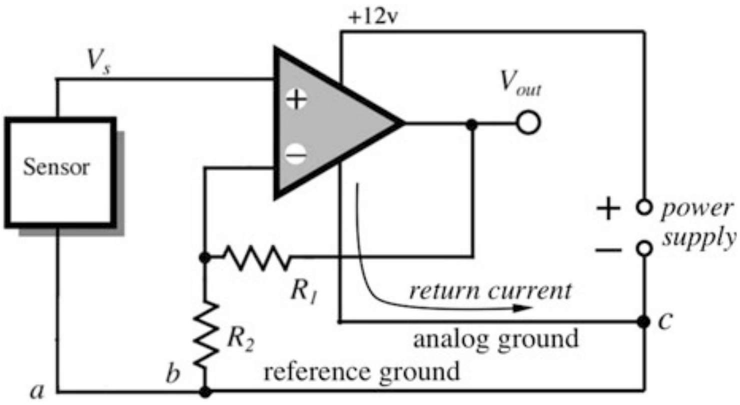


Fig. 6.50 Correct grounding of a sensor and interface circuit

As an example, consider a thermocouple temperature sensor which generates voltage corresponding to 50 μV/°C of the object's temperature. A low noise amplifier has quiescent current, $i = 1$ mA, which passes through the ground loop having resistance $R_g = 0.2\,\Omega$. The ground-loop voltage $V_g = iR_g = 0.2$ mV corresponding to an error of −4 °C! The cure is usually quite simple—ground loops must be broken. The most critical rule of the circuit board design: *never use the same conductor for the reference potential and power supply currents*. A circuit designer should always separate a reference ground from the current carrying grounds, especially serving digital devices. Thus, it is advisable to have at least three grounds: reference, analog, and digital.

The *reference ground* shall be used only for connecting the sensor components that produce low-level input signals, all front stage amplifier input components that need be referenced to a ground potential, and the reference input of an ADC. The *analog ground* shall be used exclusively for returning stronger currents from the analog interface circuits. And the *digital ground* shall be used only for binary signals, like microprocessors, digital gates, etc. There may be a need for additional "grounds", for example those that carry relatively strong currents, especially containing high-frequency signals (LEDs, relays, motors, heaters, etc.). Figure 6.50 shows that moving the ground connection from the sensor's point *a* to the power terminal point *c* prevents formation of spurious voltages across the ground conductor connected to the sensor and a feedback resistor R_2.

A rule of thumb is to *join all "grounds" on a circuit board only at one point,* preferably at the power source. Grounding at two or more spots may form ground loops which often is very difficult to diagnose.

6.7.9 Seebeck Noise

This noise is a result of the Seebeck effect (Sect. 4.9.1) which is manifested as generation of an *electromotive force* (*e.m.f.*) when two dissimilar metals are joined together. The Seebeck *e.m.f.* is small and for many sensors may be simply ignored. However, when absolute accuracy on the order of 10–100 μV is required, that noise

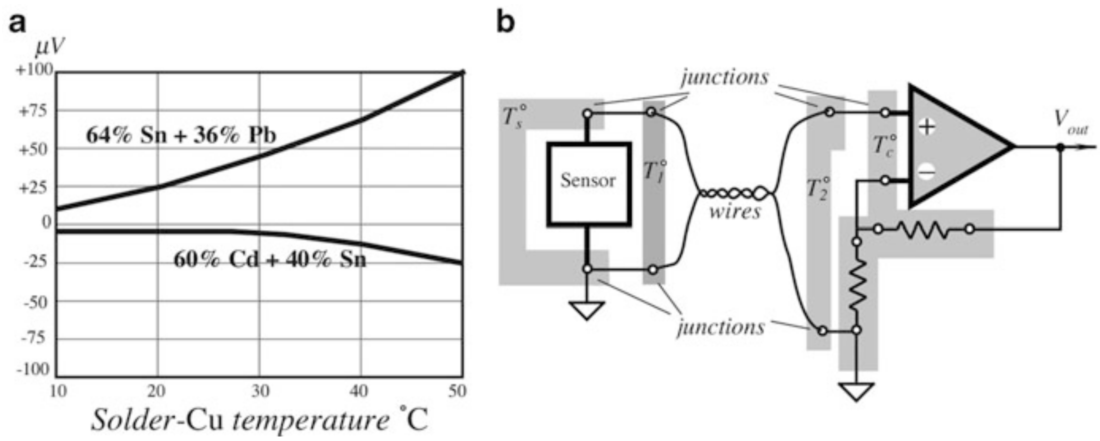


Fig. 6.51 Seebeck *e.m.f.* developed by solder-copper joints (**a**) (adapted from [14]); maintaining joints at the same temperature reduces Seebeck noise (**b**)

should be taken into account. A connection of two dissimilar metals produces a temperature sensor. However, when temperature sensing is not a desired function, a thermally induced *e.m.f.* is a spurious signal. In electronic circuits, connection of dissimilar metals can be found everywhere: connectors, switches, relay contacts, sockets, wires, etc. For instance, the copper PC board cladding connected to kovar™[4] input pins of an integrated circuit creates an offset voltage of 40 μV × ΔT where ΔT is the temperature gradient in °C between two dissimilar metal contacts on the board. The common lead-tin solder, when used with the copper cladding, creates a thermoelectric voltage between 1 and 3 μV/°C. There are special cadmium-tin solders available to reduce these spurious signals down to 0.3 μV/°C. Figure 6.51a shows Seebeck *e.m.f.* for two types of solder. Connection of two identical wires fabricated by different manufacturers may result in voltage having slope on the order of 200 nV/°C.

In many cases, Seebeck *e.m.f.* may be eliminated by a proper circuit layout and thermal balancing. It is a good practice to limit the number of junctions between the sensor and front stage of the interface circuit. Avoid connectors, sockets, switches, and other potential sources of *e.m.f.* to the extent possible. In some cases this will not be possible. In these instances, attempt to balance the number and type of junctions in the circuit's front stage so that differential cancellations occur. Doing this may involve deliberately creating and introducing junctions to offset necessary junctions. Junctions which intent to produce cancellations must be maintained at the same temperature. Figure 6.51b shows a remote sensor connection to an amplifier where the sensor junctions, input terminal junctions, and amplifier components junctions are all maintained while at different but properly arranged

[4] Trademark of Westinghouse Electric Corp.

temperatures. Such thermally balanced junctions must be maintained at a close physical proximity and preferably on common heat sinks. Air drafts and temperature gradients in the circuit boards and sensor enclosures must be avoided.

6.8 Batteries for Low-Power Sensors

Modern development of integrated sensors and need for long-term remote monitoring and data acquisition demand use of reliable and high-energy density power sources. History of battery development goes back to Volta and shows a remarkable progress during last decades. Well-known old electrochemical power sources improve dramatically. Examples are C-Zn, alkaline, Zn-air, NiCd, and lead-acid batteries. Nowadays, Zn-air, Ni-metal-hydride, and especially lithium batteries (such as Li-MnO_2) are the most popular energy sources.

All batteries can be divided into two groups: *primary*—single use devices, and *secondary* (rechargeable)—multiple use devices.

Often, batteries are characterized by energy per unit weight, however, for miniature sensor applications energy per unit volume often becomes more critical (see Table A.20)

In general, energy delivered by a battery depends upon the rate at which power is withdrawn. Typically, as the current increases the amount of energy delivered decreases. Battery energy and power are also affected by construction of battery, the size, and the duty cycle of current delivery. Manufacturers usually specify batteries as ampere-hours or watt-hours when discharged at a specific rate to a specific voltage cutoff.

If the battery capacitance is C (in mA h) and the average current drain is I (mA), the time of a battery discharge (lifetime for a primary cell when in use) is defined as

$$t = \frac{C}{In}, \tag{6.68}$$

where n is a duty cycle. For instance, if the battery is rated as having capacity of 100 mA h, the load current consumption is 5 mA and the circuit works only 3 min every hour (duty cycle is 3/60), the battery will last approximately for

$$t = \frac{C}{In} = \frac{100}{5\frac{3}{60}} = 400\,\mathrm{h}$$

Yet, the manufacturer's specification shall be used with a grain of salt and *only as a guideline*, because the specified discharge rate rarely coincides with the actual power consumption. Also, a capacity is rated for a specific cutoff voltage because when a battery discharges its output voltage drops. For example, a fresh battery capacity is specified as 220 mA h for a cutoff voltage 2.6 V, while the load needs minimum 2.8 V for its operation. Thus, the actual battery capacity will be less then specified by the manufacturer.

It is highly recommended to determine battery life experimentally, rather than rely on calculation. When designing electronic circuit, its power consumption shall be determined during various operating modes and over the operating temperature range. Then, these values of power consumption should be used in simulation of the battery load to determine the useful life with a circuit-specific cutoff voltage in mind. The accelerated life tests of a battery shall be used with caution, since the useful capacity of a battery greatly depends on the load, operational current profile, and a real duty cycle.

Sometimes, a circuit draws high currents during short times (pulse mode) and the battery ability to deliver such pulse current should be evaluated since the battery internal resistance may be a limiting factor, so the battery would not be able to deliver as much current as needed. A solution is to augment the battery with a large parallel capacitor, for example 10–100 μF, that can be used as a charge storage tank for quick delivering the current bursts. Another advantage of the parallel capacitor is that it prolongs the battery life in pulsing applications [15].

6.8.1 Primary Cells

The construction of a battery cell determines its performance and cost. Most primary cells (disposable batteries) employ single thick electrodes arranges in parallel or concentric configuration and aqueous electrolytes. Most small secondary cells (rechargeable batteries) are designed differently—they use “wound” or “jelly roll” construction, in which long thin electrodes are wound into a cylinder and placed into a metal container. This results in a higher power density, but with decreased energy density and higher cost. Due to low conductivity of electrolytes, many lithium primary cells also use “wound” construction [16].

6.8.1.1 Alkaline Manganese Batteries

Demand for these batteries grew significantly, especially after a major improvement—elimination of mercury from the zinc anode. The alkaline batteries are capable of delivering high currents, have improved power/density ratio and at least 5 years of shelf life (Table A.20)

6.8.1.2 Primary Lithium Batteries

Most of these batteries are being produced in Japan and China. Popularity of lithium-manganese dioxide cells grows rapidly since they were first introduced by Sony in 1991. They have higher operating voltage, wide range of sizes and capacities, and excellent shelf life (Table A.21). Lithium iodine cells have very high-energy density and allow up to 10 years of operation in a pacemaker (implantable heart rate controller). However, these batteries are designed with a low conductivity solid-state electrolyte and allow operation with very low current drain (in the order of microamperes), which is quite sufficient for many passive sensors.

Amount of lithium in batteries is quite small, because just 1 g is sufficient for producing capacity of 3.86 A h. Lithium cells are exempt from environmental regulations, but still are considered hazardous because of their flammability and thus restricted for transporting by aircrafts. Lithium-ion cells with cobalt cathodes should never rise above 130 °C (265 °F) because at 150 °C (302 °F) the cell becomes thermally unstable.

For portable equipment, thin batteries are highly desirable due to their small thickness. There is a tradeoff between thickness and surface area—the area becomes larger for thinner batteries. For example, lithium/manganese dioxide battery CP223045 has thickness 2.2 mm with the surface area 13.5 cm^2. It has an impressive capacity of 450 mA h with only 2 % self-discharge per year.

6.8.2 Secondary Cells

Secondary cells (Tables A.22 and A.23) are *rechargeable batteries*.

Sealed lead acid batteries offer small size at large capacities and allow about 200 cycles of life at discharge times as short as 1 h. The main advantages of these cells are low initial cost, low self-discharge, and an ability to support heavy loads and to withstand harsh environments. Besides, these batteries have long life. The disadvantages include relatively large size and weight as well as potential environmental hazard due to presence of lead and sulfuric acid.

Sealed *Nickel-Cadmium* (NiCd) and *Nickel-Metal Hydrate (Ni-MH)* are the most widely used secondary cells, being produced at volumes over one billion cells per year. Typical capacity for an “AA” cell is about 2000 mA h and even higher from some manufacturers. The NiCd cells are quite tolerant of overcharge and overdischarge. An interesting property of NiCd is that charging is the endothermic process that is the battery absorbs heat, while other batteries warm up when charging. Cadmium, however, presents potential environmental problem. Bi-MH and modern NiCd do not exhibit “memory” effect, that is, a partial discharge does not influence their ability to fully recharge. The Nickel-Metal Hydrate cells are nearly direct replacement for NiCd, yet they yield better capacity, but have somewhat poorer self-discharge.

A *Lithium Ion Polymer battery* (LiPo or LIP) contain a nonliquid electrolyte, which makes it a solid-state battery. This allows fabricating it in any size and shape, however, these batteries are most expensive.

Rechargeable *alkaline batteries* have low cost and good power density. However, their life cycles are quite low.

6.8.3 Supercapacitors

Supercapacitors (SC) or ultracapacitors fill the gap between secondary batteries and regular capacitors. They are characterized by very large capacitances ranging from 1 to 200 F and a low internal resistance ranging from 0.07 to

0.7 Ω (www.maxwell.com). While supercapacitors have energy densities that are approximately 10 % of the conventional batteries, their power density is generally 10–100 times greater. This results in much shorter charge/discharge times than for batteries. While working in tandem with a regular primary or secondary battery for applications that require both a constant low-power discharge for continual function and a pulse power for peak loads they relieve batteries of peak power functions resulting in an extension of battery life and reduction of the overall battery size and cost. The SC main advantage is a long life—about 0.5 million cycles, much exceeding that of a secondary cell that has a typical charge-discharge cycles ranging from 500 to 10,000 (for lithium-ion cells). On a negative side, a SC has a relatively large leakage current and its typical operating voltage is not exceeding 2.8 V. This precludes the SC from replacing batteries for a prolonged storage of electric charge. Thus, these capacitors are the most beneficial when their charge can be constantly replenished either from a battery or energy harvester.

6.9 Energy Harvesting

Providing electric power for a sensing module does not always have a simple solution. When electric power line is available or a battery can be periodically replaced—there is no issue. However, replacing a battery is not always easy or even possible. In such special cases, energy should be obtained or harvested from ambient sources [21]. This requires receiving one type of energy and converting it to d.c. electric power. Examples of the potential sources for energy harvesting are:

Thermal, where energy can be obtained from thermal gradients. For example, a temperature difference between a human body and environment can be converted into electricity by thermoelectric elements.

Mechanical, when energy is obtained by stressing a special transducer or moving a coil or magnet. For example, a piezoelectric element may be built into a shoe sole and stressed at every step to generate voltage spikes that can charge a capacitor or battery. Another example is a floating sensing module with a built-in electromagnetic transducer for converting mechanical energy of water waves to electric power.

Light of different wavelengths can be converted to electricity by use of a photoeffect or thermoelectric effect.

Acoustic, when sound (pressure waves) can be converted to electricity by use of special microphones or hydrophones.

Electromagnetic (RF) sources (far-field), such as radio stations emitting electromagnetic fields that can be detected and converted to a d.c. power.

Magnetic (RF) sources (near field), such as power transformers and variable magnetic filed transmitters in close proximities.

6.9.1 Light Energy Harvesting

Quanta of light carries a substantial energy depending on its wavelength, Eq. (5.3). Portion of that energy can be captured and used for powering electronic circuits. From low-energy light known as thermal radiation, electric power can be harvested by using pyroelectric cells [17]. In these cells, light first is converted into heat and subsequently heat is converted to electricity. For visible light, the most common type of a conversion device is a photovoltaic (PV) or solar cell. These cells are similar in many ways to a battery because they supply direct current. A PV cell has a positive and a negative side, just like a battery.

Photovoltaic cells are made from single crystal silicon *pn* junctions, the same as photodiodes with a very large light sensitive region. But unlike photodiodes, they are used without the reverse bias. When illuminated, the light energy causes electrons to flow through the *pn* junction and generates an open circuit voltage of about 0.58 V (for an individual solar cell). To boost output voltage (and power), individual solar cells can be connected together in series to form a solar panel. The amount of available current from a PV cell depends upon the light intensity, size of the cell, and its efficiency which is generally rather low: 15–20 %. To increase the overall efficiency of the cell, commercially available solar cells use polycrystalline silicon or amorphous silicon, which have no crystalline structure, and can generate maximum currents between 20 and 40 mA/cm^2.

For a given illumination, a cell has a flat ampere-voltage characteristic shown in Fig. 6.52a. The optimal operating point is where the cell can deliver maximum power, that is where the product of current and voltage is the highest. These points are indicated by the dots. The higher illumination the stronger the PV current, i. For illustration of the PV energy harvesting, consider a panel of seven solar cells connected in series as shown in Fig. 6.52b. They deliver about 4 V that charge a secondary battery whose purpose is to provide power in darkness, when the PV cells output no current. To prevent a current reverse flow in darkness, a diode D is used in the current path. Whenever ambient light comes on, the battery is trickle-charged. The battery output may be regulated to provide a fixed voltage to the sensing

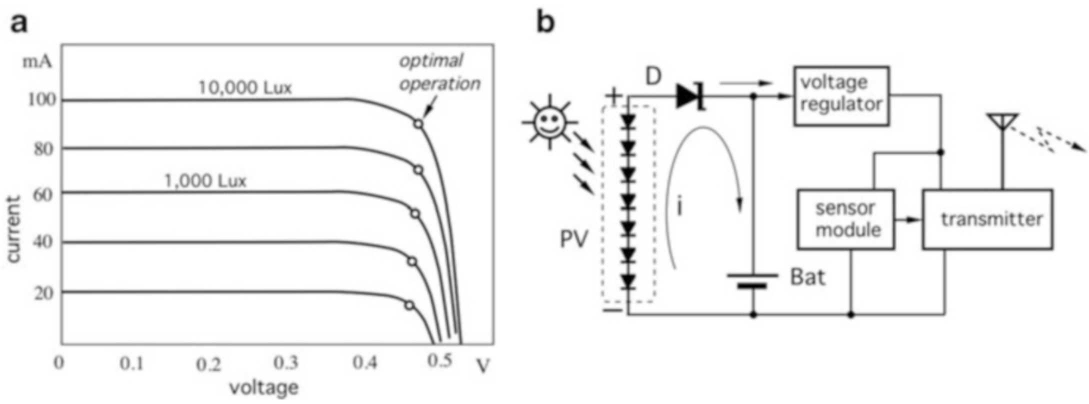


Fig. 6.52 Volt-ampere characteristic of photovoltaic cell (**a**) and use of photovoltaic battery to power sensing module (**b**)

module and data transmitter, for example Bluetooth. Obvious disadvantages of light harvesting include a need for the solar panels being exposed to bright ambient light and a potential soiling of the panels by airborne dirt.

6.9.2 Far-Field Energy Harvesting

Ambient space around us is “packed” with electromagnetic fields (EMF) of endless frequencies having substantial combined energy. Harvesting that entire energy is not possible with modern technologies, yet tapping of small incident EMF at selected frequencies is quite practical [18, 19]. The key point is—what frequency? The answer depends on proximity of EMF sources, such as radio stations, wireless routers, etc. Since the EMF strength drops dramatically when moving away from the source, typical distances between the EMF source and harvester preferably are no more than 3 m (10 ft). The harvesting range must be at least 70 % of the wavelength of EMF or longer. Therefore, this type of the electromagnetic reception is called a *far-field* harvesting. In some occasions, when the transmitter emits significant EMF power (e.g., broadcasting or communication station), the harvesting range may be as long as 40 m. The RF to d.c. power conversion system operates more efficiently in UHF frequencies in the industrial, scientific, and medical bands (ISM band) of 902–928 MHz. In this frequency range, RF power is transmitted more efficiently for longer distances and experiences lower propagational losses than higher frequency bands (i.e., 2.4 GHz).

Harvested EMF energy is used for charging battery as shown in Fig. 6.53a. The circuit includes the antenna with an LC_1 resonant tank that is tuned to a selected frequency. The tuning may be fixes or automatically adjustable to maximize the output power. The antenna preferably should be placed on the outside surface of the converter. The RF voltage from the resonant tank is rather small for rectification and charging the battery, thus first it passes through a voltage multiplier called the Villard cascade, consisting of several diodes (D) and capacitors (C) that rectify and multiply the d.c. output to the level that is sufficient for charging the battery.

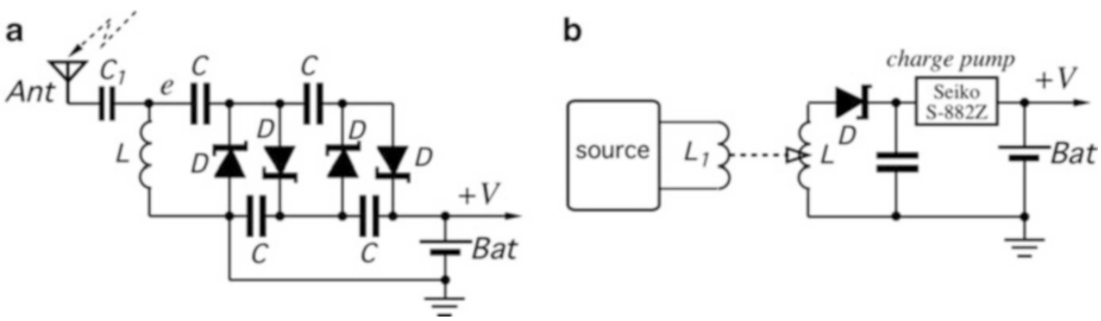


Fig. 6.53 Far-field energy harvesting circuit (**a**) and near-field magnetic coupling (**b**) with charge pump

6.9.3 Near-Field Energy Harvesting

In the near-field harvester, only the magnetic component of the EMF vector is employed, thus this method of energy harvesting resembles a transformer with two magnetically coupled coils [20] as illustrated in Fig. 6.53b. Note that voltage after the rectifier is rather small (about 0.5 V), thus a charge pump is employed to increase that voltage to about 2 V. This type of a coupling dramatically limits the range of energy transfer—typically it does not exceed 5 cm and must be less that 70 % of the wavelength. Due to its magnetic nature, a near-filed energy transfer sometimes is called a coupling to H-field. These fields are abundant in close vicinity of many appliances that use high electric currents. In some cases, H-filed in the RF range is specially generated by a radio antenna for providing power to a wireless circuit, for example for RFID tags. Thus, one coil that emanates EMF (for example in the NFC frequency of 13.56 MHz) has to be inductively coupled with a receiving coil and battery charger. The RF signal may be modulated to enable transfer of data along with supplying energy.

References

1. Widlar, R. J. (1980). Working with high impedance Op Amps, AN24. *Linear application handbook*. National Semiconductor.
2. Park, Y. E., et al. (1983). An MOS switched-capacitor readout amplifier for capacitive pressure sensors. *IEEE Custom IC Conf.* (pp. 380–384).
3. Cho, S. T., et al. (1991). A self-testing ultrasensitive silicon microflow sensor. *Sensor Expo Proceedings* (p. 208B-1).
4. Ryhänen, T. (1996). Capacitive transducer feedback-controlled by means of electrostatic force and method for controlling the profile of the transducing element in the transducer. *U.S. Patent 5531128*.
5. Pease, R. A. (1983, January 20). Improve circuit performance with a 1-op-amp current pump. *EDN* (pp. 85–90).
6. Bell, D. A. (1981). *Solid state pulse circuits* (2nd ed.). Reston, VA: Reston Publishing Company.
7. Sheingold, D. H. (Ed.). (1986). *Analog-digital conversion handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
8. Johnson, C., et al. (1986). Highly accurate resistance deviation to frequency converter with programmable sensitivity and resolution. *IEEE Transactions on Instrumentation and Measurement, IM-35*, 178–181.
9. AVR121: Enhancing ADC resolution by oversampling. (2005). *Atmel Application Note 8003A-AVR-09/05*.
10. Coats, M. R. (1991). New technology two-wire transmitters. *Sensors, 8*(1).
11. Johnson, J. B. (1928). Thermal agitation of electricity in conductors. *Physical Review, 32*, 97–109.
12. Rich, A. (1991). Shielding and guarding. In: *Best of analog dialogue*. ©Analog Devices.
13. Ott, H. W. (1976). *Noise reduction techniques in electronic systems*. New York: John Wiley & Sons.
14. Pascoe, G. (1977, February 6). The choice of solders for high-gain devices. *New Electronics* (U.K.).
15. Jensen, M. (2010). White Paper SWRA349. *Texas Instruments*.

16. Powers, R. A. (1995). Batteries for low power electronics. *Proceedings of the IEEE, 83*(4), 687–693.
17. Batra, A., et al. (2011). Simulation of energy harvesting from roads via pyroelectricity. *Journal of Photonics for Energy, 1*(1), 014001.
18. Le, T. T., et al. (2009). RF energy harvesting circuit. *U.S. Patent publ. No. 2009/0152954*.
19. Mickle, M. H., et al. (2006). Energy harvesting circuit. *U.S. Patent No. 7084605*.
20. Butler, P. (2012). Harvesting power in near field communication (NFC) device. *U.S. Patent No. 8326224*.
21. Safak, M. (2014). Wireless sensor and communication nodes with energy harvesting. *Journal of Communication, Navigation, Sensing and Services, 1*, 47–66.