Measurements & Instrumentation

1 INTRODUCTION

Measurement defined as branch of engineering that deals with measuring devices that used to determine various parameters of a system or a process. It is the process of obtaining a quantitative comparison between a predefined standard and measured one (i.e., the measured quantity). International Organization of Standardization (ISO), American National Standards Institute (ANSI), National Bureau of Standards (NBS), Bureau of Indian Standards (BIS) and other such organizations, has set the standards. Hence, measurement also is define as ascertainment of extent of certain quantity by comparison with a standard. In addition, it also provides the information of condition of a system in the form of visual indication, monitoring or suitable transmission according to the needs and requirements of the system. The application of measurement systems can be classify to:

1- Monitoring of processes and operations
2- Control of processes and operations
3- Experimental engineering analysis

1.1 Monitoring op processes and operations

Example of measurement instruments is wristwatch, an instrument for measuring time. The fuel level meter, an instrument to gives information about the level of fuel in vehicle. Other examples are thermometer, barometer, anemometer … etc.

All the mentioned devices could be use just for monitoring and they do not have any control functions.

1.2 Control of processes and operations

In another extremely important type of application for measuring instruments, the instrument serves as a component of an automatic control system. Clearly, to control any variable using feedback scheme, it is necessary to measure it; thus all closed loop control systems must incorporate at least one measurement instrument.

A familiar example is the typical home heating system employing some type of thermostatic control. For modern manufacturing and processing industry, the use of measuring instruments is imperative. For example, in thermal power stations, it is critical to keep track of the pressure built up in a steam boiler using pressure-measuring devices. The temperature of the furnace has to be measure by using some non-contact method. Measuring instruments form an integral part of automatic control system. For example, in a computerized numerically controlled (CNC) machine the displacement of slides measured and compared with the desired value continuously.

2 GENERALISED MEASURING SYSTEM AND FUNCTIONAL ELEMENTS

2.1 Experimental engineering analysis

In solving engineering problems, two general methods are available: theoretical and experimental. Many problems require the application of both methods. The relative amount of each depends on the nature of the problem. Problems on the frontiers of knowledge often require very extensive experimental studies since adequate theories are not available yet. Thus theory and experiment should thought of as complementing each other, and the engineer who takes this

Dr. Mohanned AL-Khafaji

attitude will, in general, be a more effective problem solver than one who neglects one or the other of these two approaches.

There are three categories of elements present in a generalized measuring instrument as shown in Figure 1. These are:

1. Initial sensing elements.
2. Signal conditioning elements.
3. Reading-recording elements.

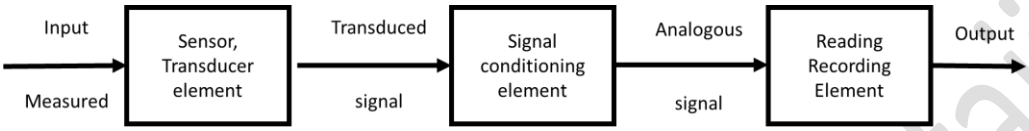


Figure 1 Signal flow in generalized measuring system

2.1.1 Initial sensing element

It is also called sensor-transducer element or simply transducer element. It is the first element, which detects or senses the measured. It is the part, which first receives energy form the measured medium and converts this input into a more practically convenient form of output. This output is the function of the measured quantity.

Sensor

A sensor is a device that detects a change in an analogue quantity, which is to be measured, and turns into another physical quantity. This one in turns is converted usually into current or voltage by a transducer.

We can distinguish parametric and self-generating sensors. In case of the former, change of the measured quantity is followed by a change of a parameter of electric circuit, for example resistance, capacitance, self-inductance or mutual inductance. In case of self-generating sensors, a measured quantity is usually changed directly into voltage, current or electric charge.

There are also coding sensors. Their digitized output goes directly towards the digital channel of a measuring system.

Transducers

Devices used to transform one kind of energy to another. When a transducer converts a measurable quantity (sound pressure level, optical intensity, magnetic field, etc) to an electrical voltage or an electrical current we call it a **sensor.** We will see a few examples of sensors shortly.

The type of a transducer depends on the kind of the output signal from a sensor. Most often bridge circuits or half-bridge circuits are applied for this purpose. They operate in connection with parametric sensors, for example strain gauges that are used for the measurement of dynamically changing strain. Other types of transducer measuring circuits are applied in connection with capacitive and inductive transducers that are used for a measurement of pressure difference and linear displacement.

Example of transducer is the solar cell, which converts the light to voltage.

2.1.2 Signal conditioning elements

Signal conditioning elements are used to modify the transduced information into a form that is acceptable to the reading-recording element.

2.1.3 Reading-recording element

It is also called data presentation element. The data in finally transmitted to data presentation element. This element is used to display information of measured quantity to the observer. The most common form of data presentation element consists of pointer moving over a scale to give the reading of measured quantity. Digital display of measured reading is also very common.

Sometimes data being read is also safely stored in the storing device like random access memory (RAM), hard disk, magnetic tape or floppy etc. Hence, these elements act as data storage element. The desired data can retrieved repeatedly whenever required.

Example:

An example to the measurement system is the Digital Thermometer.

Figure 2 shows an example of a measurement system. The thermocouple is a transducer that converts temperature to a small voltage; the amplifier increases the magnitude of the voltage; the A/D (analog-to-digital) converter is a device that changes the analog signal to a coded digital signal; and the LEDs (light emitting diodes) display the value of the temperature.

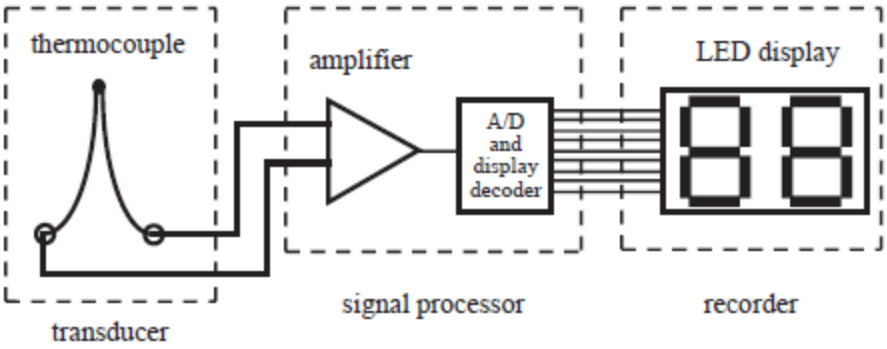


Figure 2 Digital thermometer

3 STRESS AND STRAIN MEASUREMENT

Measurement of stress in a mechanical component is important when assessing whether or not the component is subjected to safe load levels. Stress and strain measurements can also be used to indirectly measure other physical quantities such as force (by measuring strain of a flexural element), pressure (by measuring strain in a flexible diaphragm), and temperature (by measuring thermal expansion of a material). The most common transducer used to measure strain is the electrical resistance strain gage. As we will see, stress values can be determined from strain measurements using principles of solid mechanics.

3.1 Electrical Resistance Strain Gage

The most common transducer for experimentally measuring strain in a mechanical component is the bonded metal foil **strain gage** illustrated in Figure 3 and Figure 4. It consists of a thin foil of metal, usually constantan, deposited as a grid pattern onto a thin plastic backing material, usually

polyimide. The foil pattern is terminated at both ends with large metallic pads that allow leadwires to be easily attached with solder. The entire gage is usually very small, typically 5 to 15 mm long.

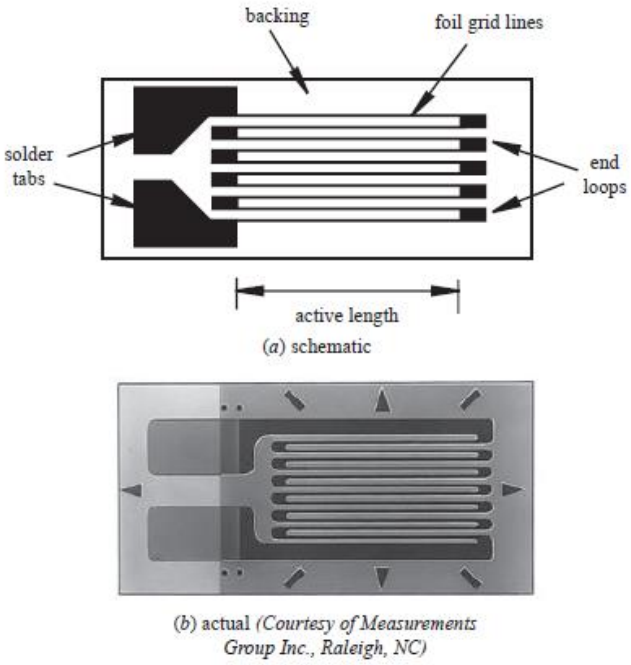


Figure 3 Metal foil strain gage construction

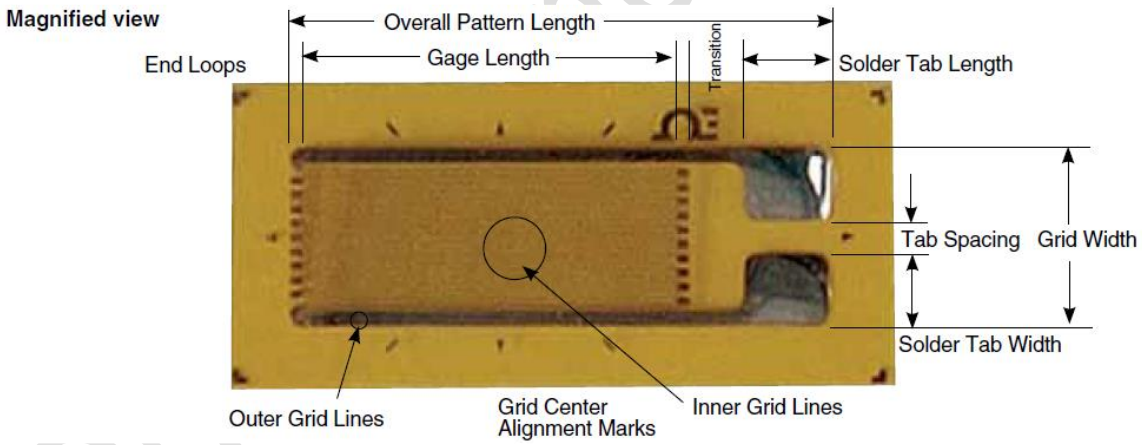


Figure 4 Magnified view of strain gauge

To measure strain on the surface of a machine component or structural member, the gage is adhesively bonded directly to the component, usually with epoxy or cyanoacrylate. The backing makes the foil gage easy to handle and provides a good bonding surface that also electrically insulates the metal foil from the component. Leadwires are then soldered to the solder tabs on the gage. When the component is loaded, the metal foil deforms, and the resistance changes in a predictable way (see below). If this resistance change is measured accurately, the strain on the surface of the component can be determined. Strain measurements allow us to determine the state of stress on the surface of the component, where stresses typically have their highest values. Knowing stresses at critical locations on a component under load can help a designer validate

analytical or numerical results (e.g., from a finite element analysis) and verify that stress levels remain below safe limits for the material (e.g., below the yield strength). It is important to note that, because strain gages are finite in size, a measurement actually reflects an average of the strain over a small area. Hence, making measurements where stress gradients are large (e.g., where there is stress concentration) can yield poor results.

Strain gauge sensors are the fundamental sensing elements for many types of sensors e.g. force sensors, torque sensors, pressure sensors, acceleration sensors, etc. They are applied to measure strain. Having strain measured and using Young's modulus *E* and geometric sizes, stress can be calculated. Finally, from these calculations an unknown and investigated quantity can be found, which is applied and acts on an object under test. Strain gauge principal of operation takes advantage of the physical property of the variety of changes of electrical resistance resulting from its elongation or shortening. If a strip of conductive material is stretched, it becomes skinnier and longer resulting in an increase of its resistance *R*, while if it is compressed, it becomes shorten and broaden resulting in decrease of its resistance. The principle of operation of a common metallic strain gauges is based on a change of a conductor resistance.

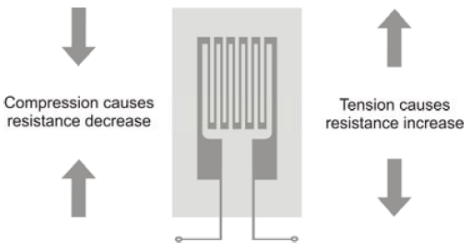


Figure 5 strain guage under loading

To understand how a strain gage is used to measure strain, we first look at how the resistance of the foil changes when deformed. The metal foil grid lines in the active portion of the gage (see Figure 3a) can be approximated by a single rectangular conductor as illustrated in Figure 6, whose total resistance is given by

$$R = \frac{\rho L}{A}$$

Eq. 1

Where ρ is the foil metal resistivity, L is the total length of the grid lines, and A is the grid line cross-sectional area. The gage end loops and solder tabs have negligible effects on the gage resistance because they typically have a much larger cross section than the foil lines.

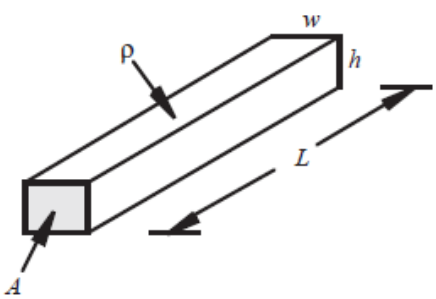


Figure 6 Rectangular conductor

Dr. Mohanned Khafaji

To see how the resistance changes under deformation, we need to take the differential of Eq. 1. If we first take the natural logarithm of both sides,

$$ln\, R = ln\, \rho + ln\, L - ln\, A$$

Eq. 2

Taking the differential yields the following expression for the change in resistance given material property and geometry changes in the conductor:

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dL}{L} - \frac{dA}{A}$$

Eq. 3

As we would expect, the signs in this equation imply that the resistance of the conductor increases ($d\,R > 0$) with increased resistivity and increased length and decreases with increased cross-sectional area. Because the cross-sectional area of the conductor is

$$A = wh$$

Eq. 4

The area differential term is

$$\frac{dA}{A} = \frac{w.dh + h.dw}{w.h} = \frac{dh}{h} + \frac{dw}{w}$$

Eq. 5

From the definition of Poisson's ratio

$$\frac{dh}{h} = -v\frac{dL}{L}$$

Eq. 6

And

$$\frac{dw}{w} = -v\frac{dL}{L}$$

Eq. 7

So

$$\frac{dA}{A} = -2v\frac{dL}{L} = -2v\,\varepsilon_{axial}$$

Eq. 8

Where ε_{axial} is the axial strain in the conductor. When the conductor is elongated ($\varepsilon_{axial} > 0$), the cross-sectional area decreases ($\frac{dA}{A} < 0$), causing the resistance to increase. Using Eq. 8 and Eq. 3 can be expressed as

$$\frac{dR}{R} = \varepsilon_{axial}(1 + 2v) + \frac{d\rho}{\rho}$$

Eq. 9

Dividing through by ε_{axial} gives

Dr. Mohanned AL-Khafaji

$$\frac{dR/R}{\varepsilon_{axial}} = 1 + 2v + \frac{d\rho/\rho}{\varepsilon_{axial}}$$

Eq. 10

The first two terms on the right-hand side, 1 and $2v$, represent the change in resistance due to increased length and decreased area. The last term $\frac{d\rho/\rho}{\varepsilon_{axial}}$ represents the **piezoresistive effect** in the material, which explains how the resistivity of the material changes with strain. All three terms are approximately constant over the operating range of typical strain gage metal foils.

Commercially available strain gage specifications usually report a constant **gage factor F** to represent the right-hand side of Eq. 10. This factor represents the material characteristics of the gage that relate the gage's change in resistance to strain:

$$\frac{dR/R}{\varepsilon_{axial}} = F$$

Eq. 11

Thus, when a gage of known resistance R and gage factor F is bonded to the surface of a component and the component is then loaded, we can determine the strain in the gage ε_{axial} simply by measuring the change in resistance of the gage ΔR:

$$\varepsilon_{axial} = \frac{dR/R}{F}$$

Eq. 12

This gage strain is the strain on the surface of the loaded component in the direction of the gage's long dimension.

Strain gauges are manufactured to various nominal values of resistance, of which 120Ω, 350Ω and 1000Ω are very common. The typical maximum change of resistance in a 120Ω device would be 5Ω at maximum deflection.

Example:

If a $120\ \Omega$ strain gage with gage factor 2.0 is used to measure a strain of $100\ \mu\varepsilon$ how much does the resistance of the gage change from the unloaded state to the loaded state?

Solution:

$$\Delta R = R.F.\varepsilon$$

Therefore, the change in resistance would be

$$\Delta R = 120\ \times 2 \times 100^{-6} = 0.024$$

3.2 Measuring Resistance Changes with a Wheatstone Bridge

To use strain gages to accurately measure strains experimentally, we need to be able to accurately measure small changes in resistance. The most common circuit used to measure small changes in resistance is the Wheatstone bridge, which consists of a four-resistor network excited by a DC voltage. A Wheatstone bridge is better than a simple voltage divider because it can be easily

Dr. Mohanned AL-Khafaji

balanced to establish an accurate zero position, it allows temperature compensation, and it can provide better sensitivity and accuracy. There are two different modes of operation of a Wheatstone bridge circuit: the **static balanced** mode and the **dynamic unbalanced** mode. For the **static balanced mode,** illustrated in Figure 7, R_2 and R_3 are precision resistors, R_4 is a precision potentiometer (variable resistor) with an accurate scale for displaying the resistance value, and R_1 is the strain gage resistance for which the change is to be measured. To balance the bridge, the variable resistor is adjusted until the voltage between nodes A and B is zero. In the balanced state, the voltages at A and B must be equal so

$$i_1 R_1 = i_2 R_2$$

Eq. 13

Also, because the high-input impedance (Z) voltmeter (VM) between A and B is assumed to draw no current,

$$i_1 = i_4 = \frac{V_{ex}}{R_1 + R_4}$$

Eq. 14

and

$$i_2 = i_3 = \frac{V_{ex}}{R_2 + R_3}$$

Eq. 15

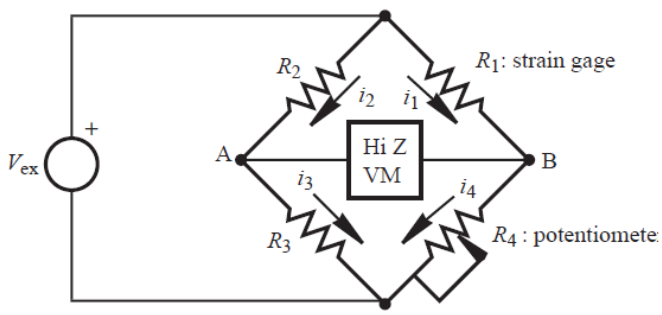


Figure 7 Static balanced bridge circuit.

Where V_{ex} is the DC voltage applied to the bridge called the **excitation voltage**. Substituting these expressions into Equation Eq. 14 and rearranging gives

$$\frac{R_1}{R_4} = \frac{R_2}{R_3}$$

Eq. 16

If we know R_2 and R_3 accurately and we note the value for R_4 on the precision potentiometer scale, we can accurately calculate the unknown resistance R_1 as

$$R_1 = \frac{R_1 R_4}{R_3}$$

Eq. 17

Note that this result is independent of the excitation voltage, V_{ex}.

Dr. Mohanned AL-Khafaji

The static balanced mode of operation can be used to measure a gage's resistance under fixed load, but usually balancing is done only as a preliminary step to measuring changes in gage resistance. In **dynamic deflection operation,** shown in Figure 8, again with R_1 representing a strain gage and R_4 representing a potentiometer, the bridge is first balanced, before loads are applied, by adjusting R_4 until there is no output voltage. Then changes in the strain gage resistance R_1 that occur under time-varying load can be determined from changes in the output voltage.

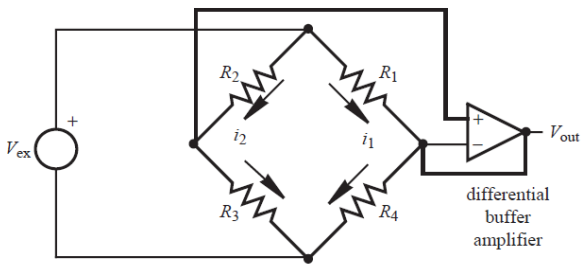


Figure 8 Dynamic unbalanced bridge circuit.

The output voltage can be expressed in terms of the resistor currents as

$$V_{out} = i_1R_1 - i_2R_2 = -i_1R_4 + i_2R_3$$

Eq. 18

and the excitation voltage can be related to the currents:

$$V_{ex} = i_1(R_1 + R_4) = i_2(R_2 + R_3)$$

Eq. 19

Solving for i_1 and i_2 in terms of V_{ex} in Eq. 19 and substituting these into the first expression in Eq. 18 gives

$$V_{out} = V_{ex}\left(\frac{R_1}{R_1 + R_4} - \frac{R_2}{R_2 + R_3}\right)$$

Eq. 20

When the bridge is balanced, V_{out} is zero and R_1 has a known value. When R_1changes value, as the strain gage is loaded, Eq. 20 can be used to relate this voltage change ΔV_{out} to the change in resistance R_1 . To find this relation, we can replace R_1 by its new resistance $R_1 + \Delta R_1$ and V_{out} by the output deflection voltage V_{out} . Then Eq. 20 gives

$$\frac{\Delta V_{out}}{V_{ex}} = \frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_4} - \frac{R_2}{R_2 + R_3}$$

Eq. 21

Rearranging this equation gives us the desired relation between the change in resistance and the measured output voltage:

$$\frac{\Delta R_1}{R_1} = \frac{\frac{R_4}{R_1}\left(\frac{\Delta V_{out}}{V_{ex}} + \frac{R_2}{R_2 + R_3}\right)}{\left(1 - \frac{\Delta V_{out}}{V_{ex}} + \frac{R_2}{R_2 + R_3}\right)}$$

Dr. Mohanned AL-Khafaji

Eq. 22

By measuring the change in the output voltage V_{out} , we can determine the gage resistance change ΔR_1 from Eq. 22 and can compute the gage strain from Eq. 12. The differential buffer amplifier shown in Figure 8 provides high input impedance (i.e., it does not load the bridge) and high gain for the small change in voltage due to the small change in resistance.

Figure 9 illustrates the effects of leadwires when using a strain gage located far from the bridge circuit. Figure 9a illustrates a two-wire connection from a strain gage to a bridge circuit. With this configuration, each of the leadwire resistances R'adds to the resistance of the strain gage branch of the bridge. The problem with this is that, if the leadwire temperature changes, it causes changes in the resistance of the bridge branch. This effect can be substantial if the leadwires are long and extend through environments where the temperature changes. Figure 9b illustrates a three-wire connection that solves this problem. With this configuration, equal leadwire resistances are added to adjacent branches in the bridge so the effects of changes in the leadwire resistances offset each other. The third leadwire is connected to the high-input impedance voltage measuring circuit, and its resistance has a negligible effect because it carries negligible current. The three wires are usually in the form of a small ribbon cable to ensure they experience the same temperature changes and to minimize electromagnetic interference due to inductive coupling.

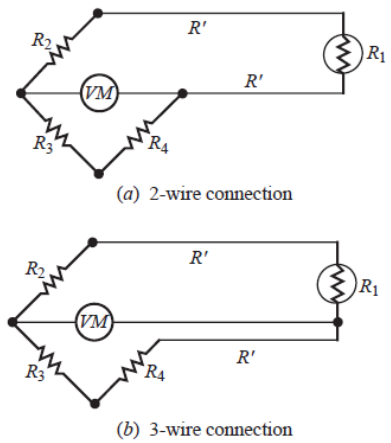


Figure 9 Leadwire effects in 1/4 bridge circuits.

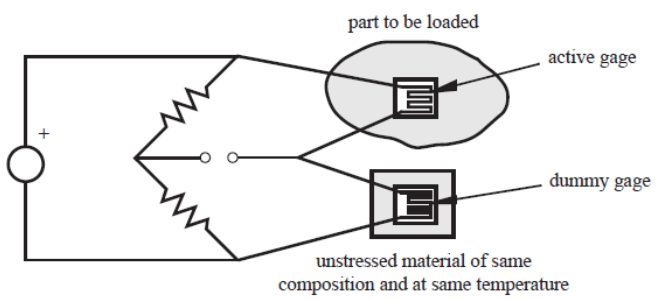


Figure 10 Temperature compensation with a dummy gage in a half bridge circuit.

In addition to temperature effects in leadwires, temperature changes in the strain gage can cause significant changes in resistance, which would lead to erroneous measurements. A convenient method for eliminating this effect is to use the circuit (called a half bridge) illustrated in Figure 10,

where two of the four bridge legs contain strain gages. The gage in the top branch is the active gage used to measure surface strains on a component to be loaded. The second "dummy" gage is mounted to an unloaded sample of material identical in composition to the component. If this sample is kept at the same temperature as the component by keeping it in close proximity, the resistance changes in the two gages due to temperature cancel because they are in adjacent branches of the bridge circuit. Therefore, the bridge generates an unbalanced voltage only in response to strain in the active gage.

3.3 Measuring Different States of Stress with Strain Gages

Mechanical components may have complex shapes and are often subjected to complex loading conditions. In these cases, it is difficult to predict the orientation of principal stresses at arbitrary points on the component. However, with some geometries and loading conditions, the principal axes are known, and measuring the state of stress is easier.

If a component is loaded uniaxially (i.e., loaded in only one direction in tension or compression), the state of stress in the component can be determined with a single gage mounted in the direction of the load. Figure 11 illustrates a bar in tension and the associated state of stress. By measuring the strain ε_x , the stress is obtained using Hooke's law:

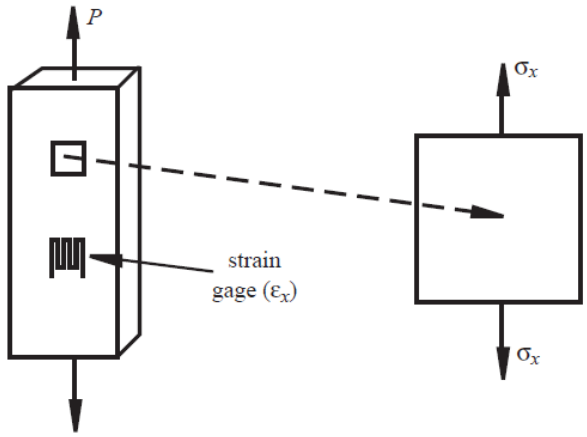


Figure 11 Bar under uniaxial stress.

$$\sigma_x = E\varepsilon_x$$

Eq. 23

Where the axial stress in the σ_x is given by

$$\sigma_x = \frac{P}{A}$$

Eq. 24

Where A is the bar's cross-sectional area. Therefore, the force P applied to the bar can be determined from the strain gage measurement:

$$P = AE\varepsilon_x$$

Eq. 25

If a component is known to be loaded biaxially (i.e., loaded in two orthogonal directions in tension or compression), the state of stress in the component can be determined with two gages aligned with the stress directions. Figure 12 illustrates a pressurized tank and the associated state of stress. By measuring the strains ε_x and ε_y , the stresses in the tank shell can be determined from Hooke's law generalized to two dimensions:

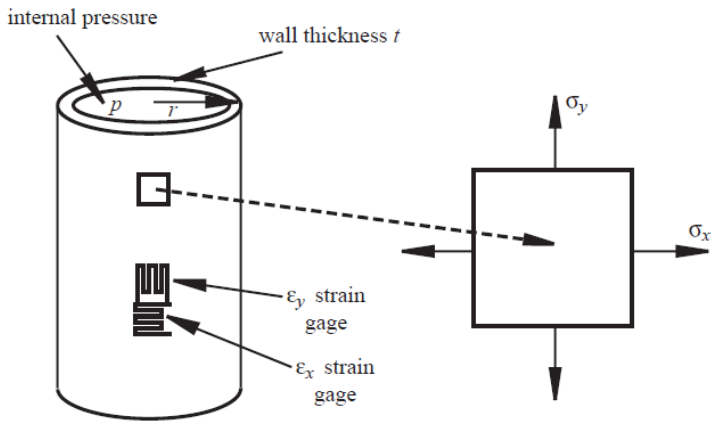


Figure 12 Biaxial stress in a long, thin-walled pressure vessel.

$$\varepsilon_x = \frac{\sigma_x}{E} - v\frac{\sigma_y}{E}$$

Eq. 26

$$\varepsilon_y = \frac{\sigma_y}{E} - v\frac{\sigma_x}{E}$$

Eq. 27

Solving for the stress components gives

$$\sigma_x = \frac{E}{1-v^2}(\varepsilon_x + v\varepsilon_y)$$

Eq. 28

$$\sigma_y = \frac{E}{1-v^2}(\varepsilon_y + v\varepsilon_x)$$

Eq. 29

For a thin-walled pressure vessel (i.e., $t / r < 1/10$), the stresses are approximated by

$$\sigma_x = \frac{pr}{t} \qquad \sigma_y = \frac{pr}{2t}$$

Eq. 30

Where *p* is the internal pressure, *t* is the wall thickness, and *r* is the radius of the vessel. The stress *x* is the transverse or hoop stress, and y is the axial or longitudinal stress. Either Eq. 28 or Eq. 29 can be used to compute the pressure in the vessel based on the strain gage measurements, yielding

$$p = \frac{t\sigma_x}{r} = \frac{tE}{r(1-v^2)}(\varepsilon_x + v\varepsilon_y)$$

Eq. 31

Dr. Mohanned AL-Khafaji

Or

$$p = \frac{2t\sigma_y}{r} = \frac{2tE}{r(1-v^2)}(\varepsilon_y + v\varepsilon_x)$$

Eq. 32

Either expression would yield the correct pressure value for an ideal thin-walled vessel and error-free measurements. In this example, the strain gages are serving as a pressure transducer.

For uniaxial and biaxial loading, we already know the directions of principal stresses in the component; hence, one or two gages needed, respectively, to determine the stress magnitudes. However, when the loading is more complex or when the geometry is more complex, which is often the case in mechanical design, we have to use three gages mounted in three different directions as illustrated in Figure 13. This assembly of strain gages is referred to as a **strain gage rosette**.

There is a wide variety of commercially available rosettes with two or more grid patterns accurately oriented on a single backing in close proximity. An assortment of rosettes and single-element gages illustrating the variety of shapes and sizes is shown in Figure 14.

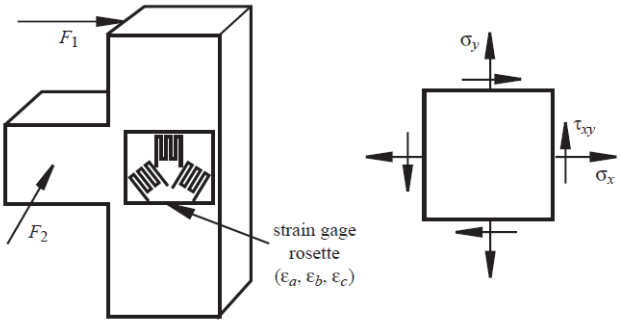


Figure 13 General state of planar stress on the surface of a component.

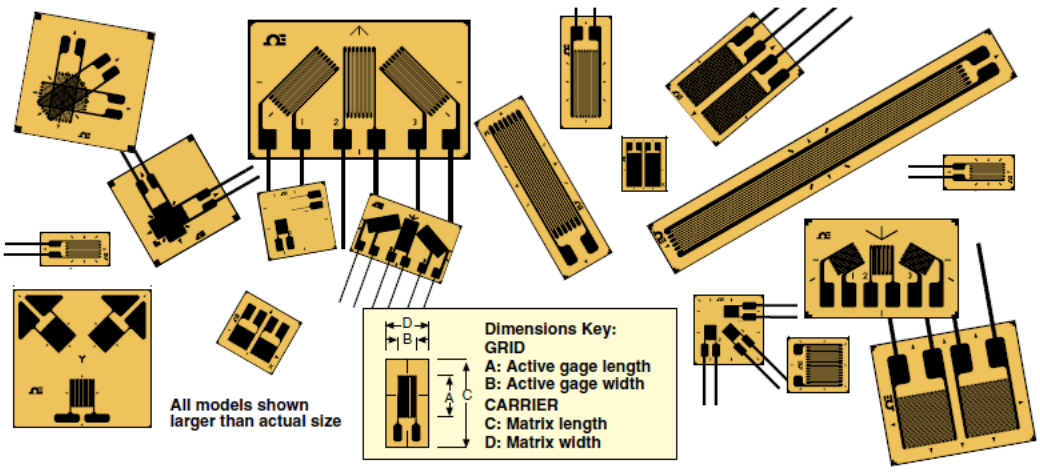


Figure 14 Assortment of different strain gage and rosette.

3.4 Force Measurement with Load Cells

A load cell is a sensor used to measure force. It contains an internal flexural element, usually with several strain gages mounted on its surface. The flexural element's shape is designed so that the

strain gage outputs can be easily related to the applied force. The load cell is usually connected to a bridge circuit to yield a voltage proportional to the load. Two commercial load cells, which are used to measure uniaxial force, are shown in Figure 15. An example of the application of load cells is in commercially available laboratory materials testing machines for measuring forces applied to a test specimen. Load cells are also used in weight scales, and they are sometimes included as integral parts of mechanical structures to monitor forces in the structures.

3.5 Torque Measurement

Torsional moment of the shaft can be measured directly by means of the appropriate location of strain gauges. The gauges are glued in along the main stress axes, where the strains have opposite signs.

Figure 16 shows how the strain gauges are glued to the surface for the measurement of torsional moment. For torque measurements of shafts, the following relations used

$$\varepsilon_1 = -\varepsilon_2 = \frac{8M}{\pi G D^3} = \frac{D}{4l}\alpha \quad and \;\; G = \frac{E}{2(1+v)}$$

Eq. 33

(a) Courtesy of MTS Systems Corp., Minneapolis, MN

(b) Courtesy of Transducer Techniques, Temecula, CA

Figure 15 Typical axial load cells

Dr. M. Al-Khafaji

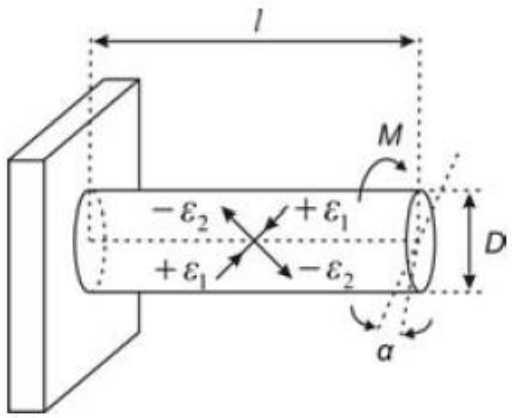


Figure 16 Location of strain gauges for the torque measurement

While for a tube

$$\varepsilon_1 = -\varepsilon_2 = \frac{8MD}{\pi G(D^4 - d^4)}$$

Eq. 34

Where G is shear modulus, v Poisson's ratio, M is the torque, α is the angle of shaft torsion, the shaft diameter is D and its length l, and d is the inside diameter of tube

4 TEMPERATURE MEASUREMENT

The temperature scales used to express temperature are:

1. Celsius (°C): Common SI unit of relative temperature.
2. Kelvin (°K): Standard SI unit of absolute thermodynamic temperature. Note the absence of the degree symbol.
3. Fahrenheit (°F): English system unit of relative temperature.

The relationships between these scales are summarized here:

$$T_c = T_K - 273.15$$

$$T_F = \frac{9}{5}T_c + 32$$

where T_C is temperature in degrees Celsius, T_K is temperature in Kelvin and T_F is temperature in degrees Fahrenheit.

4.1 Liquid-in-Glass Thermometer

A simple nonelectrical temperature-measuring device is the liquid-in-glass thermometer. It typically uses alcohol or mercury as the working fluid, which expands and contracts relative to the glass container. The upper range is usually on the order of 315.5°C (600°F). When making measurements in a liquid, the depth of immersion is important, as it can result in different measurements. Because readings are made visually, and there can be a meniscus at the top of the working fluid, measurements must be made carefully and consistently.

4.2 Bimetallic Strip

Another nonelectrical temperature-measuring device used in simple control systems is the bimetallic strip. As illustrated in Figure 17, it is composed of two or more metal layers having

Dr. Mohanned AL-Khafaji

different coefficients of thermal expansion. The strip can be straight, as shown in the figure, or coiled for a more compact design. Because these layers are permanently bonded together, the structure will deform when the temperature changes. This is due to the difference in the thermal expansions of the two metal layers. The deflection can be related to the temperature of the strip. Bimetallic strips are used in household and industrial thermostats where the mechanical motion of the strip makes or breaks an electrical contact to turn a heating or cooling system on or off.

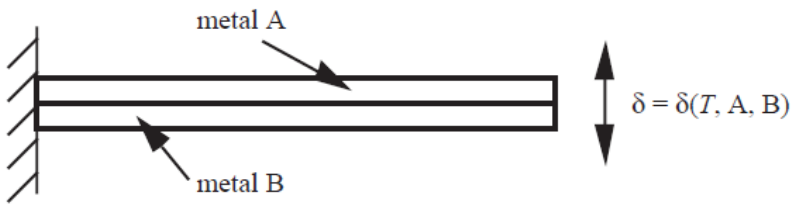


Figure 17 Bimetallic strip

4.3 Electrical Resistance Thermometer

A **resistance temperature device (RTD)** constructed of metallic wire wound around a ceramic or glass core and hermetically sealed. The resistance of the metallic wire increases with temperature. The resistance-temperature relationship usually approximated by the following linear expression:

$$R = R_0[1 - \alpha(T - T_0)]$$

Eq. 35

Where T_0 is a reference temperature, R_0 is the resistance at the reference temperature, and α is a calibration constant. The sensitivity ($\frac{dR}{dT}$) is $R_0\alpha$. The reference temperature is usually the ice point of water (0^oC).

The most common metal used in RTDs is **platinum** because of its high melting point, resistance to oxidation, predictable temperature characteristics, and stable calibration values. The operating range for a typical platinum RTD is -220° C to 750° C.

Lower cost nickel and copper types are also available, but they have narrower operating ranges

A **thermistor** is a semiconductor device, available in probes of different shapes and sizes, whose resistance changes exponentially with temperature. Its resistance temperature relationship expressed in the form

$$R = R_0 e^{\left[\beta\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]}$$

Eq. 36

Where T_0 is a reference temperature, R_0 is the resistance at the reference temperature, and β is a calibration constant called the **characteristic temperature** of the material. A well-calibrated thermistor can be accurate to within 0.01 °C or better, which is better than typical RTD accuracies. However, thermistors have much narrower operating ranges than RTDs.

Note in Eq. 36 that a thermistor's resistance actually decreases with increasing temperature. This is very different from metal conductors that experience increasing resistance with increasing temperature

Dr. Mohanned AL-Khafaji

4.4 Thermocouple

Two dissimilar metals in contact Figure 18 form a thermoelectric junction that produces a voltage proportional to the temperature of the junction. This known as the **Seebeck effect**.

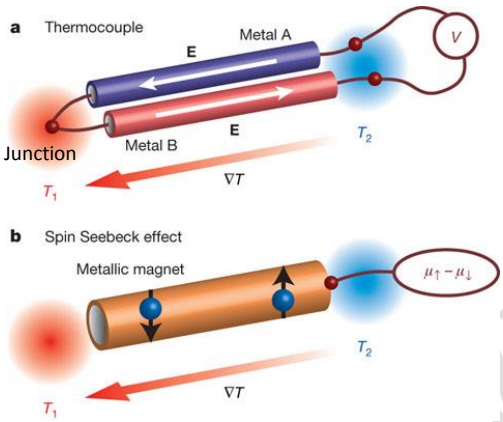


Figure 18 Thermoelectric junction.

Because an electrical circuit must form a closed loop, thermoelectric junctions occur in pairs, resulting in what called a thermocouple. We can represent a thermoelectric circuit containing two junctions as illustrated in Figure 19. Here we have wires of metals A and B forming junctions at different temperatures T_1 and T_2, resulting in a potential V that can be measured. The thermocouple voltage V depends on the metal properties of A and B and the difference between the junction temperatures T_1 and T_2. The thermocouple voltage is directly proportional to the junction temperature difference:

$$V = \alpha(T_1 - T_2)$$

Eq. 37

Where α is called the **Seebeck coefficient.**

The relationship between voltage and temperature difference is not exactly linear. However, over a small temperature range, α is nearly constant.

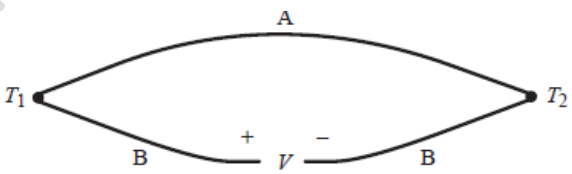


Figure 19 Thermocouple circuit

To properly design systems using thermocouples for temperature measurement, it is necessary to understand the basic laws that govern their application. The five basic laws of thermocouple behavior follow.

1. **Law of leadwire temperatures:**
   The thermoelectric voltage due to two junctions in a circuit consisting of two different conducting metals depends only on the junction temperatures T_1 and T_2. As illustrated in Figure 20, the temperature environment of the leads away from the junctions (T_3, T_4, T_5)

Dr. Mohanned AL-Khafaji

does not influence the measured voltage. Therefore, we need not be concerned about shielding the leadwires from environmental conditions.

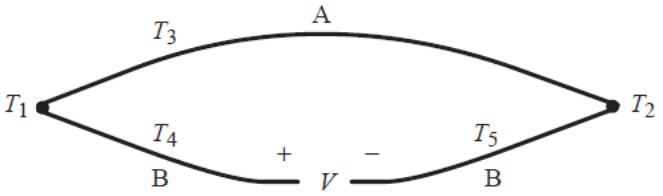


Figure 20 Law of leadwire temperatures

2. **Law of intermediate leadwire metals:**
As illustrated in Figure 21, a third metal C introduced in the circuit constituting the thermocouple has no influence on the resulting voltage as long as the temperatures of the two new junctions (A-C and C-A) are the same ($T_3 = T_4$). As a consequence of this law, a voltage measurement device that creates two new junctions can be inserted into the thermocouple circuit without altering the resulting voltage.

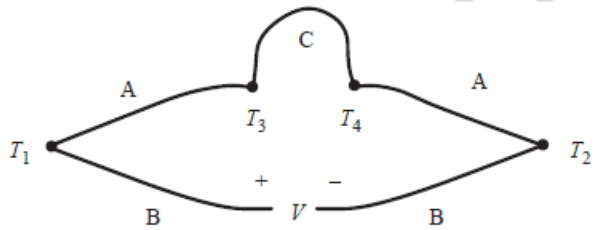


Figure 21 Law of intermediate leadwire metals.

3. **Law of intermediate junction metals:**
As illustrated in Figure 22, if a third metal is introduced within a junction creating two new junctions (A-C and C-B), the measured voltage will not be affected as long as the two new junctions are at the same temperature ($T_1 = T_3$). Therefore, although soldered or brazed joints introduce thermo junctions, they have no resulting effect on the measured voltage. If $T_1 \neq T_3$, the effective temperature at C is the average of the two temperatures($c = \frac{T_1+T_3}{2}$).

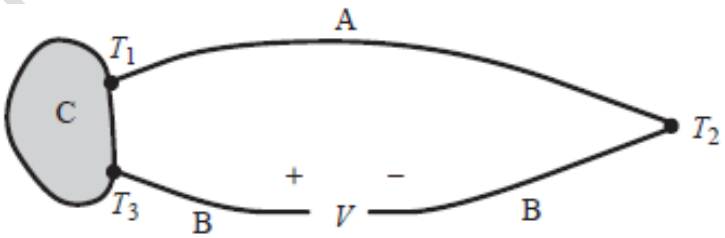


Figure 22 Law of intermediate junction metals.

4. **Law of intermediate temperatures:**
Junction pairs at T_1 and T_3 produce the same voltage as two sets of junction pairs spanning the same temperature range (T_1 to T_2 and T_2 to T_3); therefore, as illustrated in Figure 23,

Dr. Mohanned AL-Khafaji

$$V_{1/3} = V_{1/2} + V_{2/3}$$

Eq. 38

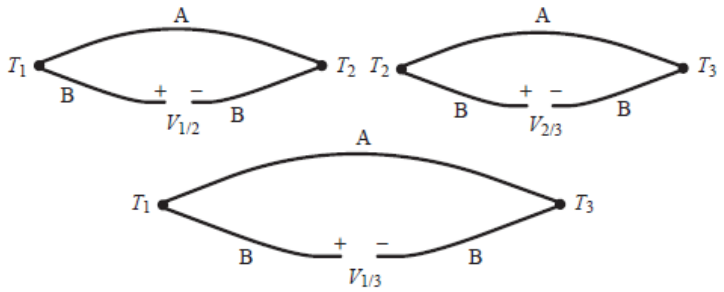


Figure 23 Law of intermediate temperatures.

This equation can be read: The voltage resulting from measuring temperature T_1 relative to T_3 is the same as the sum of the voltages resulting from T_1 relative to T_2 and T_2 relative to T_3. This result supports the use of a reference junction to allow accurate measurement of an unknown temperature based on a fixed reference temperature (described below).

5. **Law of intermediate metals:**
   As illustrated in Figure 24, the voltage produced by two metals A and B is the same as the sum of the voltages produced by each metal (A and B) relative to a third metal C:

$$V_{A/B} = V_{A/C} + V_{B/C}$$

Eq. 39

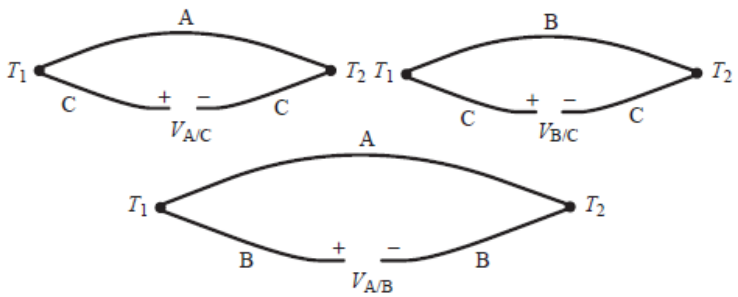


Figure 24 Law of intermediate metals

This result supports the use of a standard reference metal (e.g., platinum) to be used as a basis to calibrate all other metals.

A standard configuration for thermocouple measurements is shown in Figure 25. It consists of wires of two metals, A and B, attached to a voltage-measuring device with terminals made of metal C. The reference junction is used to establish a temperature reference for one of the junctions to ensure accurate temperature measurements at the other junction relative to the reference. A convenient reference temperature is 0° C, because this temperature can be accurately established and maintained with a distilled water ice bath (i.e., an ice-water mixture). If the terminals of the voltage-measuring device are at the same temperature, the law of intermediate leadwire metals ensures that the measuring device terminal metal C has no effect on the measurement. For a given pair of thermocouple metals and a reference temperature, a standard reference table can be compiled for converting voltage measurements to temperatures.

An important alternative to using an ice bath is a semiconductor reference (e.g., a **thermistor**), which electrically establishes the reference temperature based on solid state physics principles. These reference devices are usually included in thermocouple instrumentation to eliminate the need for an external reference temperature.

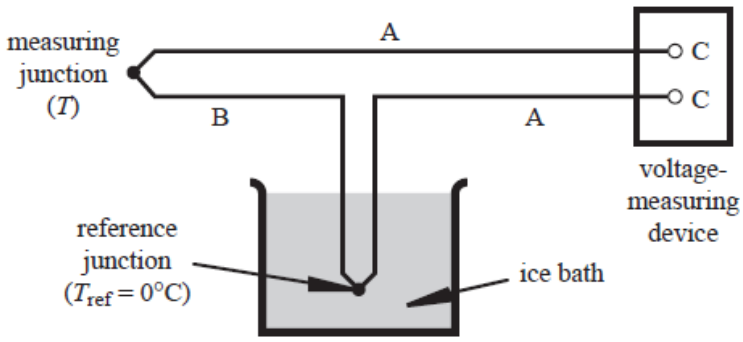


Figure 25 Standard thermocouple configuration

Example:

A standard two-junction thermocouple configuration is being used to measure the temperature in a wind tunnel. The reference junction is held at a constant temperature of 10° C. The thermocouple table referenced to 0° C is given. A portion of the table follows. Determine the output voltage when the measuring junction is exposed to an air temperature of 100° C.

Junction Temperature (°C)	Output Voltage (mV)
0	0
10	0.507
20	1.019
30	1.536
40	2.058
50	2.585
60	3.115
70	3.649
80	4.186
90	4.725
100	5.268

Solution:

By applying the law of intermediate temperature for this example, we can write

$$V_{100/0} = V_{100/10} + V_{10/0}$$

The required voltage is $V_{100/10}$, the voltage measured for a temperature of 100° C relative to a reference junction at 10° C. We can get the other voltages in the equation, $V_{100/0}$ and $V_{10/0}$, from the table because both are referenced to 0° C. Therefore,

$$V_{100/10} = V_{100/0} - V_{\frac{10}{0}} = 5.268 - 0.507 = 4.761\ mV$$

The letters E, J, K, R, S, and T denote the six most commonly used thermocouple metal pairs. The 0° C reference junction calibration for each of the types is nonlinear and can be approximated with a polynomial. The metals in the junction pair, the thermoelectric polarity, the commonly used color

Dr. Mohanned AL-Khafaji

code, the operating range, the accuracy, and the polynomial order and coefficients are shown for each type in Table 1. The general form for the polynomial using the coefficients in the table is

$$T = \sum_{i=0}^{9} c_i V^i = c_0 + c_1 V + c_2 V^2 + c_3 V^3 + c_4 V^4 + c_5 V^5 + c_6 V^6 + c_7 V^7 + c_8 V^8 + c_9 V^9$$

Eq. 40

Where V is the thermoelectric voltage measured in volts and T is the measuring junction temperature in °C, assuming a 0° C reference junction. Figure 26 shows the sensitivity curves for some commercially available thermocouple pairs. Even though we use a ninth-order polynomial to represent the temperature voltage relation, providing an extremely close fit, the relationship is close to linear as predicted by the Seebeck effect.

Table 1 Thermocouple data

	Type E	Type J	Type K	Type R	Type S	Type T
Metal pair	Chromel (+) and constantan (−)	Iron (+) and constantan (−)	Chromel (+) and alumel (−)	87% platinum, 13% rhodium (+) and platinum (−)	90% platinum, 10% rhodium (+) and platinum (−)	Copper (+) and constantan (−)
Color code	Purple	Black	Yellow	Green	Green	Blue
Operating range	−100°C to 1000°C	0°C to 760°C	0°C to 1370°C	0°C to 1000°C	0°C to 1750°C	−160°C to 400°C
Accuracy	±0.5°C	±0.1°C	±0.7°C	±0.5°C	±0.1°C	±0.5°C
Approximate sensitivity (mV/°C)	0.079	0.054	0.042	0.012	0.011	0.049
Polynomial order	9	5	8	8	9	7
c_0	0.104967	−0.0488683	0.226585	0.263633	0.927763	0.100861
c_1	17,189.5	19,873.1	24,152.1	179,075.	169,527.	25,727.9
c_2	−282,639.	−218,615.	67,233.4	-4.88403×10^7	-3.15684×10^7	−767,346.
c_3	1.26953×10^7	1.15692×10^7	2.21034×10^6	1.90002×10^{10}	8.99073×10^9	7.80256×10^7
c_4	-4.48703×10^8	-2.64918×10^8	-8.60964×10^8	-4.82704×10^{12}	-1.63565×10^{12}	-9.24749×10^9
c_5	1.10866×10^{10}	2.01844×10^9	4.83506×10^{10}	7.62091×10^{14}	1.88027×10^{14}	6.97688×10^{11}
c_6	-1.76807×10^{11}	—	-1.18452×10^{12}	-7.20026×10^{16}	-1.37241×10^{16}	-2.66192×10^{13}
c_7	1.71842×10^{12}	—	1.38690×10^{13}	3.71496×10^{18}	6.17501×10^{17}	3.94078×10^{14}
c_8	-9.19278×10^{12}	—	-6.33708×10^{13}	-8.03104×10^{19}	-1.56105×10^{19}	—
c_9	2.06132×10^{13}	—	—	—	1.69535×10^{20}	—

Dr. Mohanned Al-Khafaji

ANSI DESIGNATION	ALLOY (Generic or Trade Names)
JP	Iron
JN, EN, or TN	Constantan, Cupron, Advance
KP or EP	Chromega, Tophel, T_1, Thermokanthal KP
KN	Alomega, Nial, T_2, Thermokanthal KN
TP	Copper
RN or SN	Pure Platinum
RP	Platinum 13% Rhodium
SP	Platinum 10% Rhodium

Trade Names: Advance T—Driver Harris Co., Chromega and Alomega—OMEGA Engineering, Inc., Cupron, Nial and Tophel—Wilbur B. Driver Co., Thermokanthal KP and Thermokanthal KN—The Kanthal Corporation.

ANSI LETTER DESIGNATIONS—Currently thermocouple and extension wire is ordered and specified by an ANSI letter designation. Popular generic and trade name examples are Chromega/Alomega—ANSI Type K: Iron/Constantan—ANSI Type J: Copper/Constantan—ANSI Type T: Chromega/Constantan—ANSI Type E: Platinum/Platinum 10% Rhodium—ANSI Type S: Platinum/Platinum 13% Rhodium—ANSI Type R. The positive and negative legs are identified by letter suffixes P and N, respectively, as listed in the tables.

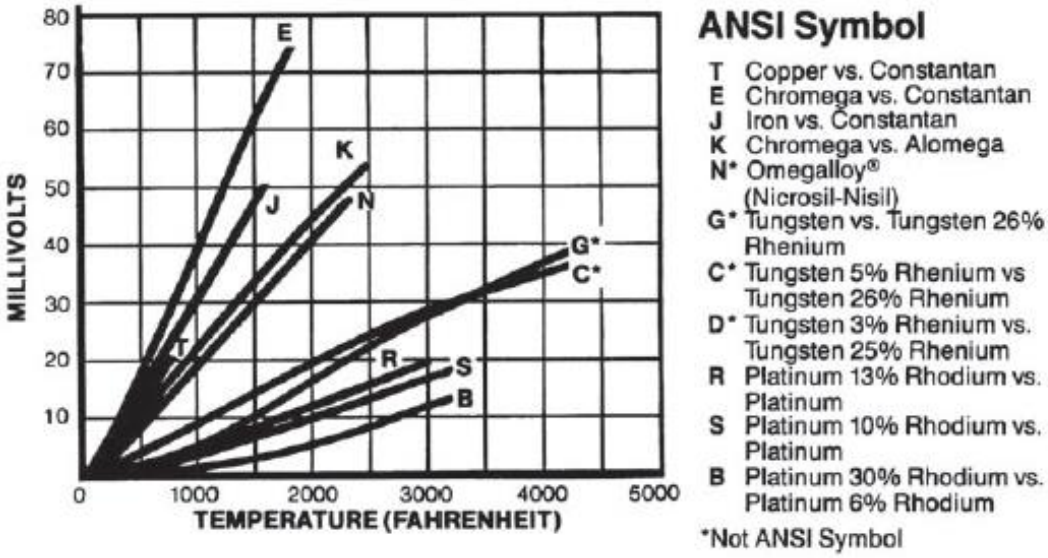


Figure 26 Thermocouple types and characteristics.

5 POSITION AND SPEED MEASUREMENT

5.1 Proximity Sensors and Switches

A proximity sensor consists of an element that changes either its state or an analog signal when it is close to, but often not actually touching, an object. Magnetic, electrical capacitance, inductance, and eddy current methods are particularly suited to the design of a proximity sensor. A **photoemitter-detector** pair represents another approach, where interruption or reflection of a beam of light is used to detect an object in a noncontact manner. The emitter can be a laser or focused LED, and the detector is usually a phototransistor or photodiode. Various configurations for photoemitter-detector pairs are illustrated in Figure 27. In the opposed and retroreflective configurations, the object interrupts the beam; and in the proximity configuration, the object reflects the beam. Figure 28 shows a commercial sensor that can be used in the retro reflective or proximity configurations. Common applications for proximity sensors and limit switches are detecting the presence of an object.

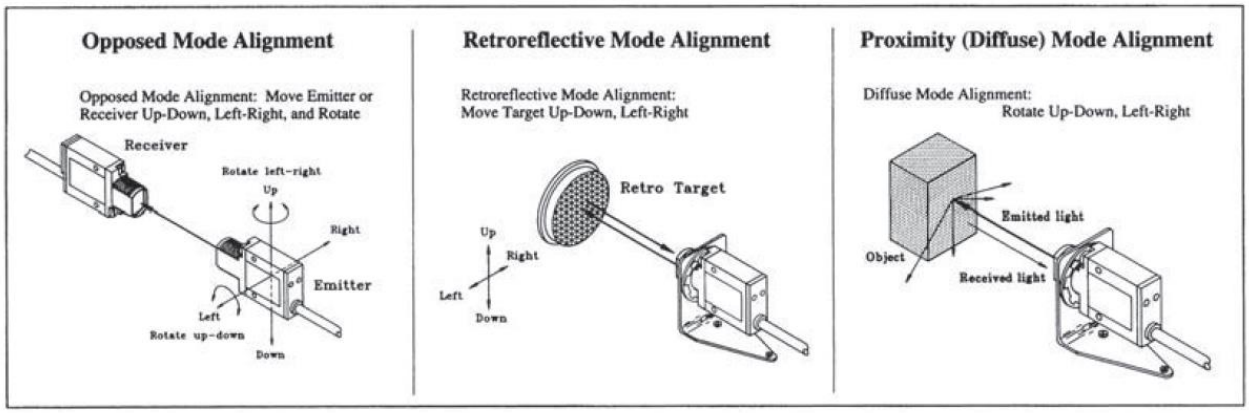


Figure 27 various configurations for photoemitter-detector pairs

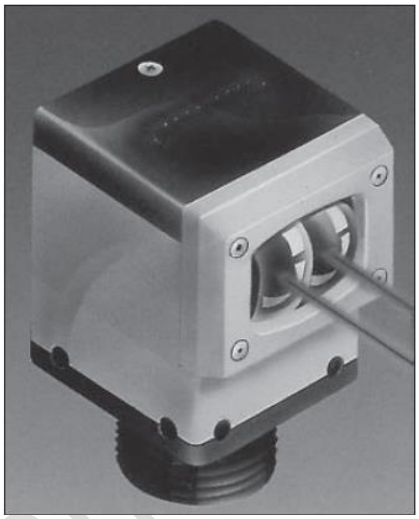


Figure 28 Example of a photoemitter-detector

There are many designs for **limit switches**, including pushbutton and levered **microswitches**. All switches are used to open or close connections within circuits. As illustrated in Figure 29, switches are characterized by the number of poles (P) and throws (T) and whether connections are **normally open** (**NO**) or **normally closed** (**NC**)

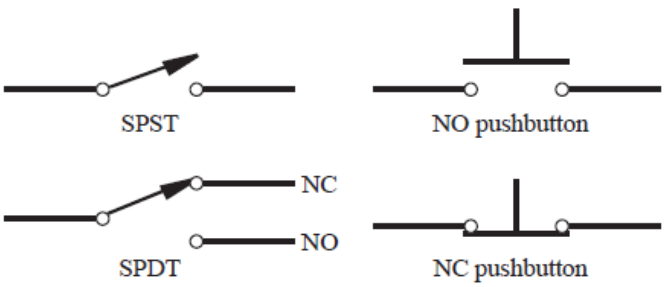


Figure 29 Switches

A **pole** is a moving element in the switch that makes or breaks connections, and a **throw** is a contact point for a pole. The SPST switch is a single-pole (SP), single-throw (ST) device that opens or loses a single connection. The SPDT switch changes the pole between two different throw positions.

There are many variations on the pole and throw configurations of switches, but their function is easily understood from the basic terminology. Figure 30 show various types of switches with the appropriate designations.

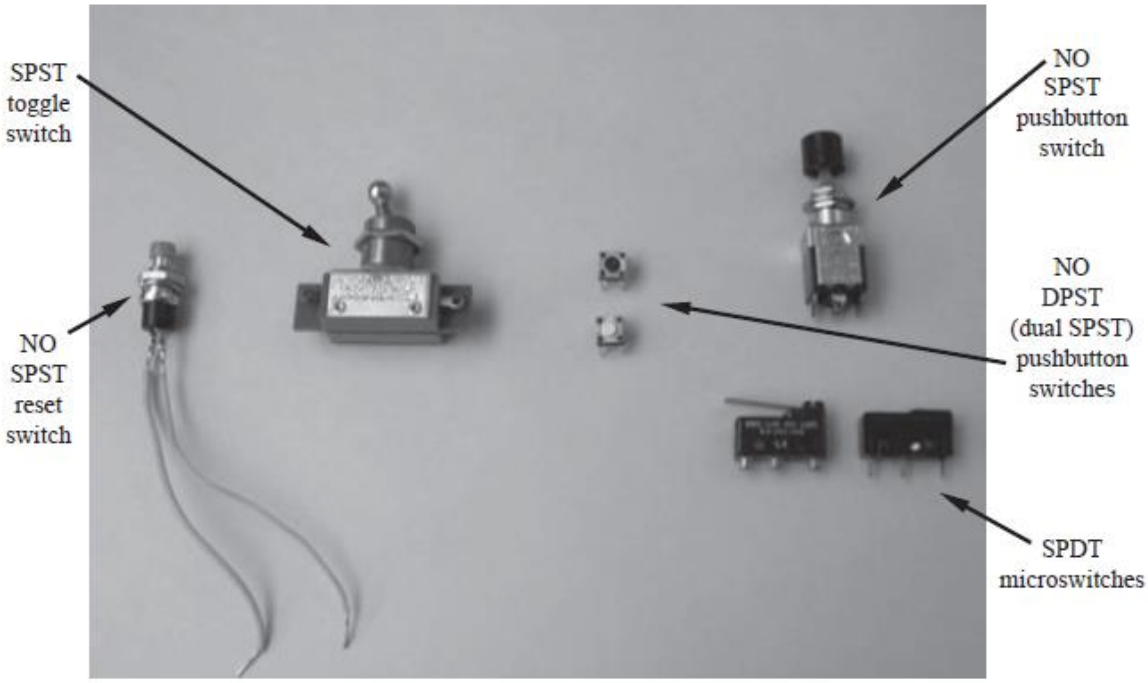


Figure 30 Photograph of various types of switches.

5.2 Potentiometer

The rotary **potentiometer** (aka **pot**) is a variable resistance device that can be used to measure angular position. It consists of a wiper that makes contact with a resistive element, and as this point of contact moves, the resistance between the wiper and end leads of the device changes in proportion to the angular displacement. Figure 31 illustrates the form and internal schematic for a typical rotary potentiometer. Figure 32 shows two common types of potentiometers. The one on the left is called a trim pot. It has a small screw on the left side that can be turned with a screwdriver to accurately make small changes in resistance (i.e. “trim” or adjust the resistance). On the right is a standard rotary pot with a knob allowing a user to easily make adjustments by hand. Through voltage division, the change in resistance of a pot can be used to create an output voltage that is directly proportional to the input displacement.

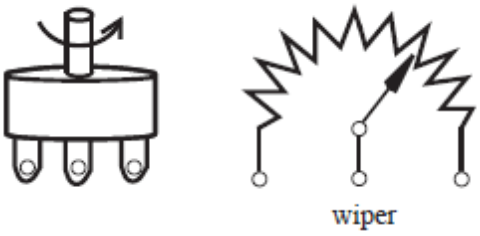


Figure 31 Potentiometer.

Dr. Mohanned

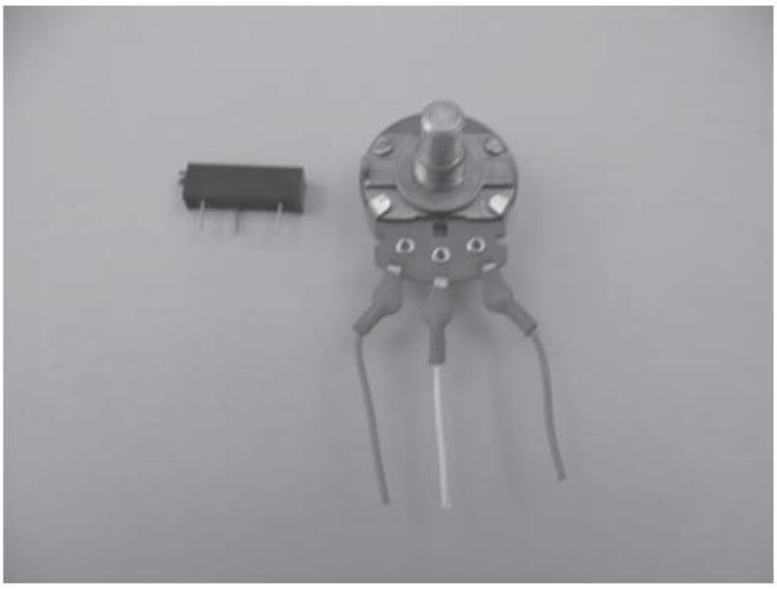


Figure 32 Photograph of a trim pot and a rotary pot.

5.3 Linear Variable Differential Transformer

The **linear variable differential transformer** (**LVDT**) is a transducer for measuring linear displacement. As illustrated in Figure 33, it consists of primary and secondary windings and a movable iron core. It functions much like a transformer, where voltages are induced in the secondary coil in response to excitation in the primary coil. The LVDT must be excited by an AC signal to induce an AC response in the secondary. The core position can be determined by measuring the secondary response.

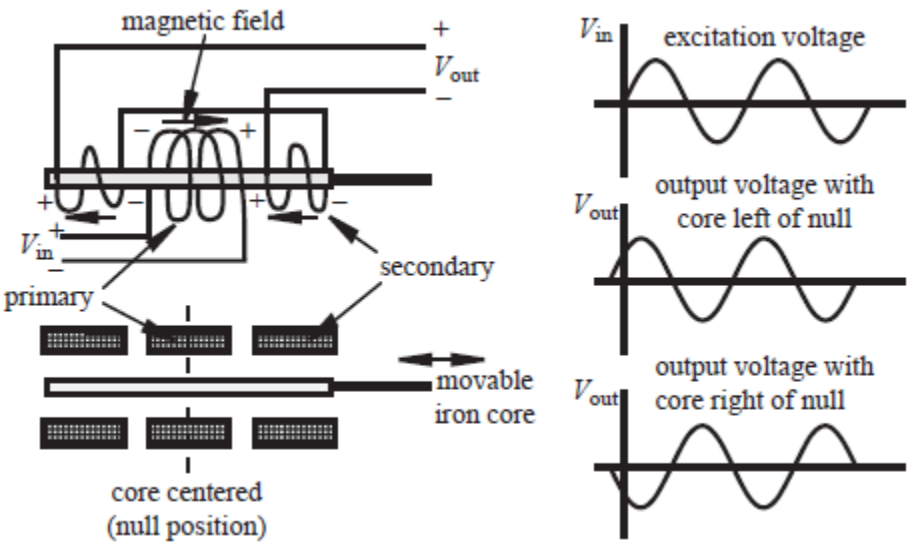


Figure 33 Linear variable differential transformer.

With two secondary coils connected in the series-opposing configuration as shown, the output signal describes both the magnitude and direction of the core motion. The primary AC excitation V_{in} and the output signal V_{out} for two different core positions are shown in Figure 33. There is a midpoint in the core's position where the voltage induced in each coil is of the same amplitude and

180° out of phase, producing a "null" output. As the core moves from the null position, the output amplitude increases a proportional amount over a linear range around the null as shown in Figure 34 . Therefore, by measuring the output voltage amplitude, we can easily and accurately determine the magnitude of the core displacement.

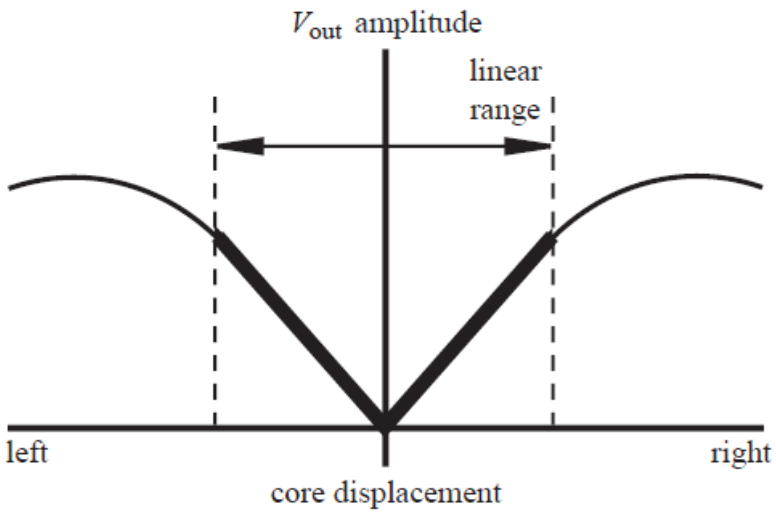


Figure 34 LVDT linear range.

To determine the direction of the core displacement, the secondary coils can be connected to a demodulation circuit as shown in Figure 35. The diode bridges in this circuit produce a positive or negative rectified sine wave, depending on which side of the null position the core is located.

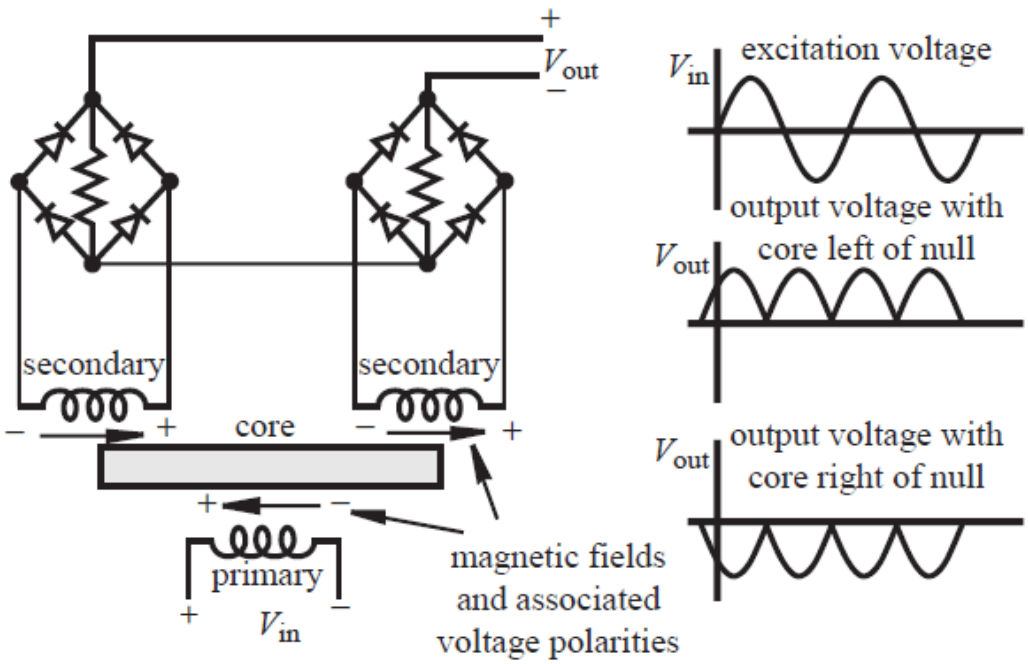


Figure 35 LVDT demodulation.

As illustrated in Figure 36, a low-pass filter used to convert the rectified output into a smoothed signal that tracks the core position. The cutoff frequency of this low-pass filter must be chosen carefully to filter out the high frequencies in the rectified wave but not the frequency components

associated with the core motion. The excitation frequency is usually chosen to be at least 10 times the maximum expected frequency of the core motion to yield a good representation of the time-varying displacement.

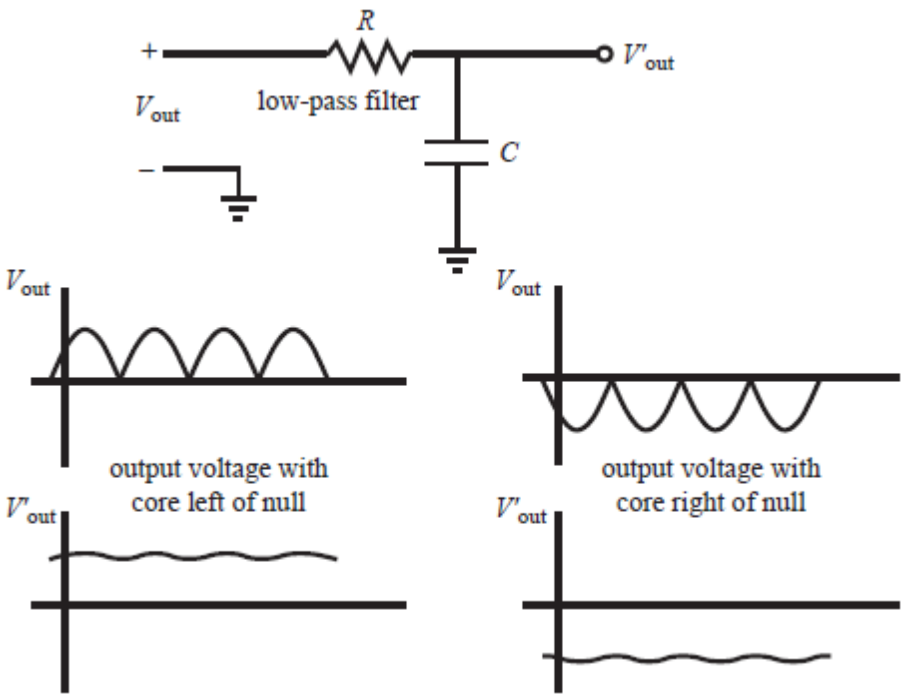


Figure 36 LVDT output filter.

5.4 Resolver

A **resolver**, also known as a **synchro-resolver**, is an analog rotary position sensor that operates very much like the **LVDT**. It consists of a rotating shaft (rotor) with a primary winding and a stationary housing (stator) with two secondary windings offset by 90°. When the primary is excited with an AC signal, AC voltages are induced in the secondary coils, which are proportional to the sine and cosine of the shaft angle. Because of this, the resolver is useful in applications where trigonometric functions of position are required.

Standard resolvers have a primary winding on the rotor and two secondary windings on the stator. Variable reluctance resolvers, on the other hand, have no windings on the rotor. Their primary and secondary windings are all on the stator, but the saliency (exposed poles) of the rotor couples the sinusoidal variation in the secondary with the angular position. Figure 37 shows classical and variable reluctance resolvers.

When the primary winding, R1–R2, is excited with a sinusoidal signal as expressed in Eq. 41, a signal is induced in the secondary windings.

$$V_{r_1-r_2} = V_0 \sin \omega t$$

Eq. 41

The amount of coupling onto the secondary windings is a function of the position of the rotor relative to that of the stator, and an attenuation factor known as the resolver transformation ratio. Because the secondary windings are displaced mechanically by 90°, the two output sinusoidal

signals are phase shifted by 90° with respect to each other. The relationships between the resolver input and output voltages are given by

$$V_{S_3-S_1} = T\,V_0 \sin(\omega t) \sin\theta$$

Eq. 42

$$V_{S_2-S_4} = T\,V_0 \sin(\omega t) \cos\theta$$

Eq. 43

Eq. 42 is the sine signal; Eq. 43 is the cosine signal.

where: θ is the shaft angle, ω is the excitation signal frequency, $V_0 = E_0$ is the excitation signal amplitude, and T is the resolver transformation ratio.

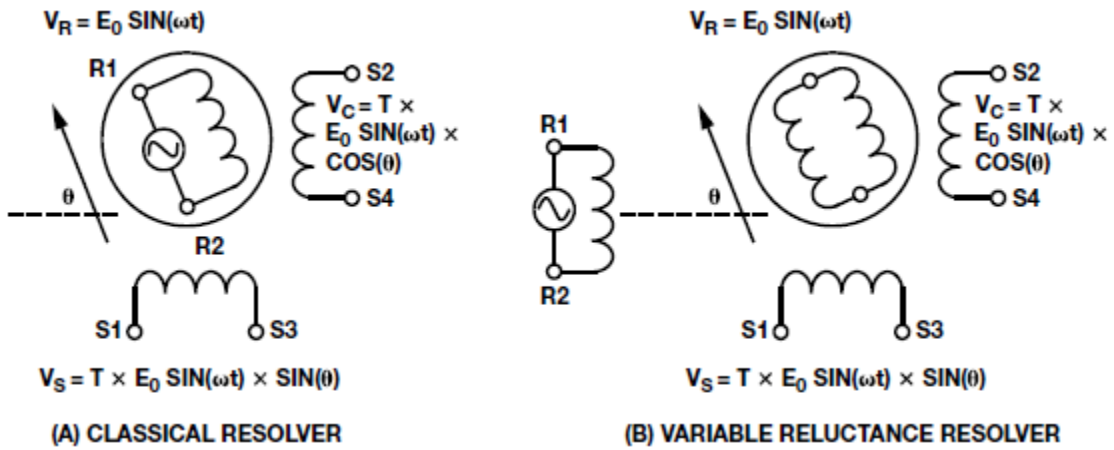


Figure 37 Classical resolver vs. variable reluctance resolver.

The two output signals are modulated by the sine and cosine of the shaft angle. A graphical representation of the excitation signal and the sine and cosine output signals is shown in Figure 38. The sine signal has maximum amplitude at 90° and 270° and the cosine signal has maximum amplitude at 0° and 180°.

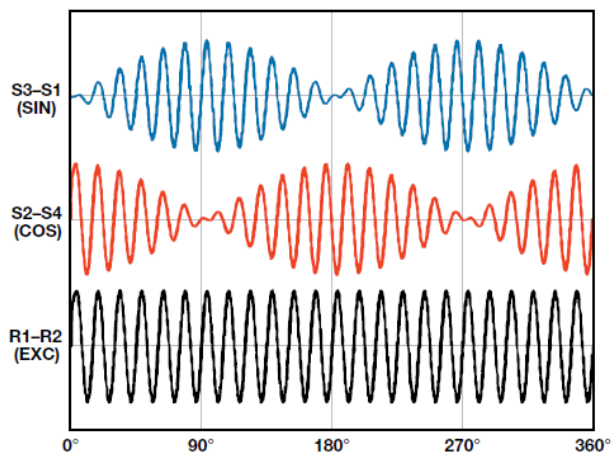


Figure 38 Resolver electrical signal representation.

The de-facto standard resolver transformation ratio is $T = 0.5$, which means that the maximum voltage produced by either secondary is half the amplitude of the reference signal.

For instance, with the rotor at 0° (called Electrical Zero or EZ), the amplitude of the sine secondary is 0 (since sin 0° = 0) and the amplitude of the cosine secondary will be at its maximum of half the reference amplitude (since cos 0° = 1):

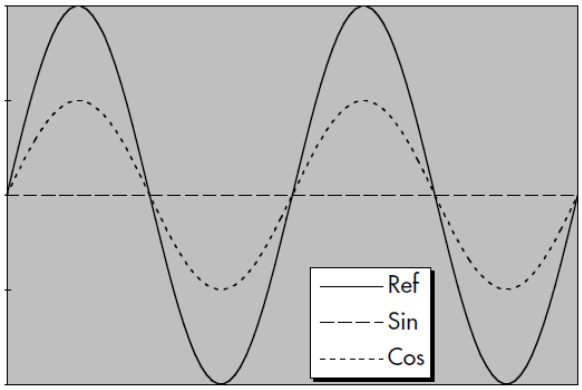


Figure 39 Resolver Signals with Rotor at 0° (EZ)

With the rotor at 45°, the secondary voltages are the same but only 70.7% of their maximum since (sin45° = cos45° = 0.707):

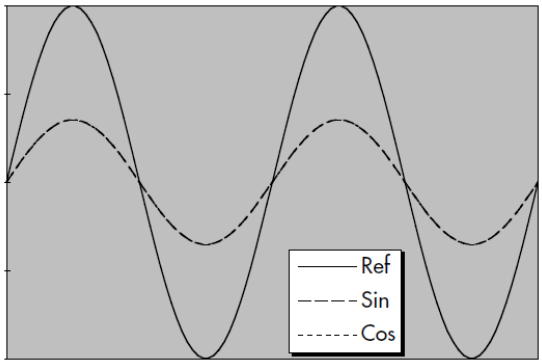


Figure 40 Resolver Signals with Rotor at 45°

With the rotor at 90°, the sin voltage is at maximum and the cosine voltage is zero:

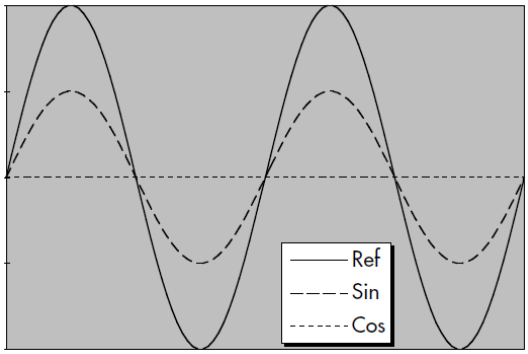


Figure 41 Resolver Signals with Rotor at 90°

Dr. Mohanned AL-Khafaji

With the rotor at 135°, the amplitudes of the secondary voltages are the same as at 45°, but the phase of the cosine voltage reverses since cos 135° is –0.707:

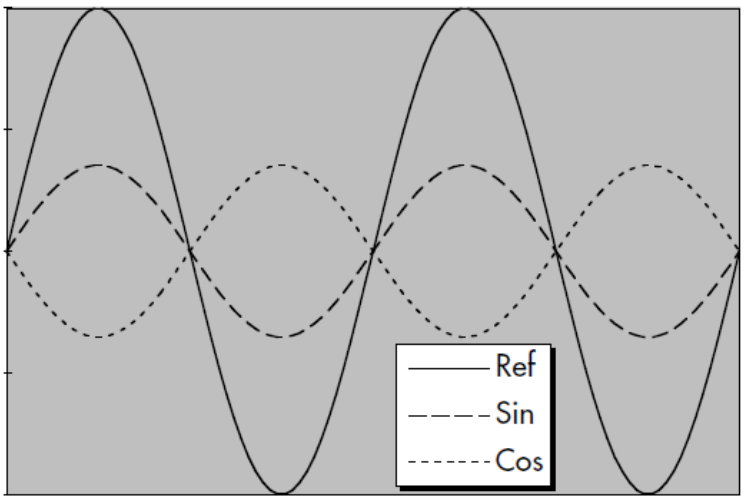


Figure 42 Resolver Signals with Rotor at 135°

5.5 Digital Optical Encoder

A digital optical encoder is a device that converts motion into a sequence of digital pulses. By counting a single bit or decoding a set of bits, the pulses can be converted to relative or absolute position measurements. Encoders have both linear and rotary configurations, but the most common type is rotary. Rotary encoders are manufactured in two basic forms: the absolute encoder where a unique digital word corresponds to each rotational position of the shaft, and the incremental encoder, which produces digital pulses as the shaft rotates, allowing measurement of relative displacement of the shaft. As illustrated in Figure 43, most rotary encoders are composed of a glass or plastic code disk with a photographically deposited radial pattern organized in tracks. As radial lines in each track interrupt the beam between a photo emitter-detector pair, digital pulses are produced.

The optical disk of the absolute encoder is designed to produce a digital word that distinguishes N distinct positions of the shaft. For example, if there are eight tracks, the encoder is capable of measuring 256 (2^8) distinct positions corresponding to an angular resolution of 1.406^o ($\frac{360^o}{256}$). The most common types of numerical encoding used in the absolute encoder are gray and natural binary codes. To illustrate the action of an absolute encoder, the gray code and natural binary code disk track patterns for a simple four-track (4-bit) encoder are illustrated in Figure 44 and Figure 45. The linear patterns and associated timing diagrams are what the photodetectors sense as the code disk circular tracks rotate with the shaft. The output bit codes for both coding schemes are listed in Table 9.1.

The gray code is designed so that only one track (one bit) changes state for each count transition, unlike the binary code where multiple tracks (bits) can change during count transitions. This effect can be seen clearly in Figure 44 and Figure 45 and in the last two columns of Table 2. For the gray code, the uncertainty during a transition is only one count, unlike with the binary code, where the uncertainty could be multiple counts.

Dr. Mohanned AL-Khafaji

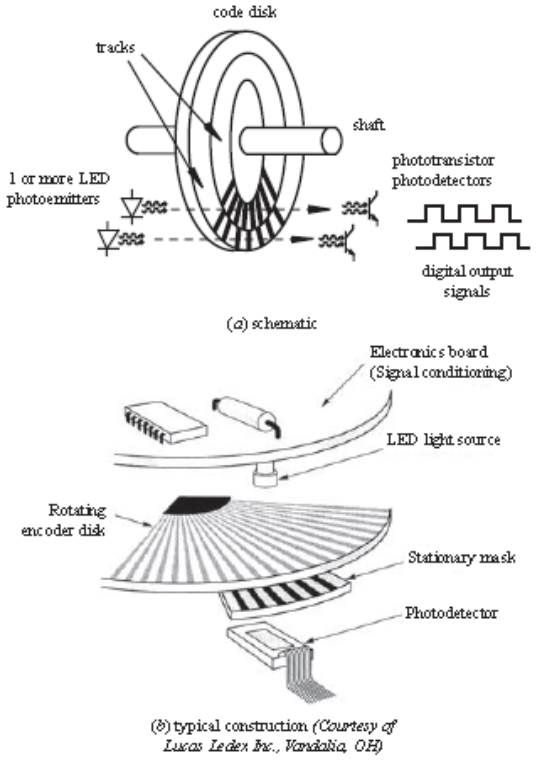


Figure 43 Components of an optical encoder.

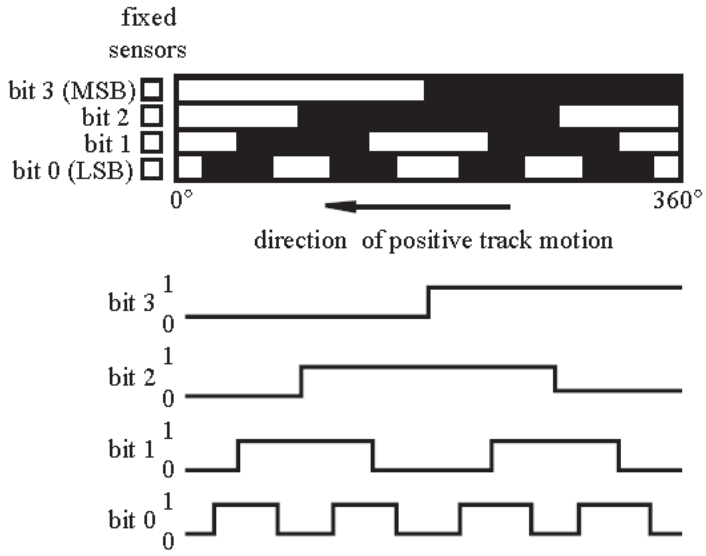
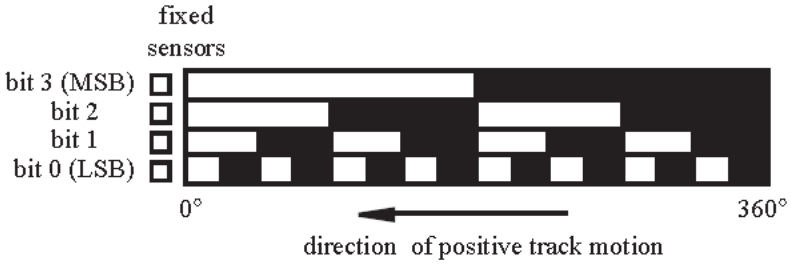
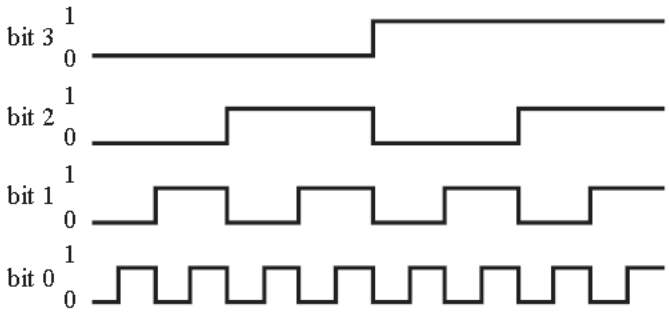


Figure 44 4-bit gray code absolute encoder disk track patterns.





(*a*) schematic and signals

(*b*) actual disk *(Courtesy of Parker Compumotor Division, Rohnert Park, CA)*

Figure 45 4-bit natural binary absolute encoder disk track patterns.

Table 2 4-bit gray and natural binary codes

Decimal Code	Rotation Range (°)	Natural binary code ($B_3B_2B_1B_0$)	Gray code ($G_3G_2G_1G_0$)
0	0–22.5	0000	0000
1	22.5–45	0001	0001
2	45–67.5	0010	0011
3	67.5–90	0011	0010
4	90–112.5	0100	0110
5	112.5–135	0101	0111
6	135–157.5	0110	0101
7	157.5–180	0111	0100
8	180–202.5	1000	1100
9	202.5–225	1001	1101
10	225–247.5	1010	1111
11	247.5–270	1011	1110
12	270–292.5	1100	1010
13	292.5–315	1101	1011
14	315–337.5	1110	1001
15	337.5–360	1111	1000

Dr. Mohanned AL-Khafaji

Because the gray code provides data with the least uncertainty but the natural binary code is the preferred choice for direct interface to computers and other digital devices, a circuit to convert from gray to binary code is desirable. Figure 9.15 shows a simple circuit that utilizes Exclusive OR (XOR) gates to perform this

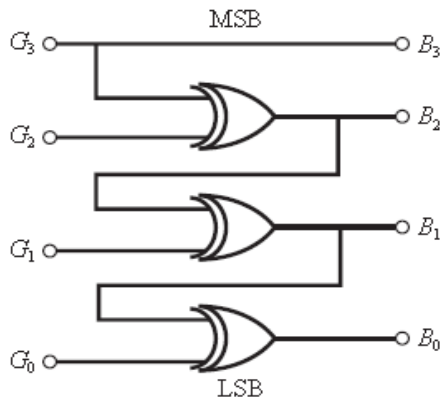


Figure 46 Gray-code-to-binary-code conversion.

For a gray-code-to-binary-code conversion of any number of bits N, the most significant bits of the binary and gray codes are always identical ($B_{N-1} = G_{N-1}$), and for each other bit, the binary bit is the XOR combination: $B_i = B_{i+1} \oplus G_i \ for\ i$ =0 to N - 2. This pattern can be easily seen in the 4-bit example above.

The **incremental encoder**, sometimes called a relative encoder, is simpler in design than the absolute encoder. It consists of two tracks and two sensors whose outputs are designated A and B. As the shaft rotates, pulse trains occur on A and B at a frequency proportional to the shaft speed, and the lead-lag phase relationship between the signals yields the direction of rotation as described in detail below. The code disk pattern and output signals A and B are illustrated in Figure 47. By counting the number of pulses and knowing the resolution of the disk, the angular motion can be measured. A and B are 1/4 cycle out of phase with each other and are known as quadrature signals. Often a third output, called INDEX, yields one pulse per revolution, which is useful in counting full revolutions. It is also useful to define a reference or zero position.

Figure 47a illustrates a configuration using two separate tracks for A and B, but a more common configuration uses a single track with the A and B sensors offset a quarter cycle on the track to yield the same signal pattern, as shown Figure 47b. A single-track code disk is simpler and cheaper to manufacture.

The quadrature signals A and B can be decoded to yield angular displacement and the direction of rotation as shown in Figure 48. Pulses appear on one of two output lines (CW and CCW) corresponding either to clockwise (CW) rotation or counterclockwise (CCW) rotation. Decoding transitions of A and B using sequential logic circuits can provide three different resolutions: 1X, 2X, and 4X. The 1X resolution provides an output transition at each negative edge of signal A or B, resulting in a single pulse for each quadrature cycle. The 2X resolution provides an output transition at every negative or positive edge of signal A or B, resulting in two times the number of output pulses. The 4X resolution provides an output pulse at every positive and negative edge of signal A or B, resulting in four times the number of output pulses. The direction of rotation is determined by the level of one quadrature signal during an edge transition of the second quadrature signal. For example, in the 1X mode, A=↓ with B = 1 implies clockwise rotation, and B = ↓ with A = 1 implies

Dr. Mohanned AL-Khafaji

counterclockwise rotation. If we only had one signal instead of both A and B, it would be impossible to determine the direction of rotation. Furthermore, shaft jitter around an edge transition in the single signal would result in erroneous pulses.

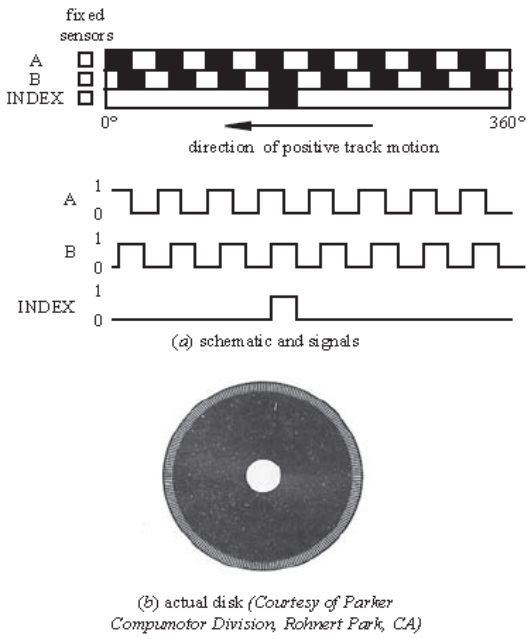


Figure 47 Incremental encoder disk track patterns

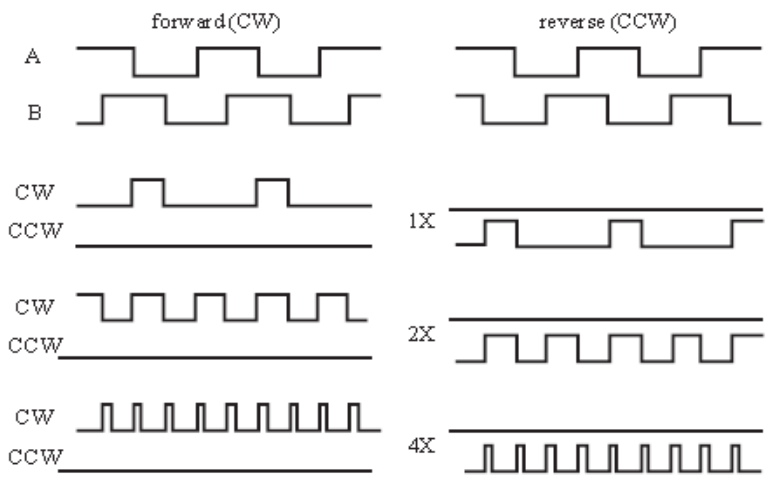


Figure 48 Quadrature direction sensing and resolution enhancement.

5.5.1 Binary Number and Code

n mathematics and digital electronics, a binary number is a number expressed in the binary numeral system, or base-2 numeral system, which represents numeric values using two different symbols: typically **0 (zero)** and **1 (one)**. More specifically, the usual base-2 system is a positional notation with a radix of 2. Because of its straightforward implementation in digital electronic

circuitry using logic gates, the binary system is used internally by almost all modern computers and computer-based devices such as mobile phones. Each digit is referred to as a bit.

Any number can be represented by any sequence of bits (binary digits), which in turn may be represented by any mechanism capable of being in two mutually exclusive states. The following symbols can be interpreted as the binary numeric value of 667:

Most Significant bit (MSB) → 1 0 1 0 0 1 1 0 1 1 ← least Significant bit (MSB)

The **most significant bit** is the one that has the highest power in the word. Whereas, the **least significant bit** is the one which has the lowest power, or it can be defined if the bit which has power of zero.

Assume that there is a binary code 11001101, in order to translate the binary code to it's equivalent decimal the following relationship can be used.

$$11001101_2 = 1 \times 2^7 + 1 \times 2^6 + 0 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$
$$= 205$$

Nate using the division by 2 method to convert decimal to binary, it can be called (Short Division by Two with Remainder)

$division$	$remindee$
205	1 LSB
102	0
51	1
25	1
12	0
6	0
3	1
1	1 MSB

The binary number can be written the MSB on the left going to right the Lsbwill be the last one, and it is 11001101. Another example taking 64

$division$	$remindee$
64	0 LSB
32	0
16	0
8	0
4	0
2	0
1	1MSB

The number will be 01000000.

Dr. Mohanned AL-Khafaji

6 DIGITAL REPRESENTATIONS

The **base** of the number system indicates the number of different symbols that can be used to represent a digit. In base 10, the symbols are 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. Each digit in a decimal number is a placeholder for different powers of 10 according to

$$d_{n-1} \cdots d_3 d_2 d_1 d_0 = d_{n-1} \cdot 10^{n-1} + \cdots + d_2 \cdot 10^2 + d_1 \cdot 10^1 + d_0 \cdot 10^0$$

Eq. 44

7 ANALOG SIGNAL PROCESSING USING OPERATIONAL AMPLIFIERS

Since electrical circuits occur in virtually all mechatronic and measurement systems, it is essential that engineers develop a basic understanding of the acquisition and processing of electrical signals. Usually these signals come from transducers, which convert physical quantities (e.g., temperature, strain, displacement, flow rate) into currents or voltages, usually the latter. The transducer output is usually described as an analog signal, which is continuous and time varying.

Often the signals from transducers are not in the form we would like them to be. They may

- Be too small, usually in the millivolt range
- Be too "noisy," usually due to electromagnetic interference
- Contain the wrong information, sometimes due to poor transducer design or installation
- Have a DC offset, usually due to the transducer and instrumentation design

Many of these problems can be solved, and the desired signal information extracted through appropriate analog signal processing. The simplest and most common form of signal processing is amplification, where the magnitude of the voltage signal is increased. Other forms include signal inversion, differentiation, integration, addition, subtraction, and comparison.

7.1 Amplifier

An amplifier increases the amplitude of a signal without affecting the phase relationships of different components of the signal.

When choosing or designing an amplifier, we must consider size, cost, power consumption, input impedance, output impedance, gain, and bandwidth. Physical size depends on the components used to construct the amplifier.

Generally, we model an amplifier as a two-port device, with an input and output voltage referenced to ground, as illustrated in Figure 49. The voltage **gain** A_v of an amplifier defines the factor by which the voltage is changed:

$$A_v = \frac{V_{out}}{V_{in}}$$

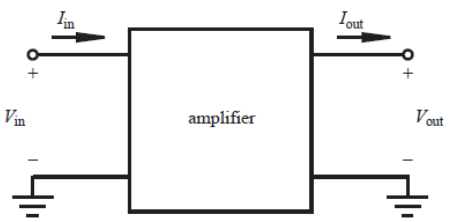


Figure 49 Amplifier model.

Dr. Mohanned AL-Khafaji

Normally we want an amplifier to exhibit amplitude linearity, where the gain is constant for all frequencies. However, amplifiers may be designed to intentionally amplify only certain frequencies, resulting in a filtering effect. In such cases, the output characteristics are governed by the amplifier's bandwidth and associated cutoff frequencies.

The input impedance of an amplifier, Z_{in}, is defined as the ratio of the input voltage and current:

$$Z_{in} = \frac{V_{in}}{I_{in}}$$

Eq. 45

Most amplifiers are designed to have a large input impedance so very little current is drawn from the input.

The output impedance is a measure of how much the output voltage drops with output current:

$$Z_{out} = \frac{V_{out}}{I_{out}}$$

Eq. 46

For most operational amplifiers, Z_{in} is larger than look and Z_{out} is a few ohms or less.

7.2 OPERATIONAL AMPLIFIERS

The **operational amplifier**, or **op amp**, is a low-cost and versatile integrated circuit consisting of many internal transistors, resistors, and capacitors manufactured into a single chip of silicon. It can be combined with external discrete components to create a wide variety of signal processing circuits. The op amp is the basic building block for

1. Amplifiers
2. Integrators
3. Summers
4. Differentiators
5. Comparators
6. A/D and D/A converters
7. Active filters
8. Sample and hold amplifiers

The **op amp** derives its name from its ability to perform so many different operations.

7.3 Ideal Model for the Operational Amplifier

Figure 50 shows the schematic symbol and terminal nomenclature for an ideal **op amp**. It is a differential input, single output amplifier that is assumed to have infinite gain. The two inputs are called the **inverting input**, labeled with a **minus** sign, and the **noninverting input**, labeled with a **plus** sign. The ∞ symbol is sometimes used in the schematic to denote the infinite gain and the assumption that it is an ***ideal op amp***. The voltages are all referenced to a common ground. The op amp is an **active device** requiring connection to an external power supply, usually plus and minus 15 V. The external supply is not normally shown on circuit schematics. Since the op amp is an active device, output voltages and currents can be larger than the signals applied to the inverting and noninverting terminals.

Dr. Mohanned AL-Khafaji

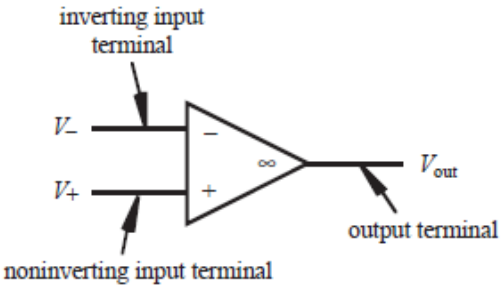


Figure 50 Op amp terminology and schematic representation.

As illustrated in Figure 51, an op amp circuit usually includes **feedback** from the **output** to the **negative (inverting)** input. This so-called closed loop configuration results in stabilization of the amplifier and control of the gain. When feedback is absent in an op amp circuit, the op amp is said to have an open loop configuration. This configuration results in considerable instability due to the very high gain, and it is seldom used.

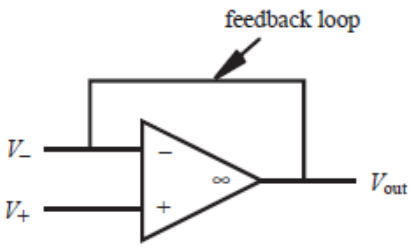


Figure 51 Op amp feedback

Figure 52 illustrates an ideal model that can aid in analyzing circuits containing op amps. This model based on the following assumptions that describe an ideal op amp:

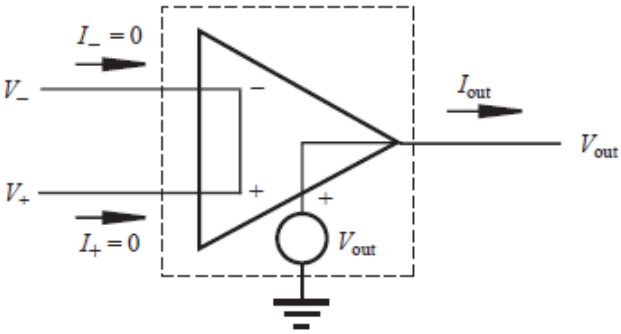


Figure 52 Op amp equivalent circuit

1- It has infinite impedance at both inputs; hence, no current drawn from the input circuits. Therefore,

$$I_+ = I_- = 0$$

Eq. 47

2- It has infinite gain. As consequence, the difference between the input voltages must be 0; otherwise, the output would be infinite. This denoted in Figure 52 by the shorting of the two inputs. Therefore,

$$V_+ = V_-$$

Eq. 48

Even though we indicate a short between the two inputs, we assume no current may flow through this short.

3- It has zero output impedance. Therefore, the output voltage does not depend on the output current.

Note that V_{out} , V_{+} , and V_{-} are all referenced to a common ground. In addition, for stable linear behavior, there must be feedback between the output and the inverting input.

These assumptions and the model may appear illogical and confusing, but they provide a close approximation to the behavior of a real op amp when used in a circuit that includes negative feedback. With the aid of this ideal model, we need only Kirchhoff's laws and Ohm's law to completely analyze op amp circuits.

7.3.1 Inverting Amplifier

An inverting amplifier is constructed by connecting two external resistors to an **op amp** as shown in Figure 53. As the name implies, this circuit inverts and amplifies the input voltage. Note that the resistor RF forms the feedback loop. This feedback loop always goes from the output to the inverting input of the op amp, implying negative feedback.

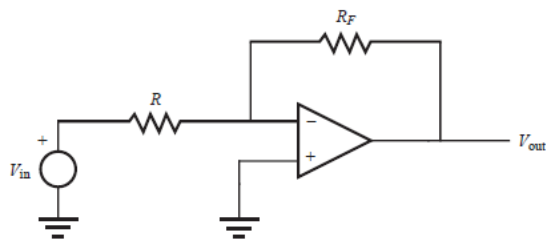


Figure 53 Inverting amplifier.

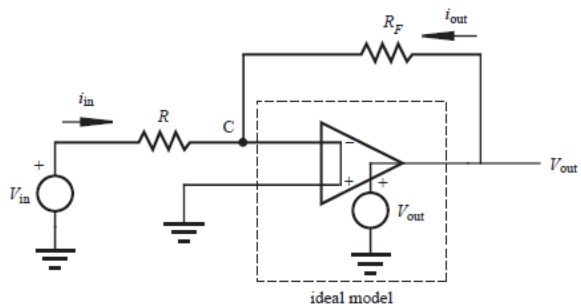


Figure 54 Equivalent circuit for an inverting amplifier.

By using Kirchhoff's laws and Ohm's law to analyze this circuit. First, replacing the op amp with its ideal model shown within the dashed box in Figure 54.

Applying Kirchhoff's current law at node C and utilizing assumption 1, that no current can flow into the inputs of the op amp,

$$i_{in} = -i_{out}$$

Eq. 49

Also, because the two inputs are assumed to be shorted in the ideal model, C is effectively at ground potential:

$$V_c = 0$$

Eq. 50

Because the voltage across resistor R is $V_{in} - V_c = V_{in}$, from Ohm's law,

$$V_{in} = i_{in}R$$

Eq. 51

and because the voltage across resistor R_F is $V_{out} - V_c = V_{out}$,

$$V_{out} = i_{out}R_F$$

Eq. 52

Substituting Eq. 49 into Eq. 51 gives

$$V_{out} = -i_{in}R_F$$

Eq. 53

Dividing Eq. 53 by Eq. 51 yields the $\frac{input}{output}$ relationship:

$$\frac{V_{out}}{V_{in}} = -\frac{R_F}{R}$$

Eq. 54

Therefore, the voltage gain of the inverted amplifier is determined simply by the external resistors R_F and R, and it is always negative. The reason this circuit called an **inverting amplifier** is that it reverses the polarity of the input signal. This results in a phase shift of 180° for periodic signals. For example, if the square wave V_{in} shown in Figure 55 is connected to an inverting amplifier with a gain of -2, the output V_{out} is inverted and amplified, resulting in a larger amplitude signal with 180° out of phase with the input.

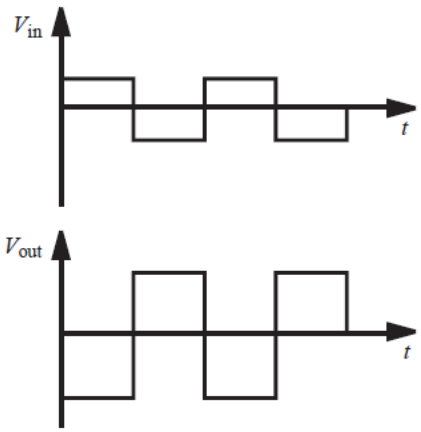


Figure 55 the input and inverted output voltage of the inverted amplifier

Dr. Mohanned AL-Khafaji

7.3.2 Noninverting Amplifier

Figure 56 shows the schematic of **noninverting amplifier**. This circuit amplifies the input voltage without inverting the signal. Kirchhoff's laws and Ohm's law could be applied to determine the voltage gain of this amplifier. Replacing the **op amp** with the ideal model shown in the dashed box in Figure 57.

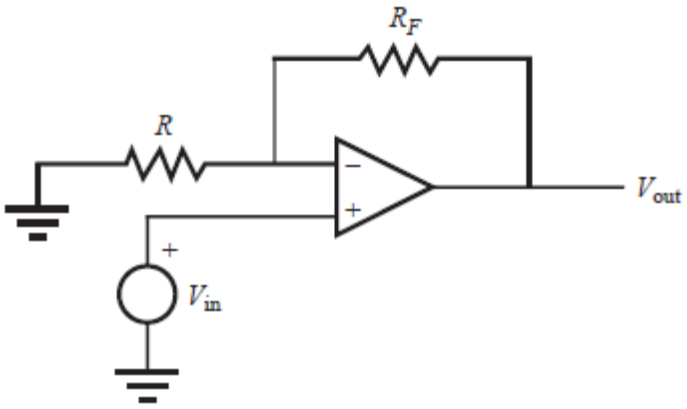


Figure 56 Noninverting amplifier.

The voltage at node C is V_{in} because the inverting and noninverting inputs are at the same voltage (assumption 2). Therefore, applying Ohm's law to resistor R,

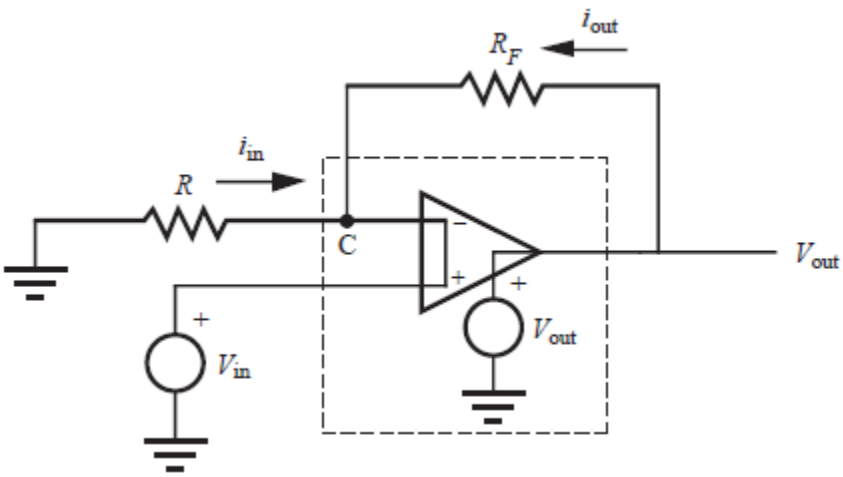


Figure 57 Equivalent circuit for a noninverting amplifier.

$$i_{in} = -\frac{V_{in}}{R}$$

Eq. 55

And also applying it on R_F too,

$$i_{out} = \frac{V_{out} - V_{in}}{R_F}$$

Eq. 56

Solving Eq. 56 for V_{out} gives

$$V_{out} = V_{in} + i_{out}\, R_F$$

Eq. 57

Applying KCL at node C gives

$$i_{in} = -i_{out}$$

Eq. 58

Hence the Eq. 55 can be rewritten as

$$V_{in} = i_{out}\, R$$

Eq. 59

Divide Eq. 57 by Eq. 59 the voltage gain of the noninverting amplifier can be written as:

$$\frac{V_{out}}{V_{in}} = \frac{V_{in} + i_{out}\, R_F}{V_{in}} = \frac{i_{out}\, R + i_{out}\, R_F}{i_{out}\, R} = 1 + \frac{R_F}{R}$$

Eq. 60

Therefore, the noninverting amplifier has a positive gain greater than or equal to one ($A_v \geq 1$). This is useful in isolating one portion of a circuit from another by transmitting a scaled voltage without drawing appreciable current.

7.3.3 Buffer (Follower) Amplifier

If we let $R_F = 0$ and $R = \infty$ in the noninverting op amp circuit the resulting circuit can be represented as shown in Figure 57. This circuit is known as a buffer or follower because $V_{out} = V_{in}$.

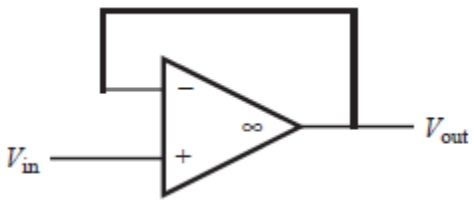


Figure 58 Buffer (Follower) Amplifier.

The buffer amplifier has a high input impedance and low output impedance. This circuit is useful in applications where you need to couple to a voltage signal without loading the source of the voltage. The high input impedance of the op amp effectively isolates the source from the rest of the circuit.

7.3.4 Summing Amplifier

The summer op amp circuit shown in Figure 59 is used to add analog signals. By analyzing the circuit with

$$R_1 = R_2 = R_F$$

Eq. 61

By drawing the equivalent summer circuit as shown in Figure 60.

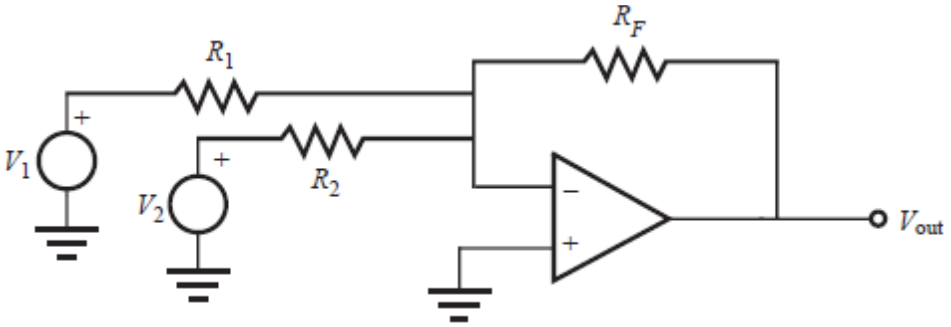


Figure 59 summer amplifier

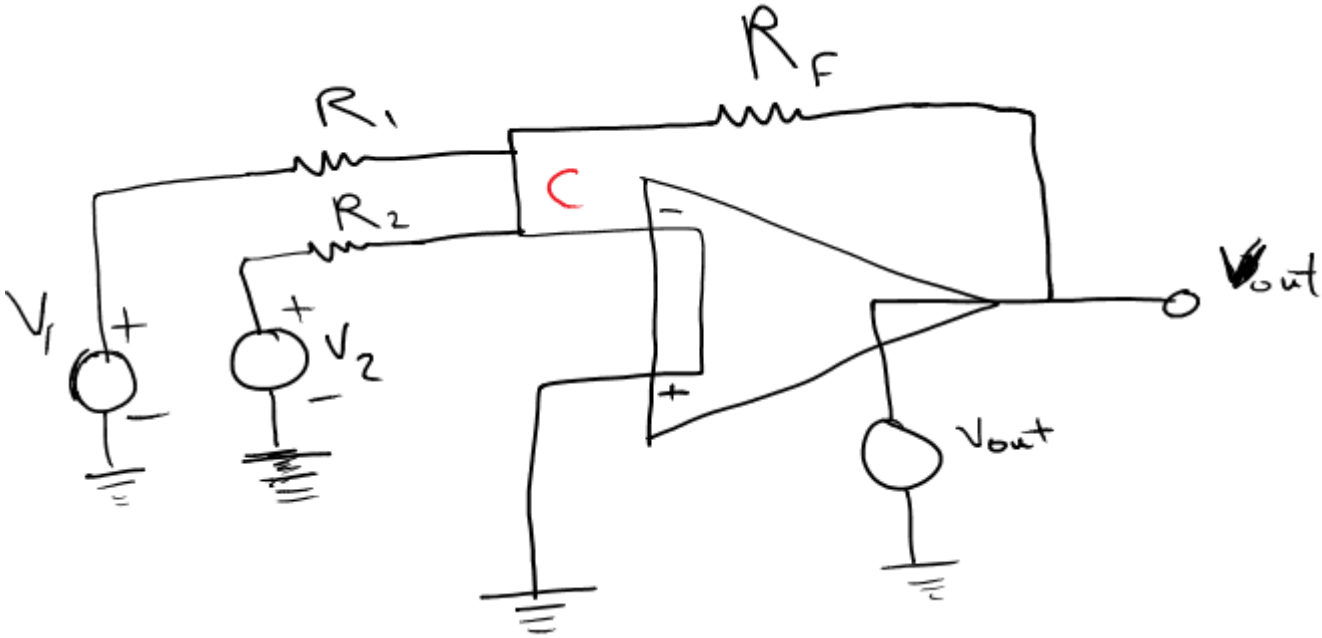


Figure 60 Equivalent summer circuit

Applying KCL at node C,

$$i_1 + i_2 = -i_{out}$$

Eq. 62

Using Ohm's law

$$i_1 = \frac{V_1}{R_1}$$

Eq. 63

$$i_2 = \frac{V_2}{R_2}$$

Eq. 64

$$i_{out} = \frac{V_{out}}{R_F}$$

Eq. 65

Substituting Eq. 63, Eq. 64 and Eq. 65 in Eq. 62

$$\frac{V_1}{R_1} + \frac{V_2}{R_2} = -\frac{V_{out}}{R_F}$$

Eq. 66

From Eq. 61 the V_{out}

$$V_{out} = -(V_1 + V_2)$$

Eq. 67

It can be noted if $R_1 = R_2 \neq R_F$ hence Eq. 66 can be written as

$$V_{out} = -\frac{R_F}{R}(V_1 + V_2)$$

Eq. 68

Therefore, the circuit output is the negative sum of the inputs.

7.3.5 Non-Inverting Summing Amplifier

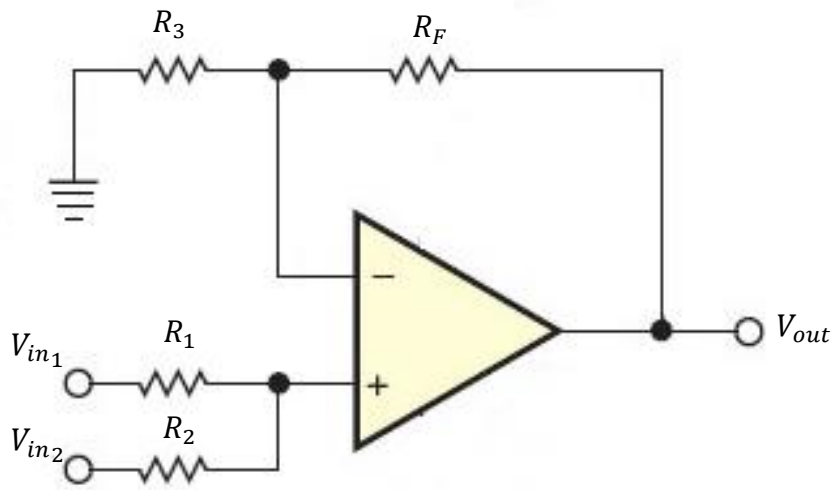


Figure 61 non-inverting summing amplifier

If $R_1 = R_2 = R_3 = R_F$

$$V_{out} = (V_{in_1} + V_{in_2})$$

Eq. 69

And if $R_1 = R_2 = R_3 \neq R_F$

$$V_{out} = \left(\frac{1}{2} + \frac{R_F}{2R_3}\right)(V_{in_1} + V_{in_2})$$

Eq. 70

7.3.6 Difference Amplifier

The **difference amplifier** circuit shown in Figure 62 is used to subtract analog signals. In analyzing this circuit, we can use the principle of superposition, which states that, whenever multiple inputs are applied to a linear system (e.g., an op amp circuit), we can analyze the circuit and determine the response for each of the individual inputs independently. The sum of the individual responses is equivalent to the overall response to the multiple inputs. Specifically, when the inputs are ideal voltage sources, to analyze the response due to one source, the other sources are shorted. If some inputs are current sources, they are replaced with open circuits.

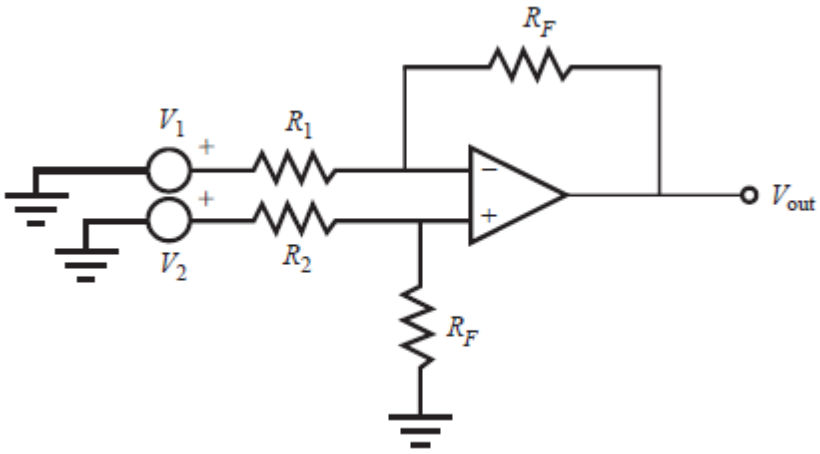


Figure 62 Difference amplifier circuit.

The first step in analyzing the circuit in Figure 62 is to replace V_2 with a short circuit, effectively grounding R_2. As shown in Figure 63, the result is an inverting amplifier. Therefore, from Eq. 54, the output due to input V_1 is

$$V_{out_1} = -\frac{R_F}{R_1} V_1$$

Eq. 71

The second step in analyzing the circuit in Figure 62 is to replace V_1 with a short circuit, effectively grounding R_1 , as shown in Figure 64a. This circuit is equivalent to the circuit shown in Figure 64b where the input voltage is

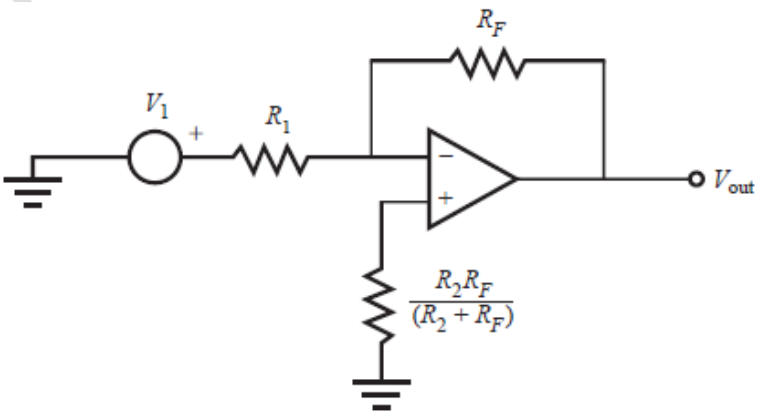
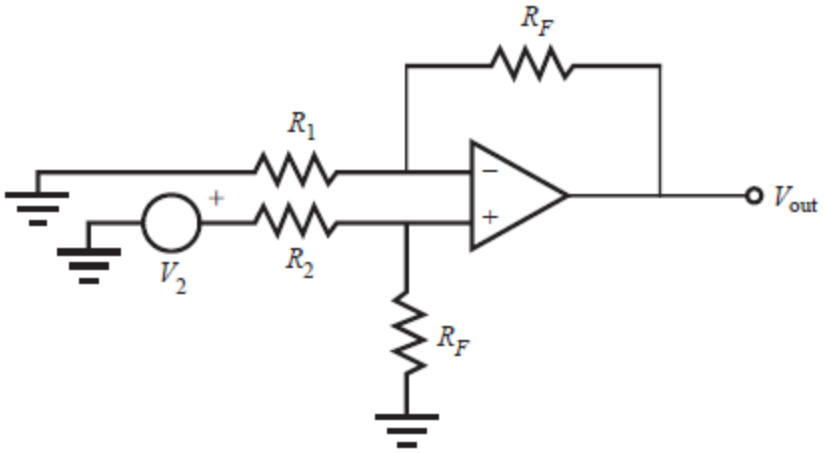
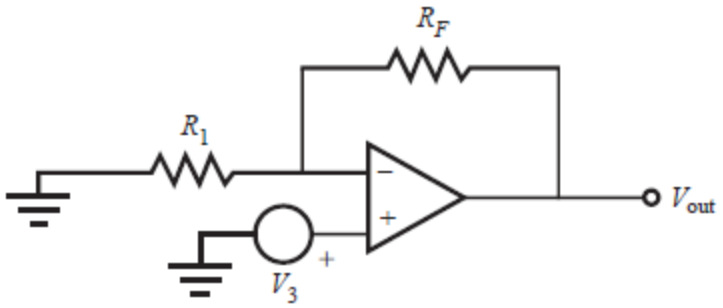


Figure 63 Difference amplifier with V_2 shorted.



(a) V_1 shorted



(b) equivalent circuit

Figure 64 Difference amplifier with V_1 shorted.

$$V_3 = \frac{R_F}{R_2 + R_F} V_2$$

Eq. 72

Because V_2 is divided between resistors R_2 and R_F.

The circuit in Figure 64b is a noninverting amplifier. Therefore, the output due to input V_2 is given by Eq. 60

$$V_{out_2} = \left(1 + \frac{R_F}{R_1}\right) V_3$$

Eq. 73

By substituting Eq. 72, this equation can be written as

$$V_{out_2} = \left(1 + \frac{R_F}{R_1}\right)\left(\frac{R_F}{R_2 + R_F}\right) V_2$$

Eq. 74

The principle of superposition states that the total output V_{out} is the sum of the outputs due to the individual inputs:

$$V_{out} = V_{out_1} + V_{out_2} = -\left(\frac{R_F}{R_1}\right) V_1 + \left(1 + \frac{R_F}{R_1}\right)\left(\frac{R_F}{R_2 + R_F}\right) V_2$$

Eq. 75

When $R_1 = R_2 = R$ he output voltage is an amplified difference of the input voltages:

$$V_{out} = \frac{R_F}{R}\,(V_2 - V_1)$$

Eq. 76

This result can also be obtained using the op amp rules, KCL, and Ohm's law

7.3.7 Instrumentation Amplifier

The difference amplifier presented in Section 7.3.6 may be satisfactory for low impedance sources, but its input impedance is too low for high-output impedance sources. Furthermore, if the input signals are very low level and include noise, the difference amplifier is unable to extract a satisfactory difference signal. The solution to this problem is the instrumentation amplifier. It has the following characteristics:

1- Very high input impedance
2- Large **common mode rejection ratio** (CMRR). The CMRR is the ratio of the **difference mode gain** to the **common mode gain**. The **difference mode gain** is the amplification factor for the difference between the **input signals**, and the **common mode gain** is the amplification factor for the average of the **input signals.** For an ideal difference amplifier, the common mode gain is 0, implying an infinite CMRR. When the common mode gain is nonzero, the output is nonzero when the inputs are equal and nonzero. It is desirable to minimize the common mode gain to suppress signals such as noise that are common to both inputs.
3- Capability to amplify low-level signals in a noisy environment, often a requirement in differential-output sensor signal-conditioning applications
4- Consistent bandwidth over a large range of gains

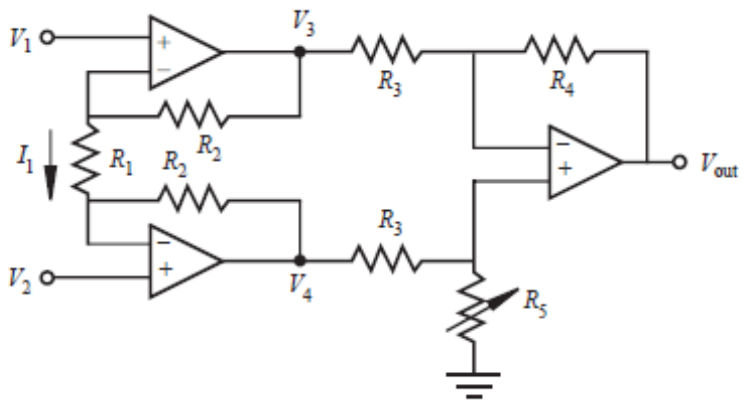


Figure 65 Instrumentation amplifier

Figure 65 shows an instrumentation amplifier schematic. This circuit could be analyzed by dividing it into two parts. The two op amps on the left provide a high-impedance amplifier stage where each input is amplified separately. This stage involves a moderate CMRR. The outputs V_3 and V_4 are supplied to the op amp circuit on the right, which is a difference amplifier with a potentiometer R5 used to maximize the overall CMRR.

Dr. Mohanned AL-Khafaji

Applying KCL and Ohm's law to the left portion of the circuit to express V_3 and V_4 in terms of V_1 and V_2, and using the assumptions and rules for an ideal op amp, it is clear that the current I_1 passes through R_1 and both feedback resistors R_2. Applying Ohm's law to the feedback resistors gives.

$$V_3 - V_1 = I_1 R_2$$

Eq. 77

$$V_2 - V_4 = I_1 R_2$$

Applying Ohm's law to R_1 gives

$$V_1 - V_2 = I_1 R_1$$

Eq. 78

To express V_3 and V_4 in terms of V_1 and V_2 , we eliminate I_1 by solving Eq. 78 for I_1 and substituting it into Eq. 76 and Eq. 77. The results are

$$V_3 = \left(\frac{R_2}{R_1} + 1\right) V_1 - \frac{R_2}{R_1} V_2$$

Eq. 79

$$V_4 = -\frac{R_2}{R_1} V_1 + \left(\frac{R_2}{R_1} + 1\right) V_2$$

Eq. 80

By analyzing the right portion of the circuit, it can be seen that:

$$V_{out} = \frac{R_5(R_3 + R_4)}{R_3(R_3 + R_5)} V_4 - \frac{R_4}{R_3} V_3$$

Eq. 81

We can substitute the expressions for V_3 and V_4 from Eq. 79 and Eq. 80 into Eq. 81 to express the output voltage V_{out} in terms of the input voltages V_1 and V_2. Assuming $R_5 = R_4$, the result is

$$V_{out} = \left[\frac{R_4}{R_3}\left(1 + 2\frac{R_2}{R_1}\right)\right](V_2 - V_1)$$

Eq. 82

A design objective for the instrumentation amplifier is to maximize the CMRR by minimizing the common mode gain. For a common mode input, $V_1 = V_2$, Eq. 82 yields an output voltage $V_{out} = 0$. Hence, the common mode gain is 0, and the CMRR is infinite if $R_5 = R_4$.

7.3.8 Integrator

If the feedback resistor of the inverting op amp circuit is replaced by a capacitor, the result is an integrator circuit. It is shown in Figure 66. Referring to the analysis for the inverting amplifier, the relationship between voltage and current for a capacitor:

$$\frac{dV_{out}}{dt} = \frac{i_{out}}{C}$$

Eq. 83

Dr. Mohanned AL-Khafaji

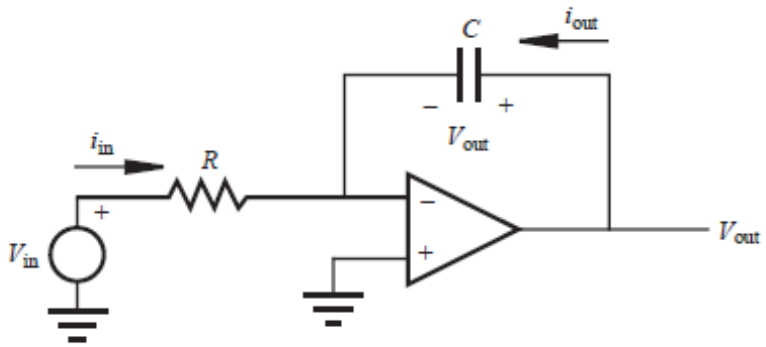


Figure 66 Ideal integrator.

Integrating gives

$$V_{out}(t) = \frac{1}{C}\int_0^t i_{out}(\tau)\, d\tau$$

Eq. 84

where τ is a dummy variable of integration. Since $i_{out} = -i_{in}$ and $i_{in} = \frac{V_{in}}{R}$,

$$V_{out} = -\frac{1}{RC}\int_0^t V_{in}(\tau)\, d\tau$$

Eq. 85

Therefore, the output signal is an inverted, scaled integral of the input signal.

A more practical integrator circuit is shown in Figure 67. The resistor R_s placed across the feedback capacitor is called a shunt resistor. Its purpose is to limit the low-frequency gain of the circuit. This is necessary because even a small DC offset at the input would be integrated over time, eventually saturating the op amp. The integrator is useful only when the scaled integral always remains below the maximum output voltage for the op amp. As a good rule of thumb, R_s should be greater than $10\ R_1$.

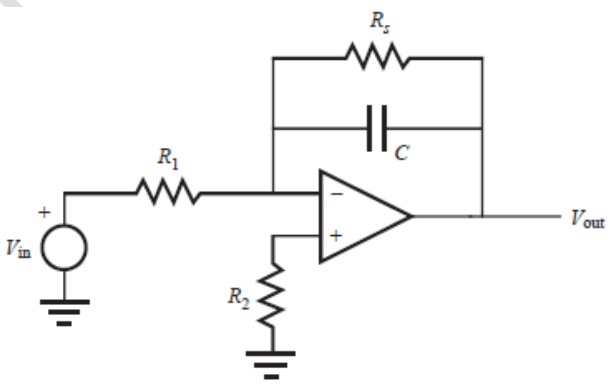


Figure 67 Improved integrator.

Because of the impedance and frequency response of the feedback circuit containing R_s and C, the circuit in Figure 67 acts as an integrator only for a range of frequencies. At very low frequencies,

Dr. Mohanned AL-Khafaji

the circuit behaves as an inverting amplifier because the impedance of the feedback loop is effectively R_s because the impedance of C is large at low frequencies. At very high frequencies (i.e.,$\omega = \frac{1}{R_1 C}$), the output is attenuated to zero because the feedback loop is effectively a short.

Any DC offset due to the input bias current is minimized by R_2 , which should be chosen to approximate the parallel combination of the input and shunt resistors:

$$R_2 = \frac{R_1 R_s}{R_1 + R_2}$$

Eq. 86

The reason for this is that the input bias current flowing into the inverting terminal is a result of the currents through R_1 and R_s , and the input bias current flowing into the noninverting terminal flows through R_2 . If the voltages generated by the bias current are the same, they have no net effect on the output.

7.3.9 Differentiator

If the input resistor of the inverting op amp circuit is replaced by a capacitor, the result is a differentiator circuit. It is shown in Figure 68. Referring to the analysis for the inverting amplifier, Eq. 51 is replaced by the relationship between voltage and current for a capacitor:

$$\frac{dV_{in}}{dt} = \frac{i_{in}}{C}$$

Eq. 87

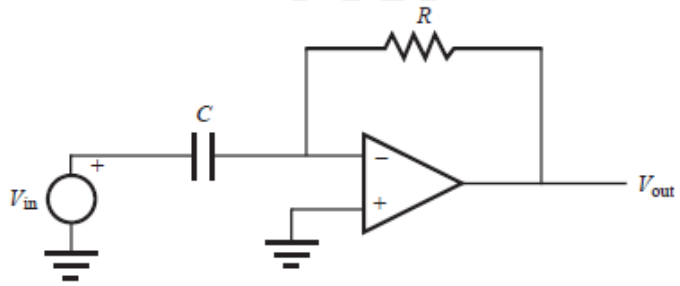


Figure 68 Differentiator.

Since $i_{in} = -i_{out}$ and $i_{out} = \frac{V_{out}}{R}$ hence

$$V_{out} = -RC\frac{dV_{in}}{dt}$$

Eq. 88

Therefore, the output signal is an inverted, scaled derivative of the input signal.

One note of caution concerning using differentiation in signal processing is that any electrical noise in the input signal will be accentuated in the output. In effect, the differentiator amplifies the fast-changing noise. Integration, on the other hand, has a smoothing effect, so noise is not a concern when using an integrator.

Dr. Mohanned AL-Khafaji

7.3.10 Sample and Hold Circuit

A sample and hold circuit is used extensively in analog-to-digital conversion where a signal value must be stabilized while it is converted to a digital representation. The sample and hold circuit illustrated in Figure 69 consists of a voltage-holding capacitor and a voltage follower. With switch S closed,

$$V_{out}(t) = V_{in}(t) \qquad \text{Eq. 89}$$

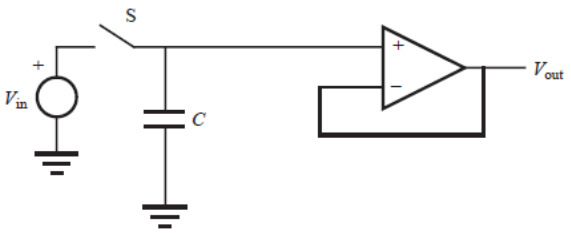


Figure 69 Sample and hold circuit.

When the switch is opened, the capacitor C holds the input voltage corresponding to the last sampled value, because negligible current is drawn by the follower. Therefore,

$$V_{out}(t - t_{sampeled}) = V_{in}(t_{sampeled}) \qquad \text{Eq. 90}$$

Where $t_{sampled}$ is the time when the switch was last opened. Often, an op amp buffer is also used on the V_{in} side of the switch to minimize current drain from the input voltage source V_{in}.

The type of capacitor used for this application is important. A low-leakage capacitor such as a polystyrene or polypropylene type would be a good choice. An electrolytic capacitor would be a poor choice because of its high leakage. This leakage would cause the output voltage value to drop during the “hold” period.

7.3.11 Comparator

The comparator circuit illustrated in Figure 5.22 is used to determine whether one signal is greater than another. The comparator is an example of an op amp circuit where there is no negative feedback and the circuit exhibits infinite gain. The result is that the op amp saturates. Saturation implies that the output remains at its most positive or most negative output value.

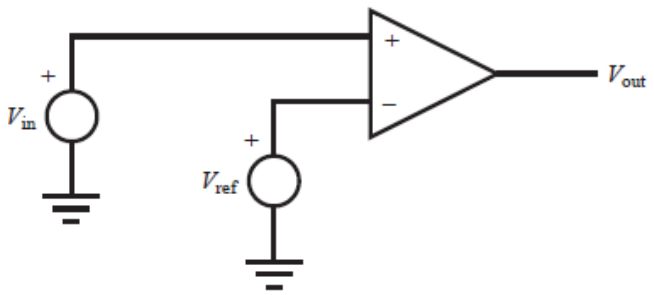


Figure 70 Comparator.

Certain op amps are specifically designed to operate in saturation as comparators. The output of the comparator is defined by

Dr. Mohanned AL-Khafaji

$$V_{out} = \begin{cases} +V_{sat} & if \quad V_{in} > V_{ref} \\ -V_{sat} & if \quad V_{in} < V_{ref} \end{cases} \qquad \text{Eq. 91}$$

Where V_{sat} is the saturation voltage of the comparator and V_{ref} is the reference voltage to which the input voltage V_{in} is being compared. The positive saturation value is slightly less than the positive supply voltage, and the negative saturation value is slightly greater than the negative supply voltage.

Often, comparators (e.g., LM339) have open-collector outputs, where the output states are controlled by an output transistor operating at cutoff or saturation. This type of output, illustrated in Figure 71, is called an open-collector output because the collector of the output transistor is not connected internally and requires an external powered circuit. The output transistor is ON (at saturation) and the output is effectively grounded when $V_{in} > V_{ref}$, and the output transistor is OFF (at cutoff) and the output is open circuited when $V_{in} < V_{ref}$.

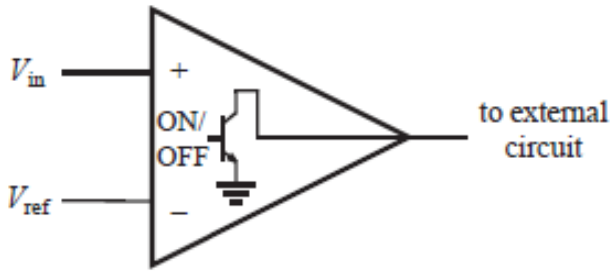


Figure 71 Comparator open-collector output.

8 Data Acquisition

Microprocessors, microcontrollers, single-board computers, and personal computers are in widespread use in mechatronic and measurement systems. It is increasingly important for engineers to understand how to directly access information and analog data from the surrounding environment with these devices. As an example, consider a signal from a sensor as illustrated by the analog signal in Figure 72. One could record the signal with an analog device such as a chart recorder, which physically plots the signal on paper, or display it with an oscilloscope. Another option is to store the data using a microprocessor or computer. This process is called **computer data acquisition**, and it provides more compact storage of the data (magnetic, optical, or flash media vs. long rolls of paper), can result in greater data accuracy, allows use of the data in a real time control system, and enables data processing long after the events have occurred.

To be able to input analog data to a digital circuit or microprocessor, the analog data must be transformed into coded digital values. The first step is to numerically evaluate the signal at discrete instants in time. This process is called **sampling**, and the result is a **digitized signal composed** of discrete values corresponding to each sample, as illustrated in Figure 72. Therefore, a digitized signal is a sequence of numbers that is an approximation to an analog signal. Note that the time relation between the numbers is an inherent property of the sampling process and need not be recorded

separately. The collection of sampled data points forms a data array, and although this representation is no longer continuous, it can accurately describe the original analog signal.

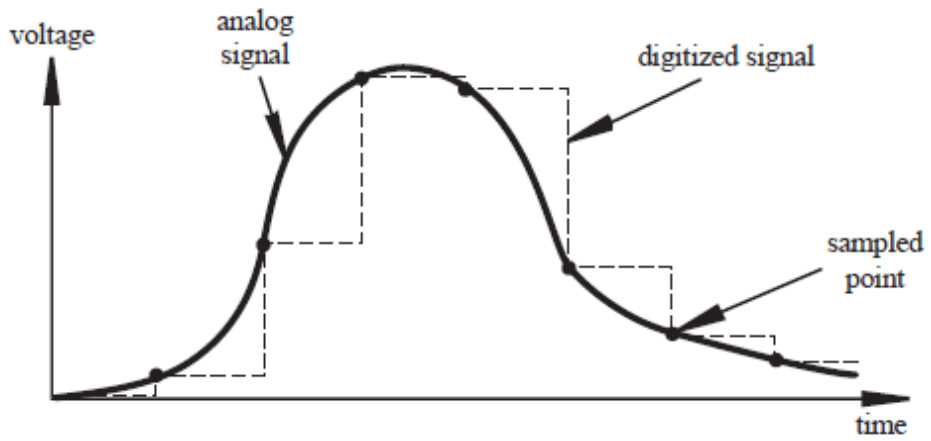


Figure 72 Analog signal and sampled equivalent.

The sampling theorem, also called **Shannon's sampling theorem**, states that we need to sample a signal at a rate more than two times the maximum frequency component in the signal to retain all frequency components. In other words, to faithfully represent the analog signal, the digital samples must be taken at a frequency f_s such that

$$f_s > 2f_{max} \qquad \text{Eq. 92}$$

Where f_{max} is the highest frequency component in the input analog signal. The term f_s is referred to as the sampling rate, and the limit on the minimum required rate ($2\ f_{max}$) is called the **Nyquist frequency**. If we approximate a signal by a truncated Fourier series, the maximum frequency component is the highest harmonic frequency. The time interval between the digital samples is

$$\Delta t = \frac{1}{f_s} \qquad \text{Eq. 93}$$

As an example, if the sampling rate is 5000 Hz, the time interval between samples would be $0.2\ ms$.

If a signal is sampled at less than two times its maximum frequency component, **aliasing** can result. Figure 73 illustrates an example of this with an analog sine wave sampled regularly at the points shown. Twelve equally spaced samples are taken over 10 cycles of the original signal. Therefore, the sampling frequency is $1.2\ f_0$, where f_0 is the frequency of the original sine wave. Because the sampling frequency is not greater than $2\ f_0$, we do not capture the frequency in the original signal. Furthermore, the apparent frequency in the sampled signal is $0.2\ f_0$ (2 aliased signal cycles for ten original signal cycles). You can think of this as a "phantom" frequency, which is an **alias** of the true frequency. Therefore, undersampling not only results in errors but also creates information that is not really there.

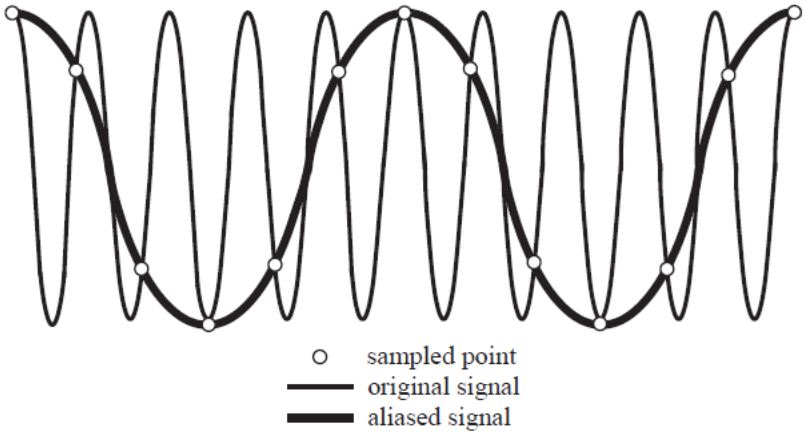


Figure 73 aliasing

Example: Sampling Theorem and Aliasing

Consider the function

$$F(t) = \sin(at) + \sin(bt)$$

Using a trigonometric identity for the sum of two sinusoidal functions, we can rewrite $F(t)$ as the following product:

$$F(t) = \left[2\cos\left(\frac{a-b}{2}t\right)\right]\sin\left(\frac{a+b}{2}t\right)$$

If frequencies a and b are close in value, the bracketed term has a very low frequency in comparison to the sinusoidal term on the right. Therefore, the bracketed term modulates the amplitude of the higher frequency sinusoidal term. The resulting waveform exhibits what is called a beat frequency that is common in optics, mechanics, and acoustics when two waves close in frequency add.

To illustrate aliasing associated with improper sampling, the waveform is plotted in the following figures using two different sampling frequencies. If a and b are chosen as

$$a = 1Hz = 2\pi\frac{rad}{sec} \qquad b = 0 \cdot 9a$$

Then to sample the signal $F(t)$ properly, the sampling rate must be more than twice the highest frequency in the signal:

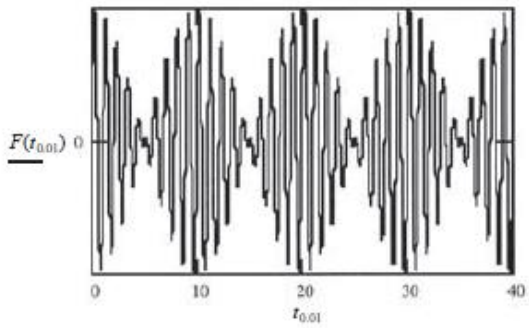
$$f_S > 2a = 2Hz$$

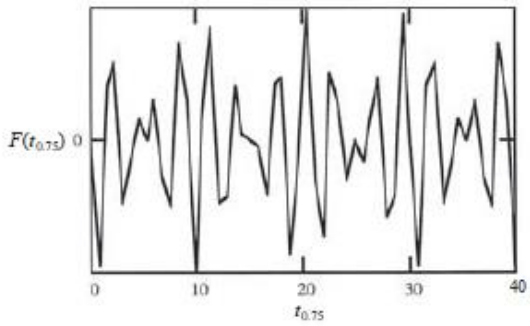
Therefore, the time interval between samples (plotted points) must be

$$\Delta t = \frac{1}{f_s} < 0.5 \ sec$$

The first data set is plotted with a time interval of 0.01 sec (100 Hz sampling rate), providing an adequate representation of the waveform. The second data set is plotted with a time interval of 0.75 sec (1.33 Hz sampling rate), which is less than twice the maximum frequency of the waveform (2 Hz). Therefore, the signal is **undersampled**, and **aliasing** results. The sampled waveform is an incorrect representation, and its observed maximum frequency appears to be approximately 0.4 Hz because there are approximately four cycles over 10 sec.

Dr. Mohanned AL-Khafaji





Unlike the second plot, the first plot retains all the frequency information in the analog signal.

8.1 Quantizing Theory

The process required to change a sampled analog voltage into digital form is called analog-to-digital conversion, conceptually involves two steps: quantizing and coding.

Quantizing is defined as the transformation of a continuous analog input into a set of discrete output states.

Coding is the assignment of a digital code word or number to each output state.

Figure 74 illustrates how a continuous voltage range is divided into discrete output states, each of which is assigned a unique code. Each output state covers a subrange of the overall voltage range. The stair-step signal represents the states of a digital signal generated by sampling a linear ramp of an analog signal occurring over the voltage range shown.

8.2 Analog-To-Digital Conversion

An **analog-to-digital** converter is an electronic device that converts an analog voltage to a digital code. The output of the A/D converter can be directly interfaced to digital devices such as microcontrollers and computers. **The resolution** of an A/D converter is the number of bits used to digitally approximate the analog value of the input. The number of possible states N is equal to the number of bit combinations that can be output from the converter:

$$N = 2^n \qquad \text{Eq. 94}$$

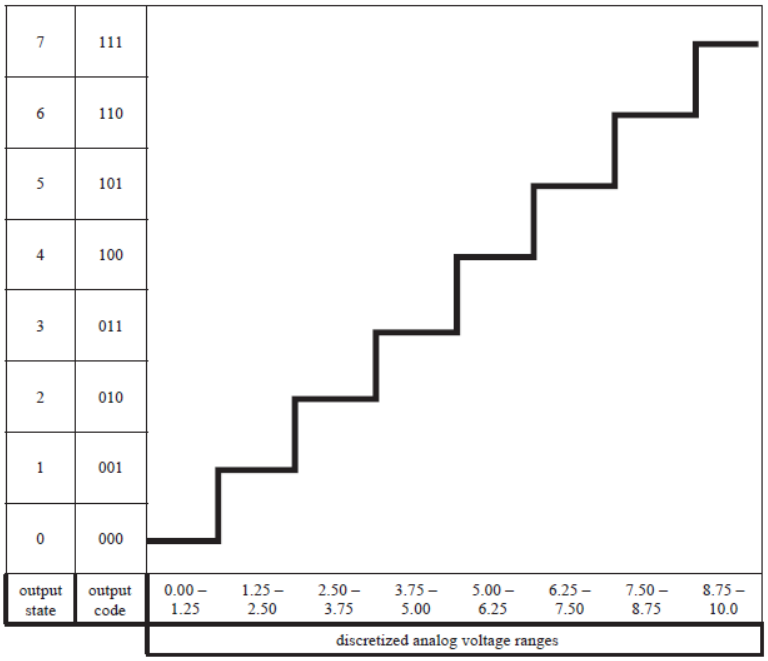


Figure 74 Analog-to-digital conversion.

Where n is the number of bits. For the example illustrated in Figure 74, the 3-bit device has 2^3 or 8 output states as listed in the first column. The output states are usually numbered consecutively from 0 to ($N - 1$). The corresponding code word for each output state is listed in the second column. Most commercial A/D converters are 8-, 10-, or 12-bit devices that resolve 256, 1024, and 4096 output states, respectively.

The number of **analog decision points** that occur in the process of quantizing is ($N - 1$). In Figure 74 , the **decision points** occur at 1.25 V, 2.50 V, . . . , and 8.75 V. The analog quantization size Q, sometimes called the code width, is defined as the full-scale range of the A/D converter divided by the number of output states:

$$Q = \frac{V_{max} - V_{min}}{N} \qquad \text{Eq. 95}$$

It is a measure of the analog change that can be resolved by the converter. Although the term resolution is defined as the number of output bits from an A/D converter, sometimes it is used to refer to the analog quantization size. For our example, the analog quantization size is $\frac{10}{8}V = 1.25$V. This means that the amplitude of the digitized signal has an error of at most 1.25 V. Therefore, the A/D converter can only resolve a voltage to within 1.25 V of the exact analog voltage.

To properly acquire an analog voltage signal for digital processing, the following components must be properly selected and applied in this sequence:

1. buffer amplifier
2. low-pass filter
3. sample and hold amplifier
4. analog-to-digital converter
5. computer

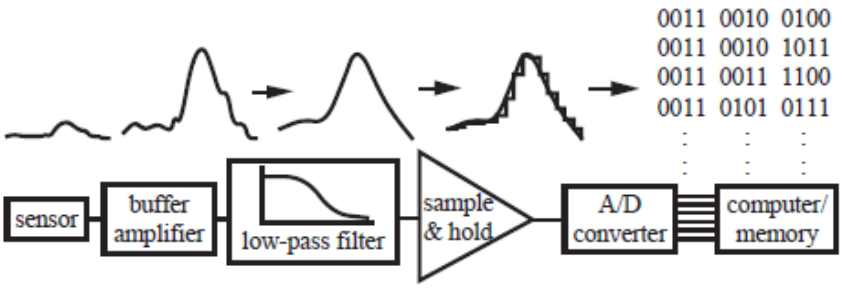


Figure 75 Components used in A/D conversion.

The components required for A/D conversion along with an illustration of their respective outputs are shown in Figure 75. The buffer amplifier isolates the output from the input (i.e., it draws negligible current and power from the input) and provides a signal in a range close to but not exceeding the full input voltage range of the A/D converter. The low-pass filter is necessary to remove any undesirable high- frequency components in the signal that could produce aliasing. The cutoff frequency of the low-pass filter should be no greater than 1/2 the sampling rate. The sample and hold amplifier maintains a fixed input value (from an instantaneous sample) during the short conversion time of the A/D converter.

The converter should have a resolution and analog quantization size appropriate to the system and signal. The computer must be properly interfaced to the A/D converter system to store and process the data. The computer must also have sufficient memory and permanent storage media to hold all of the data.

Analog-to-Digital Converters

A/D converters are designed based on a number of different principles: **successive approximation**, **flash or parallel encoding**, single-slope and dual-slope integration, switched capacitor, and delta sigma. The first two will be considered because they occur most often in commercial designs. The successive approximation A/D converter is very widely used because it is relatively fast and cheap. As shown in Figure 76, it uses a D/A converter in a feedback loop. D/A converters are described in the next section. When the start signal is applied, the sample and hold (S&H) amplifier latches the analog input. Then the control unit begins an iterative process, where the digital value is approximated, converted to an analog value with the D/A converter, and compared to the analog input with the comparator. When the D/A output equals the analog input, the end signal is set by the control unit, and the correct digital output is available at the output.

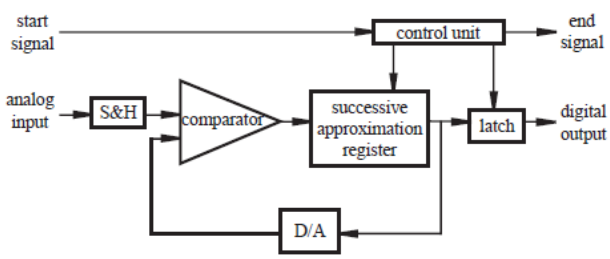


Figure 76 Successive approximation A/D converter.

If n is the resolution of the A/D converter, it takes n steps to complete the conversion. More specifically, the input is compared to combinations of binary fractions $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n})$ of the full-scale (FS) value of the A/D converter. The control unit first turns on the most significant bit (MSB)

Dr. Mohanned AL-Khafaji

of the register, leaving all lesser bits at 0, and the comparator tests the DAC output against the analog input. If the analog input exceeds the DAC output, the MSB is left on (high); otherwise, it is reset to 0. This procedure is then applied to the next lesser significant bit and the comparison is made again. After n comparisons have occurred, the converter is down to the least significant bit (LSB). The output of the DAC then represents the best digital approximation to the analog input. When the process terminates, the control unit sets the end signal signifying the end of the conversion.

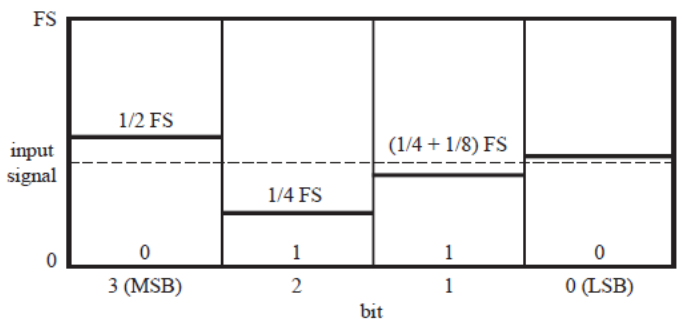


Figure 77 4-bit successive approximation A/D conversion.

As an example, a 4-bit successive approximation procedure is illustrated graphically in Figure 77. The MSB is $\frac{1}{2}FS$, which in this case is greater than the signal; therefore, the bit is turned off. The second bit is $\frac{1}{4}\ FS$ and is less than the signal, so it is left on. The third bit gives $\frac{1}{4} +\ 1/8$ of FS, which is still less than the analog signal, so the third bit is left on. The fourth provides $\frac{1}{4} + \frac{1}{8} + \frac{1}{16}$ of FS and is greater than the signal, so the fourth bit is turned off and the conversion is complete. The digital result is 0110. Higher resolution would produce a more accurate value.

An n -bit successive approximation A/D converter has a conversion time of $n\Delta T$, where ΔT is the cycle time for the D/A converter and control unit. Typical conversion times for 8-, 10-, and 12-bit successive approximation A/D converters range from 1 to 100 μs.

The fastest type of A/D converter is known as a flash converter. As Figure 78 illustrates, it consists of a bank of input comparators acting in parallel to identify the signal level. The output of the latches is in a coded form easily converted to the required binary output with combinational logic. The flash converter illustrated in Figure 78 is a 2-bit converter having a resolution of four output states.

Table 3 2-bit flash converter output

State	Code ($G_2G_1G_0$)	Binary (B_1B_0)	Voltage Range
0	000	00	0–1
1	001	01	1–2
2	011	10	2–3
3	111	11	3–4

Table 3 lists the comparator output codes and corresponding binary outputs for each of the states, assuming an input voltage range of 0 to 4 V. The voltage range is set by the V_{min} and V_{max} supply voltages shown in Figure 78 (0 V and 4 V in this example). The code converter is a simple combinational logic circuit. For the 2-bit converter, the relationships between the code bits G_i and the binary bits B_i are

Dr. Mohanned AL-Khafaji

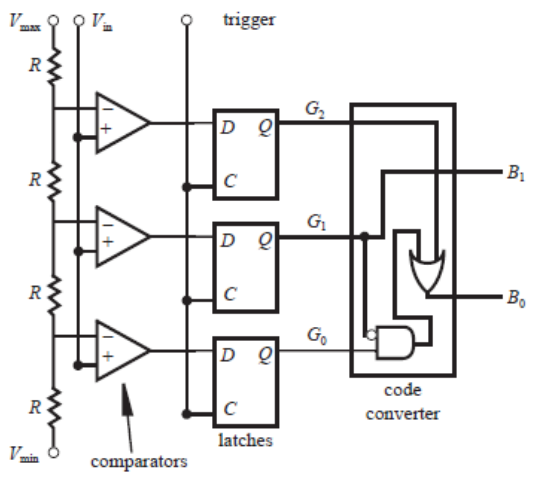


Figure 78 A/D flash converter.

$$B_0 = G_0\overline{G_1} + G_2$$
$$B_1 = G_1$$

Eq. 96

Adding more resolution is a simple matter of adding more resistors, comparators, and latches. The combinational logic code converter would also be different. Unlike with the successive approximation converter, adding resolution does not increase the time required for a conversion.

A single A/D converter can digitize several analog signals if the analog signals are multiplexed at the input to the A/D converter. An analog multiplexer simply switches among several analog inputs using transistors or relays and control signals. This can significantly reduce the cost of a system's design. In addition to cost, other parameters important in selecting an A/D converter are the input voltage range, output resolution, and conversion time.

8.3 Digital-To-Analog Conversion

In order to reverse the process of A/D conversion by changing a digital value to an analog voltage. This is called **digital-to-analog (D/A) conversion**. A D/A converter allows a computer or other digital device to interface with external analog circuits and devices.

The simplest type of **D/A converter** is a resistor ladder network connected to an inverting summer op amp circuit as shown in Figure 79. This particular converter is a 4-bit $R-2R$ resistor ladder network, which requires only two precision resistance values (R and $2R$). The digital input to the DAC is a 4-bit binary number represented by bits b_0 , b_1 , b_2 , and b_3 , where b_0 is **the least significant bit** and b_3 is **the most significant bit**. Each bit in the circuit controls a switch between ground and the inverting input of the op amp. To understand how the analog output voltage V_{out} is related to the input binary number, we can analyze the four different input combinations 0001, 0010, 0100, and 1000 and apply the principle of superposition for an arbitrary 4-bit binary number.

If the binary number is 0001, the b_0 switch is connected to the op amp, and the other bit switches are grounded. The resulting circuit is as shown in Figure 80. Because the noninverting input of the op amp is grounded, the inverting input is also at ground. The equivalent resistance between node V_0 and ground is R, which is the parallel combination of two $2R$ values. Therefore, V_0 is the result of voltage division of V_1 across two series resistors of equal value R:

Dr. Mohanned AL-Khafaji

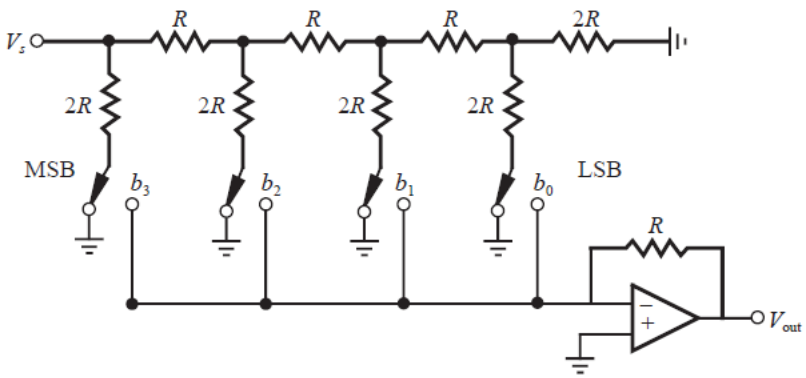


Figure 79 4-bit resistor ladder D/A converter.

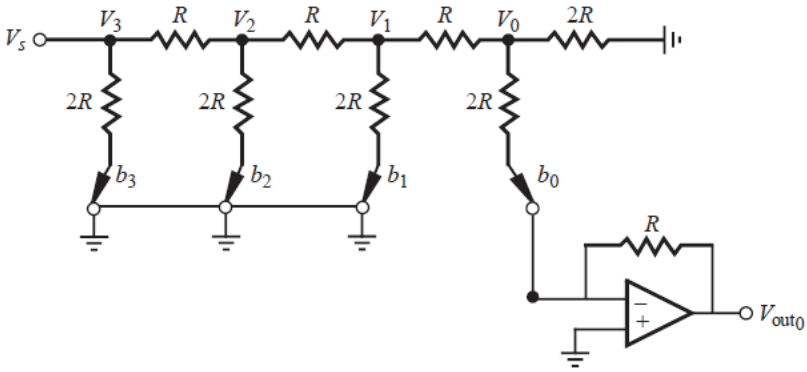


Figure 80 4-bit resistor ladder D/A with digital input 0001.

$$V_0 = \frac{1}{2}V_1 \qquad \text{Eq. 97}$$

Similarly, we can show that

$$V_1 = \frac{1}{2}V_2 \qquad and \qquad V_2 = \frac{1}{2}V_3 \qquad \text{Eq. 98}$$

Therefore,

$$V_0 = \frac{1}{8}V_3 = \frac{1}{8}V_s \qquad \text{Eq. 99}$$

V_0 is the input to the inverting amplifier circuit, which has a gain of

$$-\frac{R}{2R} = -\frac{1}{2} \qquad \text{Eq. 100}$$

Therefore, the analog output voltage corresponding to the binary input 0001 is

$$V_{out_0} = -\frac{1}{16}V_s \qquad \text{Eq. 101}$$

Similarly, we can show that, for the input 0010,

$$V_{out_1} = -\frac{1}{8}V_s \qquad \text{Eq. 102}$$

and for the input 0100,

$$V_{out_2} = -\frac{1}{4}V_s \qquad \text{Eq. 103}$$

and for the input 1000

$$V_{out_3} = -\frac{1}{2}V_s \qquad \text{Eq. 104}$$

The output for any combination of bits comprising the input binary number can now be found using the principle of superposition:

$$V_{out} = b_3V_{out_3} + b_2V_{out_2} + b_1V_{out_1} + b_0V_{out_0} \qquad \text{Eq. 105}$$

If V_s is $10\ V$, the output ranges from $0\ V\ to\ (-\frac{15}{16})10\ V$ for the 4-bit binary input, which has 16 values ranging from $0000\ (0)\ to\ 1111\ (15)$. A negative reference voltage V_s can be used to produce a positive output voltage range. Either case yields a unipolar output, which is either positive or negative but not both. A bipolar output, which ranges over negative and positive values, can be produced by replacing all ground references in the circuit with a reference voltage whose sign is opposite to V_s .

Dr. Mohanned AL-Khafaji