A SunCam online continuing education course

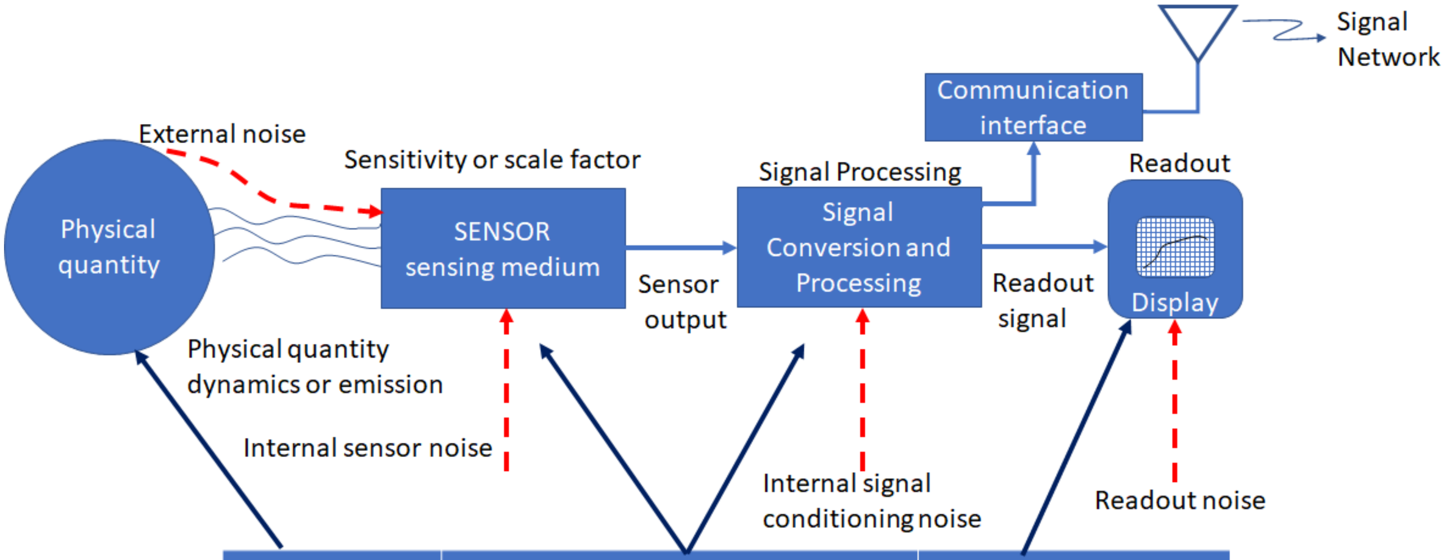
Understanding Sensors Part 1

Sensor Technology

by

Peter Kennedy

1.0 Overview: Sensors provide status information on our physical environment, homes, cars, and equipment we use. They are a part of nearly all walks of life and essential elements of control and safety systems. Understanding sensor technology (Part 1 of the course), and sensor networks with the fusion of sensor network data, (Part 2 of the course), is of increasing importance in most engineering and scientific applications. The sensor is a device that detects and/or measures the state of a physical quantity such as temperature, pressure, force, flow, or level. Measurements are converted to an observation media such as an electrical signal or mechanical, hydraulic, or pneumatic motion providing the status of the physical quantity's state. They may also interface directly to an actuator. The sensor measurement function is performed by several components that constitute a sensing system. This system, termed sensor node when integrated into a network, is comprised of a sensing element, signal conditioning and processing components, power supply and some form of output. This can be a simple display or meter, or now with the internet many sensors connect to a mini-processor or microcontroller for signal processing and then wireless communication over a network. This sensing system concept is illustrated in Figure 1.0 with the quantity, sensor and display medium listed for various sensors in the table below the block diagram. The sensor, when used as a threshold detector provides information of when a physical quantity's state has exceeded or gone below some significant threshold; i.e., true/false, on/off type information. For example, a bistatic thermal switch is normally designed for activation at a specified temperature. A wide measurement range is not necessarily required for detection and performance is generally based upon minimizing the number of false detections. Measurement is more complex. Performance is defined by accuracy which is a function of sensitivity, resolution, repeatability, dynamic range, response time, signal to noise ratio (SNR) and other quantities to be described. Because sensors are used for so many applications, there are many implementations; more than can be covered in one course. Transducer is a term often used in conjunction with sensors. It is defined as a device that converts the energy of a measured quantity, such as motion, light, or heat to some form of output energy for measurement and/or actuation. The automatic car lock control and keyless ignition are common examples. A radio frequency (RF) transponder is often located inside the plastic head or case of the key. Pushing the door lock/unlock button generates an RF signal coded to perform each function. With keyless ignition a receiver the car connected to the car computer must detect the key coded RF transponder signal for this purpose before it allows the ignition sequence to commence. Another example is the automatic vent opener used in green houses or sun rooms that is a pneumatic transducer containing a gas in a tube connected to a piston actuator. The gas senses heat and expands or contracts pneumatically moving the piston up/down or in/out.



Physical Quantity Measured	Sensor	Output Device (Display, Actuator)
Light Intensity	Photo-Resistor, Photodiode Photo-transistor, Solar Cell	Lights, Lamps, LED's, Displays, Fiber Optics
Temperature	Thermocouple, Thermistor, Thermostat Temperature sensing IC, MEMS temperature sensor	Heater, Fan
Force/Pressure	Strain Gauge, Pressure Switch, Load Cells, MEMS pressure sensor	Lifts & Jacks, Electromagnet, Vibration
Position	Potentiometer, Encoders, Resolver, Inductosyn, LVDT Inductive proximity sensor, capacitive displacement sensor	Motor, Solenoid, Display/meter
Speed, rate	Tachometer, Reflective/Slotted Opto-coupler, Angular Rate Sensors, Inertial Measurement Unit	AC and DC Motors, Stepper Motor, Brake
Sound	Carbon Microphone, Piezo-electric Crystal	Bell, Buzzer, Loudspeaker
Level	Float, conductance, diff pressure, acoustic (sonar)	Gauge, meter

Figure 1.0 Block Diagram of Sensing System and Types

Since there are many sensor types and applications, nomenclature describing sensor performance can also vary so it is important to fully understand the definition of performance specifications. The variation in complexity of a sensor covers a wide range, from the bi-static thermal switch to an optical video system requiring several cameras and inertial sensors as might be used by the military for situational awareness. Many applications require several networked sensors of the same and different types that provide information of several quantities for evaluation of a final condition. Sensor networking and fusion of multi-sensor data is described in Part 2 of this two-part course.

2.0 Sensor Types [1, 2]: Sensors provide a measure of an electrical, mechanical, thermal, chemical, light intensity, biological or magnetic energy quantity. The materials used for a sensor exhibit a response characteristic necessary for an application. The old spring scales used the spring constant that defined spring change in length as a function of applied force. The old mercury thermometers used the thermal expansion coefficient of mercury, enclosing it in a thin tube so as temperature changed it would expand or contract relative to a calibrated scale providing the temperature reading. Silicon is a material that has numerous micro-electro-mechanical systems (MEMS) applications. It is also widely used for optical detection and sensing applications in the visible and near-infrared (Vis/NIR), being able to convert photons incident on the detector surface to electrons; generating a current proportional to the incident optical energy. Applications include single element detectors, multi-element detectors as well as pixel elements in Vis/NIR cameras. It can also be used as a temperature sensor. Another area in optical sensing and detection receiving much interest in recent years are fiber optic sensors which will be described in Part 2 of the course. The majority of sensors operate in a passive mode, with an intrinsic material characteristic responding to a specific stimulus. A passive sensor combined with a sensor specific emitter can also operate in an active mode. The emitter transmits a signal to the environment and the passive sensor detects the return, processing it to obtain the desired information such as object detection and distance. This concept can be applied to small objects at short distances as well as objects at longer range; being the basis for RADAR and LIDAR systems. The landscape for sensing has changed dramatically in recent years. The transition to sensor elements imbedded within integrated circuits began nearly 40 years ago. Now with the proliferation of low-cost MEMS sensing elements which include a communication interface, many sensors can be connected in a distributed network via the internet of things (IoT). Large amounts of information can be quickly evaluated and used in sensor fusion algorithms increasing the knowledge base of information far beyond the scope of an individual sensor, and then transmitting the information rapidly to anywhere on the network.

3.0 Sensor Performance [3]: The performance criterion for a sensor is to provide an accurate measurement of the quantity being measured; accuracy quantified as the error between the measured quantity property and its actual value. Performance defined by accuracy, is a function of sensitivity, resolution, repeatability, dynamic range, response time, signal to noise ratio (SNR) and other quantities to be described as well as internal and external noise sources. If we are only trying to detect, then the metric would be to do so consistently without false detections. To estimate sensor performance, a simple performance model can be constructed from the basic algorithm:

measured quantity output = [sensitivity * physical quantity] + [sensitivity error * physical quantity] + [sensitivity * (external noise + internal sensor noise + internal sensor drift)] + [readout noise+ readout sampling error]

The physical quantity is in physical units (watts, deg/sec, etc.) while the measured quantity output is converted to a readout signal such as volts with the conversion factor being the sensor sensitivity. Terminology can vary depending on the sensor technology, sometimes the conversion factor is called the scale factor or also sensor responsivity. Sensitivity error could be an inherent non-linearity in the sensitivity or a variation of it with changes in the environment. In some cases, the baseline sensitivity is non-linear, a function of the measured quantity's magnitude as is the case for some temperature sensors, and errors in characterizing this non-linearity occur. External noise sources refer to noise measured by the sensor but not due to the quantity. For optical sensors this can be background and clutter, with some temperature sensors that require a current to generate a signal, the self-heating of the device can cause an error. Some sensors also have internal noise sources such as the shot noise and thermal random noise inherent to many optical sensors. Internal readout noise primarily refers to noise on the output signal, readout electronic noise, quantization, signal interference, power supply/ground loop noise, loading on the output that impacts signal level. The noise terms in general will have random and varying (drift) or constant (offset bias) deterministic components. Deterministic components can often be calibrated out. Precision measurements may require a more complex model of the quantity dynamics and sensor to estimate performance, as shown in Figure 2.0. This model may also be integrated with an error filter that further enhances performance; which will be described section 3.2. The sensor must be sensitive to the measured physical quantity; while having a minimal effect on the measurement. For example, old thermometers tended to be large. If the mass of the quantity being measured was not substantially greater than the thermometers, it could absorb enough heat to increase cooling of the quantity corrupting the measurement. For many applications, smaller sensors may minimize the mass impact on the measurement. In addition, sensor installation must be performed so it measures the property directly without impacting the measurement. For example, when mounting accelerometers to measure vibration one must be sure the mount does not add damping to the measurement or possibly even excite it beyond the true value.

3.1 Sensing Performance and Error [3, 4]: Sensor errors can generally be classified as systematic or random. Systematic errors are deterministic and can be modeled by an equation, algorithm, and/or or look up tables, therefore calibration or error correction algorithms can often be used for compensation, although there will still be some residual error. Noise is a random error that can be reduced by signal processing, such as filtering, usually at the expense of the dynamic response of the sensor. If the random noise can be statistically described by a probability distribution; more sophisticated filtering techniques can be used for compensation. Filtering will be described in the next section. In general, with most sensor applications, an error budget is defined to estimate performance based upon all the error sources. Many of the critical sensor performance characteristics along with noise sources that impact and limit a sensor accuracy are listed below. Their instantiation of how they impact a sensing system is often sensor type and application dependent.

Performance Characteristics:

- Signal to Noise Ratio (SNR): The amplitude ratio of the signal induced by the measured quantity to the signal induced by internal and external noise sources. It is one of the most important performance metrics. Performance varies proportional to SNR.
- Sensitivity: This term can be ambiguous in definition. It is the smallest variation of the input quantity detectable by the sensor; however, this could be interpreted as related to the sensor noise equivalent quantity (i.e., power, intensity, temperature, pressure, etc.) or to the ratio between the sensor output signal and the measured quantity; being equivalent to sensor scale factor or responsivity. In this case, with a linear response, it is the slope of the response curve, higher slope means higher sensitivity. For a non-linear response it is the slope of the response curve at a specific input.
- Resolution: The smallest detectable measured output variation. The ratio of the output to input resolution equates to sensitivity. Output resolution is often constrained so an estimate of achievable input resolution is (output resolution)/sensitivity. For a digital output, quantization error sets resolution. For example, the error for 8 bits is 0.39% or 1 part in 2^8.
- Sensor precision: Describes the reproducibility of the measurement or how closely several measurements of a steady state signal or quantity match, values close together indicate a sensor with good precision.

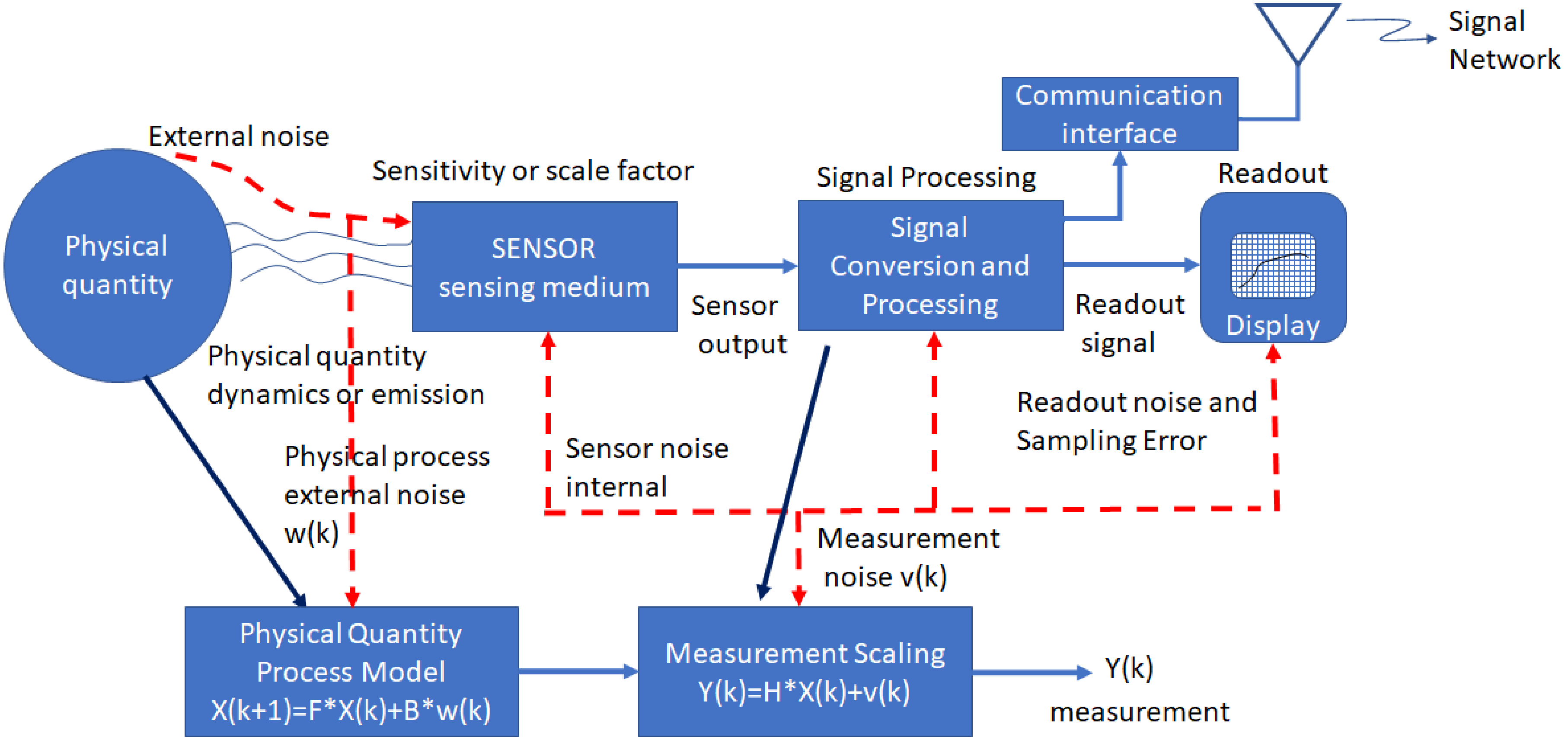


Figure 2.0 Sensing System Block Diagram and Corresponding Mode

- Sensor dynamic range: Is the ratio between the largest and smallest signal amplitude measured by the sensor.
- Repeatability: Is the sensors ability generate an identical output measurement to the same input, similar to precision.
- Response time: the time for the sensor to respond equates to bandwidth (BW); the sensor BW should be high enough to measure variations in the physical quantity property being measured.
- Saturation: The range of the output signal is always limited, when the output signal reaches its maximum level it saturates at that limit.

Noise:

- Internal Noise: Systematic deterministic, random, and offset noise inherent in a sensor due to the sensor material structure and impact of the environment on sensor structure that corrupts the measured sensor output. It includes hysteresis effects; this error occurs when output depends on both the present input and also the previous input values.
- External Noise: Clutter and background noise quantities external and similar to the quantity being measured that corrupt the true measurement.
- Sensor Drift and Bias: This can be short- or long-term offset error and is usually independent of the measured property. Long term drift can occur over several hours or even days.
- Sensitivity Error: The sensors sensitivity deviation from a specified linear response; both scaling and non-linear deviations of the response slope. It is the amount the output varies from a linear response over the full sensor measurement range.
- Readout Noise: Application dependent, generally several common issues such as sensor output loading by an interface drive circuit; interface circuit noise, power supply noise, and proper grounding to avoid ground loops.
- Readout Sampling Error
    - Sampling Frequency: The sampling frequency must at least greater than twice the highest frequency exhibited by the quantity, ideally much higher for good signal fidelity and no aliasing effects
    - Sampling Latency: Any delays associated with the sampled measurement
    - Sampling Quantization: Discussed above, effectively the resolution error
- Calibration: The compensation of systematic sensor measurement errors.

3.2 Improved Sensor Performance [5, 6]: Sensor performance can be improved by calibration, averaging, signal conditioning filters, and processing algorithm filters. Averaging improves accuracy by reducing random noise but will impact measurement bandwidth. Random noise can be reduced by the square root of the number of samples (i.e., $\sqrt{n}$) averaged. Multiple measurements are summed and divided by the total number of samples, n, reducing noise but also the effective information rate or bandwidth by ~n. Averaging, which is most effective with white noise, is a basic form of filtering and there are many other techniques that provide better performance. Besides averaging, signal amplification, and filtering can improve the sensor signal to noise ratio. Many sensor signals are very small and can be filtered and amplified by an operational amplifier circuit to produce a much larger voltage signal. If dominant noise sources come after the sensor amplifier or are outside the signal bandwidth it can enhance SNR considerably, otherwise if sensor noise dominates it is also amplified although the absolute margin between signal and noise increases. Besides amplification and filtering, the impact of noise and interference can often be reduced by impedance matching, and isolation between the input and output. Basic low pass filtering (LPF) can be implemented in hardware with a simple electronic circuit and many digital filtering algorithms can be implemented in processor software to reduce the impact of noise. High pass filters (HPF) reduce the effect of background and low frequency noise while notch filters will reduce noise at a specific frequency.

For more precision applications that include sensor signal processing, the Kalman Filter is a very useful and powerful tool for improving sensor measurement performance. The Kalman Filter has been utilized in numerous applications since its development in the early 1960s. It was named after one of its primary developers, Dr. Rudolf E. Kalman. With this filter, the process state is estimated using sensor measurements up to the present time. The filter is effectively a two-step algorithm, illustrated in Figure 3.0; using the present measurement to improve or correct the estimate of the current process state and then prediction of the next process state based upon a model of the physical process. When each measurement is available, the correction is determined by calculating the measurement error between the sensor measurement and the predicted measurement and multiplying this error by a gain matrix, often termed the Kalman gain. This gain is a weighted average of the error covariance, applying more weight to estimates with greater error in the state correction update error term. The weighting factors are based upon the calculated covariance of the process state and associated variance matrix of the process noise and measurement noise, theoretically assumed to be zero mean gaussian white noise distributions. The filter theoretical foundation is based in statistical and control theory. It has been derived in several ways since its original inception. From a sequence of measurements observed over time, corrupted with white noise, it provides optimal estimates of the process state variables more accurate than can be obtained with a single stand-alone measurement by using an estimate of the process covariance state at each time step. The filter can also be used in sensor networking as applied to sensor fusion described in Part 2 of the course.

Many forms of the filter exist including those for a non-linear process model. That version is often called an Extended Kalman Filter (EKF). Besides filtering, it can be used for prediction as well as process identification. A simple implementation of the filter is provided below with Figure 4.0 being a block diagram of the algorithm implementation.

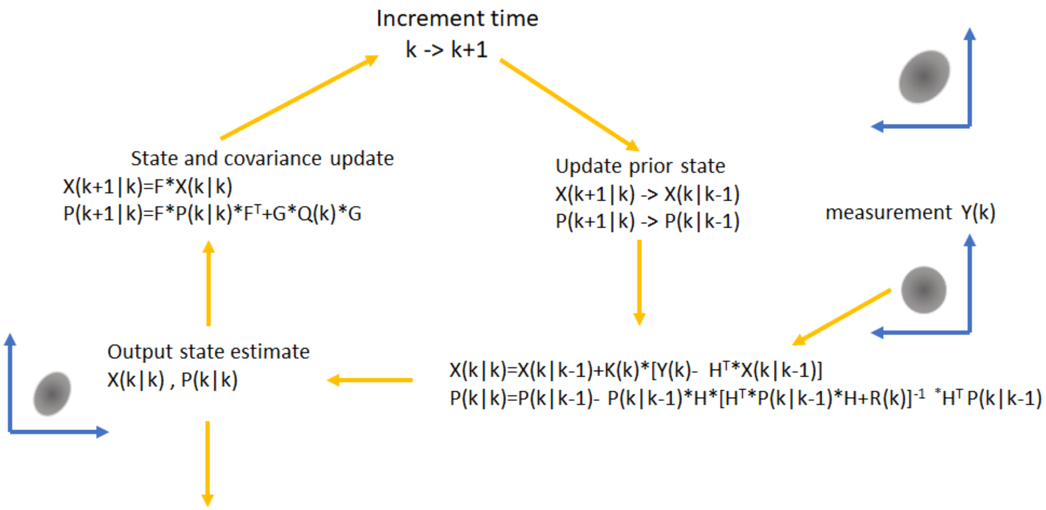


Figure 3.0 Kalman Filter Process

State Measurement update

$X_E(k|k) = X_E(k|k-1)+K(k)*[Y(k)- H^T*X_E(k|k-1)]$

State Update

$X_E(k+1|k) = F*X_E(k|k)$

State Measurement update

$X_E(k|k) = X_E(k|k-1)+K(k)*[Y(k)- H^T*X_E(k|k-1)]$

State Update

$X_E(k+1|k) = F*X_E(k|k)$

Total Update

$X_E(k+1|k) = [F-K(k)*H^T]*X_E(k|k-1)+K(k)*Y(k)=F*X_E(k|k-1)+K(k)*[Y(k)- H^T*X_E(k|k-1)]$

$K(k)$-gain matrix$=P(k|k-1) *H*[H^T *P(k|k-1) *H+R(k)]^{-1}$

$R(k)=E(v(k)*v(l)^T]*\delta_{kl}$ measurement noise covariance

Covariance Measurement update, often termed the Riccati Equation

$P(k|k) = P(k|k-1)- P(k|k-1)*H*[H^T*P(k|k-1)*H+R(k)]^{-1}*H^T*P(k|k-1)$

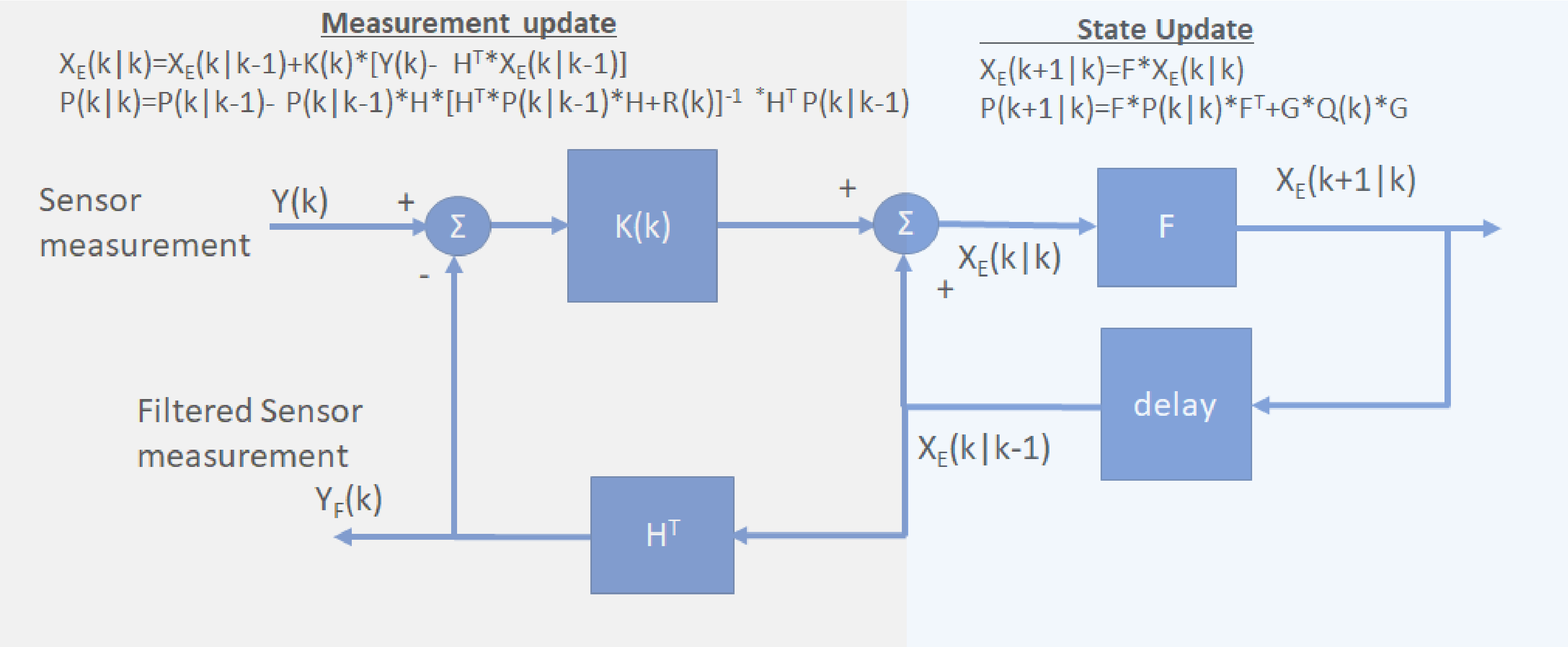
Covariance update

$P(k+1|k) = F*P(k|k)*F^T+G*Q(k)*G^T$

Total Update

$P(k+1|k) = F*[P(k|k-1)- P(k|k-1)*H*[H^T*P(k|k-1)*H+R(k)]^{-1} H^T P(k|k-1)]*F^T+G*Q(k)*G^T$

$Q(k)=E(w(k)*w(l)^T]*\delta_{kl}$ physical process noise covariance



Total Update

$X_E(k+1|k)=[F-K(k)*H^T]*X_E(k|k-1)+K(k)*Y(k)=F*X_E(k|k-1)+K(k)*[Y(k)- H^T*X_E(k|k-1)]$

$P(k+1|k)=F*[P(k|k-1)- P(k|k-1)*H*[H^T*P(k|k-1)*H+R(k)]^{-1} *H^T P(k|k-1)]*F^T+G*Q(k)*G^T$

$K(k)$-gain matrix$=P(k|k-1)*H*[H^T*P(k|k-1)*H+R(k)]^{-1}$

$R(k)=E(v(k)*v(l)^T]*\delta_{kl}$ measurement noise covariance

$Q(k)=E(w(k)*w(l)^T]*\delta_{kl}$ physical process noise covariance

Figure 4.0 Kalman Filter Algorithms Block Diagram

4.0 Sensor Types and Applications: Sensors can be classified in several ways; the quantity it measures, the sensor conversion or transduction process, sensor material/technology, sensor properties, and application. There can be many different types of sensing technologies used for one application. This section will provide an overview several sensor types and applications.

4.1 Temperature Sensing [7]: Temperature sensors are integrated with nearly every aspect of our daily living. Room temperature, outdoor temperature, car vehicle and engine temperature, temperature control in buildings, water temperature regulation, and control of refrigerators. The list is nearly endless when one considers all the applications in consumer, medical, and industrial electronics as well as all the variations in the temperature of the quantity measured; air or gas, liquid, and mass. Some sensors must be made to withstand very high temperatures. There are several types of temperature sensors: bistatic thermal switches (thermostat), RTDs (resistance temperature detectors), thermistors, thermocouples and semiconductor based integrated circuits (IC). Prior to the boon in consumer electronics, the mercury glass thermometer was the mainstay of most household temperature sensors. The mercury confined within a glass tube expands and contracts with temperature causing its level to rise and fall with a calibrated temperature scale behind the clear glass tube that provided the temperature for the mercury level. Most temperature sensors need to be in contact with the quantity being measured but there are some non-contact thermal radiation sensors that can measure the heat radiated from a source directly without contact. A description of several types of temperature sensors follows.

4.1.1 The Thermostat: This is an electro-mechanical temperature sensor or bi-metallic switch. It uses two different metals with differing thermal expansion coefficients, such as nickel, copper, tungsten or aluminum to sense temperature and convert it to mechanical motion. Strips of each metal are bonded together to form a composite bi-metallic strip to convert thermal to mechanical energy. Due to the difference in linear expansion rates of the two metals, a mechanical bending moment occurs when the strip is heated causing the strip to bend. The response time of these devices is generally slow due to the mass involved. The bi-metallic strip can be configured as a stand-alone electrical switch or for mechanical operation of an electrical switch in thermostatic controls. The switch contacts can be configured to carry low to high current. Thermostats are used in many applications for control of hot water heating elements in boilers, furnaces, etc. A simple diagram of the device concept is shown in Figure 5.0.

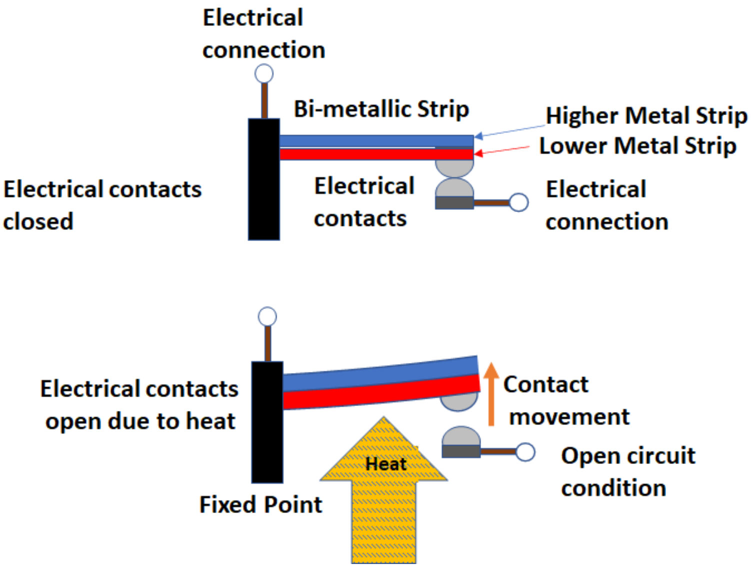


Figure 5.0 Thermostatic Switch

When cold the contacts are closed and current passes through the thermostat but as it heats up one metal expands more than the other and the bi-metallic strip bends opening the contacts. There are two versions of the bi-metallic strips; the snap-action that instantaneously switches on/off due to mechanical bending motion opening or closing the electrical contacts at a temperature setpoint, and the creep-action strip that changes position gradually as the temperature changes. Snap-action type thermostats are common in our homes for controlling the temperature set point of appliances and the heating system. They have a hysteresis so often open slightly above the setpoint and then close slightly below the setpoint. The slower creep types use a bi-metallic coil that slowly expands or contracts as the temperature changes. They often have more sensitivity to temperature changes and are used in temperature gauges and dials like the old meat temperature cooking thermometers.

4.1.2 The Resistance Temperature Detector (RTD): These are very stable and accurate temperature sensors when instrumented and implemented correctly. They are fabricated using high purity conducting metals such as platinum, copper or nickel. Platinum is very accurate and commonly used. They are wound into a coil with a nominal resistance proportional to length. As temperature changes, the material resistivity changes and causing the coil resistance to change which is the basis for RTD temperature sensors. The temperature to resistance relationship is non-linear with a positive temperature coefficient (PTC), but it can nearly be linearized using a Wheatstone bridge circuit.

RTDs produce very accurate temperature measurements and Platinum RTDs are available with a 100 Ω (PT100) and 1000 Ω (PT1000) resistance at 0°C. Platinum RTDs resistance vs temperature relationship is a near linear response over small temperature ranges, stable, accurate, repeatable, and has a wide overall temperature operating range, between -200 to +600°C. RTD elements have high thermal mass resulting in a slow response time to temperature changes when compared to thermocouples. The RTD thermal sensitivity is low; a change in temperature producing only a very small output change (i.e.,1Ω/°C) so that signal conditioning is required to obtain a high-fidelity measurement. An excitation current is required to flow through the RTD to observe a voltage output and resistance can be calculated from the known current. The impact of the excitation current needs to be accounted for since any variation in resistance due to self-heating of the resistive wires due to current flow causes measurement error. Compensation for lead length can be obtained by connecting the RTD to a Wheatstone Bridge network and connecting additional feedback wires to the bridge. Configurations include two (A), three (B), and four (C) wire as shown in Figure 6.0. The two-wire option works when lead length is short enough so that its resistance doesn't affect measurement accuracy. A three-wire adds an RTD probe carrying excitation current. providing a way to cancel wire resistance. Four-wire is the most accurate since separate feed and sense leads eliminate the effect of wire resistance.

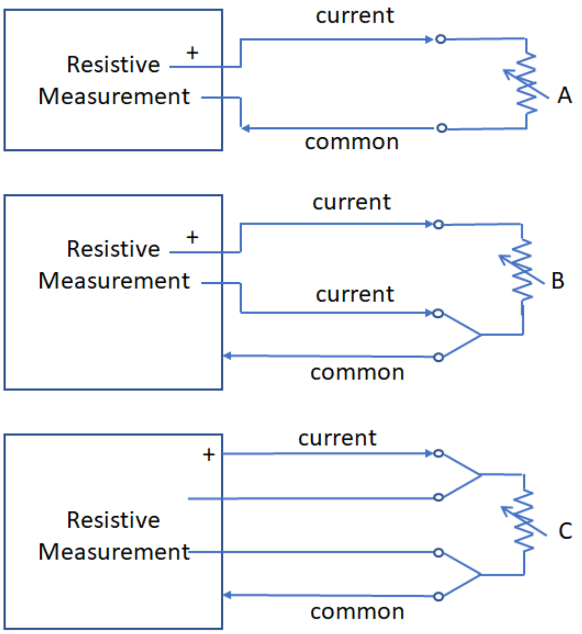


Figure 6.0 Two, Three, and Four Wire RTDs

4.1.3 The Thermistor: The thermistor, sketched in Figure 7.0, is a solid semiconductor with a high negative temperature coefficient (NTC) whose resistance varies with temperature change similar to RTDs.

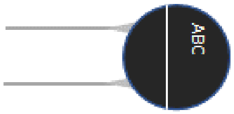


Figure 7.0 Thermistor

Polymer or ceramic materials such as oxides of nickel, manganese or cobalt coated in glass are used to fabricate thermistors. They can damage easily, are less expensive than RTDs but also less accurate. They are formed into small pressed hermetically sealed disks or balls. Their low mass allows for a relatively fast response time to a temperature change and have good repeatability and high responsivity. Most thermistors are available in two wire configurations. They have an exponential temperature resistance relationship requiring calibration and data interpretation when used directly. When integrated with a voltage divider circuit in series with a resistor and connected directly to an analog to digital converter (ADC), as shown in Figure 8.0, they have a near linear response over a limited temperature range. With a Wheatstone Bridge type arrangement, the current obtained in response to a voltage applied to the divider/bridge network is linear with temperature. An NTC thermistor's resistance decreases as the temperature increases. Thermistors are rated at room temperature for their resistive value, time constant and power rating for the current flowing through them. Like resistors, thermistors resistance values at room temperature range from 10's of MΩ to just a few ohms; with values in the kilo-ohms are often used.

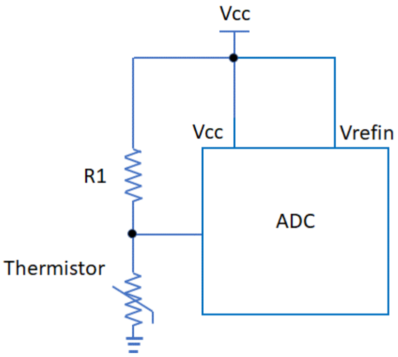


Figure 8.0. Typical Thermistor Interface

4.1.4: Thermocouples [8, 9]: These thermo-electric sensors find use in many applications. They are simple to use, small in size, require no external power supply, and have fast response times relative to most thermal time constants. They also have a wide temperature operating range from below -200°C to well over 2000°C. Thermocouples measure temperature based upon the Seebeck effect. A temperature difference across two dissimilar conductive metals, in isothermal contact at one end, generate a voltage proportional to the temperature difference. The device has two junctions, cold and hot, between which there is the temperature delta. The cold junction is maintained at a constant temperature with one wire connecting to the measurement circuitry. The hot measurement junction is where the two dissimilar metals are bonded together. When the two junctions are at different temperatures, a small voltage (milli-volts) develops at the hot junction between the two metals. Electronic amplification of the small voltage is required to provide the measured temperature from the sensor as shown in Figure 9.0 below.

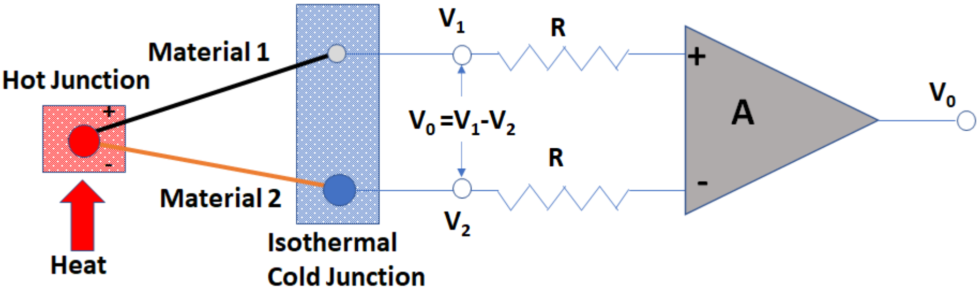


Figure 9.0 Thermocouple Measurement Configuration

When the junctions are at the same temperature V_1 equals V_2 so there is no output other than noise. The amplifier must be chosen for very low noise and drift. The small voltage generated by the sensor makes the signal suspectable to external noise and requires precise signal amplification. At the cold junction where the thermocouple wires connect to signal circuit copper traces the wires must be maintained at the same temperature otherwise another Seebeck Effect occurs requiring cold junction compensation. There are many types of thermocouples, designated by letters, fabricated from a variety of materials providing for different temperature ranges and sensitivities. The most commonly used is the K type followed by Types J and T. Table 1 shows characteristics of a few common types of thermocouples. With the large choice of materials and temperature range, international standards evolved, including thermocouple color codes, to allow a user to select the best thermocouple sensor for a specific application.

Table 1.0 Some Thermocouple Types and Characteristics [8]

Letter Code	Materials	Temperature range	Sensitivity (uV/°C)
E	Nickel Chromium/Constantan	-40°C to 900°C	68
J	Iron/Constantan	-180°C to 800°C	55
K	Nickel Chromium/Nickel Aluminum	-180°C to 1300°C	41
N	Nicrosil/Nisil	-270°C to 1300°C	39
T	Copper/Constantan	-250°C to 400°C	43
R/S	Copper/ Copper Nickel Compensating	-50°C to 1750°C	10
B	Platinum Rhodium	0°C to 1820°C	10

4.1.5 Semiconductor based ICs: Semiconductor based temperature sensors are usually based upon the temperature sensitivity of silicon. There are three types of semiconductor temperature sensors; bulk silicon resistors, junction semiconductor devices (diodes, transistors), and integrated circuits (IC). The bulk resistor is a piece of silicon; physically similar to a ¼ watt resistor, with a

PTC of ~ 0.7%/°C and nominal resistance between from 10Ω to 10KΩ. They can be used directly in a bridge circuit. Nominal values are specified at zero power so when used in a circuit, self-heating effects must be accounted for. With diodes or transistors, the potential across the junction varies with temperature at about 2.2 mV/ΩC. They have a fast response time and are low cost but accurate measurement requires calibration or implementation in pairs. A two terminal temperature sensitive current source integrated circuits (IC) can be used as a temperature sensor. Typical devices are the Analog Devices AD590 and the Texas Instruments LM334. They must operate within the device temperature limits. They are linear and easy to use, come in several levels of accuracy and do not require a bridge circuit. The output is current so it can be remoted without any voltage drop or induced voltage noise issues. Semiconductor and IC temperature sensors can be used as a local or remote temperature sensor. Local temperature sensors can use either analog (voltage or current) or digital outputs. Local IC temperature sensors are used to sense the temperature on printed circuit boards or the ambient air around it, for example equipment rack air channel flow. Remote digital temperature sensors use a digital output to transmit temperature data from the remote location.

4.1.6 Summary: Thermostats, RTDs, thermistors, thermocouples and semiconductor-based ICs are the main types of temperature sensors used today. Thermocouples are inexpensive, durable, and can measure a wide range of temperatures. RTDs also measure over a wide temperature range (although less than thermocouples) and provide accurate, repeatable measurements, but they are slower, require an excitation current, and signal conditioning. Thermistors are durable and small, but they are less accurate than RTDs and need more data corrections to interpret temperature. Semiconductor based ICs are used for many electronics and medical applications, can be used for implantation, can be integrated into networks and come in extremely small packages, but they have a limited temperature range. Table 2.0 summarizes these sensors listing temperature conversion approach and characteristics.

Table 2.0 Common Temperature Sensors Summarized [6]

Type	Temperature to Electrical I/O	Notes
Thermostat	Switch closure simple on/off output	Many types available covering a wide range of temperatures, contact configurations and current handling capacity
RTD	Resistance changes with temperature Positive temperature coefficient. Typical impedance (0°C) 20Ω to 2KΩ, typical sensitivities 0.1%/°C to 0.66%/°C dependent on material	Mid-level sensitivity and repeatability; good linearity over wide ranges; requires bridge or other circuit for typical interface
Thermistor	Resistance changes with temperature negative temperature coefficient. Typical impedance (25°C) 50Ω to 1MΩ, typical sensitivities at 25°C 4%/°C linearized circuit available with 0.4%/°C sensitivity	High sensitivity relative to common temperature sensors; Inherently non-linear (exponential function) but accurate linearized interface circuit available
Thermocouple	Low source impedance, typically 10Ω; voltage input devices; output shift is 10s of uV/°C outputs typically in millivolts at room temperature	Low voltage output requires low drift signal compensation, small size and wide temperature range are advantages, requires reference to a known temperature; non-linear response
Semiconductor	Voltage, current, or resistance functions. Voltage types (diodes) require excitation current. Current types (AD590) require excitation voltage; Resistive types (bulk silicon) may use either excitation	Many devices are uncalibrated and require significant signal conditioning; AD590 is calibrated, linear, and requires minimal signal conditioning

4.2 Optical Sensors [10]: Optical detectors are used in numerous applications and range from single element detectors to camera focal plane arrays. The spectral band and the application will dictate the sensor type used as follows:

- Visible (VIS) and near infrared (NIR) (0.38um-1um) detector applications using single or multiple element detectors usually fabricated from silicon such as silicon (Si) PIN photodiodes and phototransistor, Si Avalanche photodiodes (SiAPD), Si photomultiplier (SiPM). These devices convert light intensity impingent on the sensor surface to current when operated in a photoconductive mode and voltage in a photovoltaic mode. Another device that can be used is a photoresistor which is a resistive element whose resistance decreases when light intensity, within the sensitive spectral band, is directed on the sensor surface. The material used for this device in visible to NIR is cadmium sulfide or cadmium selenide. In the mid-wave infrared usually lead sulfide and lead selenide are used

- Visible (VIS) and near infrared (NIR) (0.38um-1um) camera applications (described in the next section) generally use a silicon-based material structured as either a:
    - charge coupled device (CCD) or
    - complementary metal oxide semiconductor (CMOS) sensor.
- Short wave infrared (SWIR) (1um-2.9um) detectors used in cameras are mostly indium gallium arsenide (InGaAs). Other detectors include germanium, lead sulfide, and lead selenide can be used in single or multi-element detectors
- Mid-Wave infrared (MWIR) (3um-5um) are either indium antimonide (InSb) or mercury cadmium telluride (HgCdTe)
- Long wave infrared (LWIR) cameras and detectors are:
    - HgCdTe photoelectric sensor arrays
    - Pyroelectric
    - Thermal micro-bolometers use temperature sensitive amorphous silicon, vanadium oxide

Optical sensors can be used in both passive and active modes and there are many applications for each, possibly more than any other type of sensor. The active mode requires an emitter and detector. Many commercial active applications work in the near infrared region such as TV remotes and LIDAR based proximity or object detection sensors. These devices emit and/or detect near infrared radiation to sense a particular quantity in the environment. For example, an NIR light emitting diode (LED) transmits a NIR beam to an object. It is reflected and a NIR photodiode detects the reflected beam off the object. When infrared light impinges on an IR photodiode receiver surface, a current is generated within the detector from which a voltage is obtained using a photoconductive amplifier configuration. The diagram of the setup is shown in Figure 10.0 below.

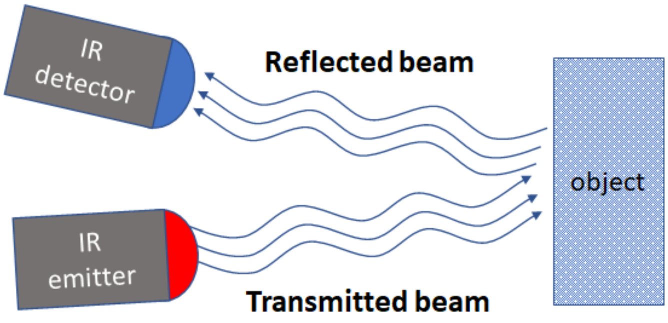


Fig. 10.0: Reflective Optical Detection Sensor

There is also a photovoltaic mode in which case the photodiode acts similar to a solar cell, converting light directly to a voltage. In the photoconductive mode, the photodiode is usually reverse biased at a voltage up to a specified maximum. Applying a reverse bias reduces response time, improves linearity, and provides some increase to responsivity but also introduces additional noise current to the generated photocurrent. Optical sensors can also be used for rate detection, counting, and coded identification. The slotted "U' channel detector is an example, as are the disc and read heads widely used in optical encoders (described in detail under positioning sensors); both convert analog rotary position to a digitized position and/or rotation rate. The channel detector has an LED in one leg of the channel and a phototransistor in the other leg. A slotted disk, attached to a rotating shaft, rotates in the channel between the legs producing an output when the LED transmits through a slot and no output between slots. This discrete digital output is a digital representation of the quantity being measured. The signal output by the phototransistor is considered a "1" bit when high and "0" when low. These pulses can be sent to a register or counter to determine the speed or revolutions of the shaft which is then displayed. Increasing the number of disk slots generates more output pulses per revolution of the shaft which then provides greater measurement resolution. A concept sketch is shown in Figure 11.0.

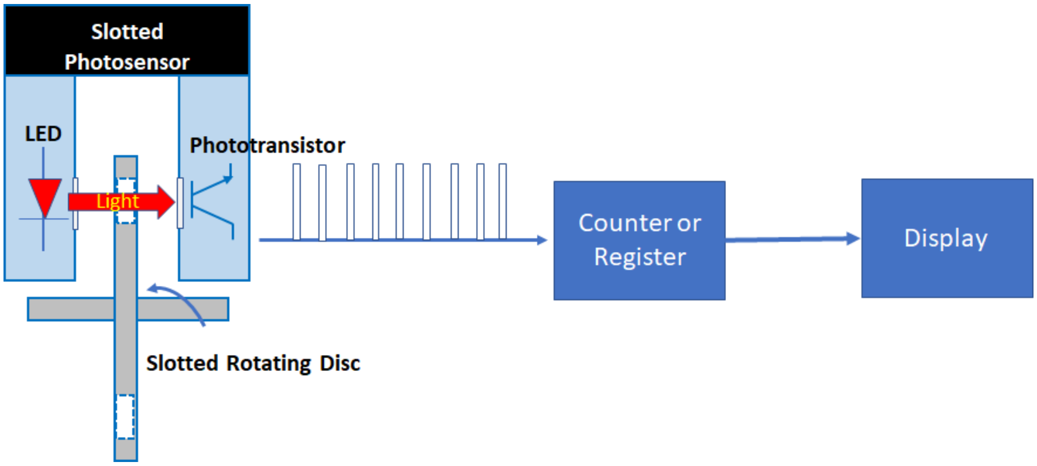


Figure 11.0 Slotted Optical Detector

Using bars or a pattern to modulate light to generate pulses has many applications including barcode readers. There are several types including laser scanners, pen type readers, and camera-based readers. Another application for optical sensors is motion detection. Many cameras now have software that detects motion within a scene. An often-used consumer application is for detecting motion around a home. Passive infrared detectors are used for both outdoor and indoor motion sensing. They will detect a change in temperature and are sensitive in the LWIR spectral region where humans and many warm-blooded animals emit radiation. Often a pyroelectric detector with a Fresnel lens in front to focus energy on the detector is used. The detector is split into two halves generating a differential signal to cancel out background. Optical sensors can be disturbed by the surrounding environment; stray radiation sources, ambient background light, atmospheric turbulence if outside will impact performance. Spectral filtering and short pulse modulation with a matched filter detector providing AC coupling at the detector can reduce the impact of many noise sources. Short pulsed modulation or frequency modulation with matched IR source and sensor components can be used to measure the distance or range to an object. LIDAR based systems can be used at both short and long range; out to several kilometers.

Ultra-violet sensors operate in the spectral region just below visible. These sensors can be used similar to the NIR sensors except they measure the intensity or power of the incident ultraviolet radiation. Electromagnetic radiation in this region has wavelengths shorter than visible radiation. UV sensors can measure an environments exposure level to ultraviolet radiation. UV radiation causes tanning and sunburn; too much exposure can be harmful to the skin and eyes. Charge coupled devices, discussed in detail in the next section, are often used to measure the intensity of incident UV light in scientific photography. Other uses include UV light detectors for solar irradiance, UV water and air treatments, and Germicidal UV detectors. Although not necessarily a direct sensor, UV radiation is also known to have a destructive impact on infections and viral germs. UVC radiation (200 nm – 280 nm) has effectively been used for decades to reduce the spread of bacteria, such as tuberculosis. The radiation, at a wavelength of ~222 nm, is finding uses during the COVID 19 Pandemic to kill the virus and disinfect an object, air or water.

4.2.1 Cameras [11]: With a camera, irradiance from a scene is captured by the camera lens optics, focused and directed onto an image sensor consisting of picture elements, or pixels that convert photons into electrical signals. The photoelectric sensing elements collect the light intensity, capturing photons within an intrinsic region of a p-type and n-type semiconductor sandwich structure. Charge carriers or electron hole pairs are created; with the number of electrons proportional to the number of photons absorbed. Electrons are converted into a voltage read by an analog to digital (A/D) converter generating a digital signal. Metal Oxide Semiconductor (MOS) technology is the basis for modern image sensors. The two main MOS technologies used to fabricate visible and NIR image sensors are CCD (Charge-coupled Device) and CMOS (Complementary Metal-oxide Semiconductor) as described in the paragraphs that follow.

CCD Technology [11]: Photons captured by a CCD pixel sensor surface are converted to electrons (charge). Each pixel sensor is configured as a p-doped MOS capacitive element storing the electron charge; effectively a capacitive well. A line of pixel terminates at an amplifier. At each update, charges shift one pixel closer to the amplifier, until a charge is amplified and output. The charges are converted to analog voltage levels, buffered, and finally converted to a digital signal by an A/D-converter external to the sensor. This process continues until all lines of pixels have had their charge amplified and output. The CCD technology was developed specifically to be used in cameras. Traditionally, CCD sensors have lower noise and higher-quality images, particularly in low-light conditions, compared to CMOS sensors. They also have had the advantages of greater sensor dynamic range, higher resolution and light sensitivity. In recent years, many of these differences have been significantly reduced as CMOS technology has advanced. CCD sensors require more electronic circuitry outside the sensor, are more expensive to produce, and can consume much more power than CMOS sensors. CCD sensors also require a higher data rate, since signals go through just one output amplifier, or a few output amplifiers.

CMOS Technology [11]: CMOS sensors convert pixel photon measurements to electrons on the sensor. All pixel measurements are processed simultaneously. An amplifier within each pixel results in a smaller, faster and more flexible sensor. Having an amplifier in each pixel, compared to the few amplifiers of a CCD, causes some blockage resulting in less photon capture area. This problem is resolved using micro lenses in front of each pixel that focus the light onto its photodiode so it is not blocked by the amplifier. Some CMOS imaging sensors also use back-side illumination to increase the number of photons collected by the photodiode. CMOS sensors can potentially be implemented with fewer components, use less power, and provide faster readout than CCD sensors. Modern CMOS sensors use improved sensor technology increasing sensor quality and light sensitivity that enhances image quality combined with lower costs. This is driving the use CMOS-based cameras for many applications previously dominated CCD cameras. Due to CCD sensor frame rate limitations at high resolutions, this trend is likely to continue.

Camera performance can be divided between camera sensor and optical system functions. Sensor parameters driving sensitivity and noise at a pixel level include quantum efficiency (QE), dynamic range, quantum well saturation, noise sources and signal to noise. Optical system parameters that are important include the optics aperture, focal length, sensor size and pixel density, exposure time, fame rate. To maintain a high signal to noise ratio (SNR) the camera sensor should have low noise and high sensitivity. A description of camera sensor performance parameters and noise follows.

- Sensitivity: A camera's absolute sensitivity threshold can be defined as the number of photons required to increase a pixel's value by one. It is the lowest detectable intensity signal above the sensor noise floor and can be divided into photon to electron conversion and collection efficiency.
- Quantum efficiency (QE_λ) is a measure the pixel photon to electron conversion efficiency, defined as the %photons converted to electrons at a specific wavelength; the ratio of photons in to electrons out. Responsivity (R_λ) is also a conversion measure but defined as amps/watt converting measured intensity (w/steradian) or irradiance (w/m^2) to photons per second based on the QE_λ to estimate electrons per second then converted to a current. Quantum efficiency is proportional to responsivity and related as:

$$R_\lambda = QE_\lambda \cdot \frac{\lambda \cdot q}{h \cdot c}$$

where λ-wavelength(m); q=1.6×10^{-19} coulombs; c-speed of light=$2.998\ 10^{8}$ m/sec and
; h-Planck's constant=$6.63\ 10^{-34}$ Joule-sec.

- Collection: The larger the pixel size, the more photon collection area resulting in more electrons generated. Collection begins at the camera aperture; camera optics transmission through the detector geometry impact collection efficiency. Larger pixels improve collection efficiency but on a negative side reduce resolution.
- Noise and Clutter: Noise sources that impact the sensor signal to noise ratio (SNR) include:
  - o Shot noise (η_{shot})- is inherent to the detector junction structure, a result of the discrete nature of the photon generation and the fundamental quantum physics of a photodetector.
  - o Dark noise (η_{dark}) – is both thermal and time dependent; due to thermally generated electrons producing a dark current within the pixel. It is a function integration time as well as temperature. Sensor cooling can be used to reducing this noise.
  - o Readout noise (η_{RO}) - occurs with every camera image frame, caused by the output electronics (amp and A/D converter) noise. It occurs prior to transmitting the digitized signal to the processor. Slower readout rates and precision clock pulses reduce this noise.
  - o Clock induced charge noise (η_{CIC}) – occurs primarily in CCD sensors, generated when clocking pixels to move the charge out of the sensor. Precision high resolution clocking pulses can reduce this noise source.

System noise can be estimated as the RSS sum of the noise current from these four sources. The photodetector response to an optical signal is a signal current generated by a photodiode as:

$$I_{\text{PIX}} = \varepsilon \cdot \eta_\lambda \cdot A_S \cdot \frac{QE_\lambda}{K_{P\lambda}} \cdot \tau_{INT} \cdot \frac{A_{CAM}}{R^2} \cdot e^{-\alpha \cdot R}$$

with: ε-emissivity, α-atmospheric transmission constant (km^{-1}),
τ_{INT} -camera integration time (s), QE_λ -quantum efficiency,
$K_{P\lambda}$ =(h*c)/ (λ*q) , A_{CAM} – camera collection aperture area (m^2),
A_S -source area (m^2), R- range (km); η_λ-integrated scene (source) spectral radiance weighted by atmosphere (W/sr-m^2).

The scene spectral radiance can be due to many sources within the camera FOV. A rough estimate of SNR, in terms of current, is given by:

$$SNR = \frac{I_{PIX}}{\sqrt{\eta_{SN}^2 + \eta_{DN}^2 + \eta_{RO}^2 + \eta_{CIC}^2}}$$

SNR is often measured in decibels (dB); the higher the SNR the better the system performance. Another major signal detection issue is clutter as it relates to extracting image features and is generally removed by spatial filtering. When comparing sensors, there are trade-offs between many sensor characteristics that must be balanced for best performance. A description of other camera parameters that impact performance follows:

- Saturation Capacity: As a pixel photodiode has a finite well capacity for holding charge which is its saturation capacity. Higher saturation capacity equates being able to capture a wider range of intensity. Larger pixels and surface areas will usually have greater saturation capacity.
- Dynamic Range: defines the range of magnitude over which a sensor can measure a signal; so is a function of saturation capacity. It is the ratio of the maximum detectable signal at saturation to the minimum signal the sensor can measure. A high dynamic range enables a sensor to measure a signal over a wide range of ambient lighting conditions.
- Gain: with a CCD camera is the conversion factor between electrons to Analog-Digital Units (ADUs) or counts. It is expressed as the #electrons converted to a digital number, or electrons per ADU (e-/ADU) and is #electrons required to induce a count change in 16-bit ADUs.
- Bit Depth: The bit depth of a camera indicates the numerical resolution to which each pixel is capable of measuring. It is effectively the pixel's quantization level.
- Exposure time, Frame Rate, and Shutter types: Exposure time is the interval a scene is exposed to the camera sensor. Frame rate is the frequency with which a scene image is recorded. Exposure time cannot exceed the frame time; often being half the frame period. Long exposure times dictate low frame rates; short exposure times allow high frame rates. In many longer wavelength cameras, exposure time is referred to as integration time performed electronically.
- Resolution, Pixel Size, and Optical Format: are inter-related parameters whose characteristics must be matched to achieve a required performance. Optical format is the sensor physical dimension measured across the sensor diagonal and defines the required lens coverage of the projected image to completely illuminate the sensor. For a given optical format, deceasing pixel size increases resolution but reduces collection capacity and efficiency.

4.3 Object Detection, Proximity, and Distance [12, 13]

Proximity sensors provide information on an object's presence and proximity without contact with the object. Some inherently provide distance information while others simply infer it from the measurement setup; detection occurring at some predetermined or maximum sensing distance. They are used for monitoring and control in production assembly applications, process control systems, alarms, and machines. Being contactless sensors improves reliability and the operating life of the sensor. There are many types of proximity sensors that include Inductive Proximity sensors, Capacitive Proximity sensors, Ultrasonic proximity sensors, and the photoelectric sensors, already described under optical sensors.

4.3.1 Inductive Proximity sensors [13]– They are also called an Eddy current sensor and detect the presence of an object within a close proximity without actually measuring its distance. An oscillator generates a magnetic field around the inductive sensor. When a ferromagnetic material, such as metal plate or metal screw, is placed within the magnetic field, eddy currents are created on the objects surface which change the reluctance of the magnetic field lowering the oscillation frequency effectively changing the inductance of the coil. The proximity sensors signal processing circuit detects this change in impedance producing an output voltage. The existence and characteristics of the eddy current field are defined by Faraday's Law of inductance which states that a loop of wire placed in a magnetic field will generate a magnetic flux through the wire is proportional to the number of magnetic fields lines through the loop. When the flux changes, due to the loop moving or blockage in the field from a moving object, Faradays law states the wire loop acquires an electromotive force. In an inductive sensor, a coil is wound around an iron core within an electromagnetic field to form an inductive loop. These sensors are only sense metallic targets.

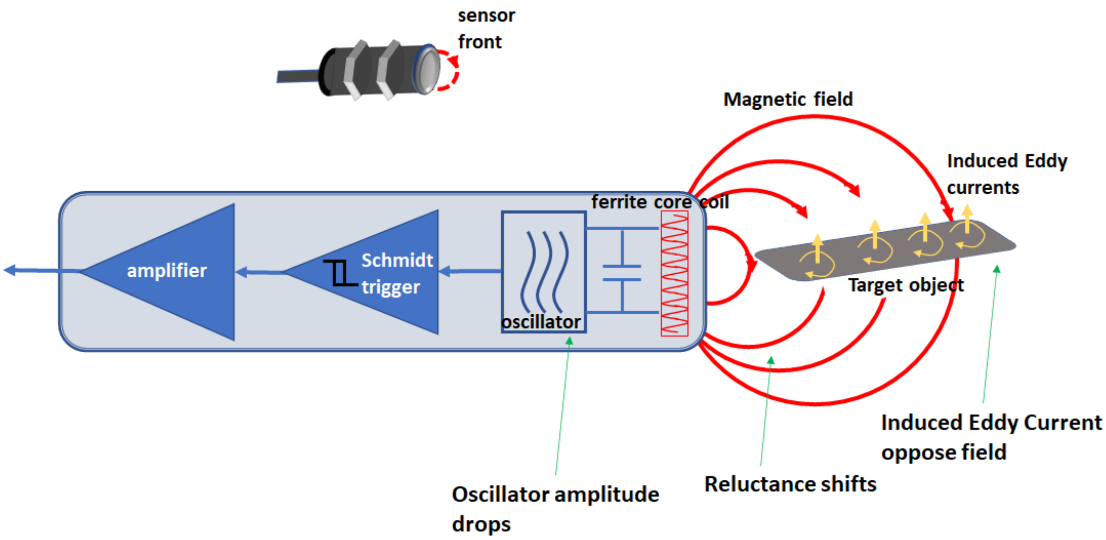


Figure 12.0 Inductive Proximity Sensor [13]

An inductive proximity sensor, as shown in Figure 12.0, has four main components; the oscillator that generates the electromagnetic field; a ferrite core coil to generate the magnetic field; the detection circuit, a Schmidt trigger as shown in the figure, responding to any change in the field when an object enters it, and the output amplifier circuit. The output circuit produces drive signal to a relay type device with either normally closed (NC) or normally open (NO) contacts. Inductive proximity sensors sense metallic objects in front of the sensor head and can be used in dirty or wet environments. The sensing range of proximity sensors is very small, typically 0.1mm to 12mm. Inductive proximity sensors are used for traffic flow control; controlling traffic lights at junctions and cross roads. Rectangular inductive loops of wire buried beneath the road surface detect the change in inductance a vehicle passes over due to the vehicle's metallic body which activates the sensor alerting the traffic light controller a vehicle waiting. The sensor does not detect non-metallic objects and are "omni-directional" since they will sense a metallic object in roughly a hemispherical region in front of the sensor head, therefore are subject to clutter detections.

4.3.2 Capacitive proximity sensors [13]: These sensors detect metallic and nonmetallic targets. They sense through nonferrous materials, solids and liquid mediums, and are a good sensor for applications such as tank liquid level detection. In capacitive sensors, as shown in Figure 13.0, two conduction plates at different potentials are enclosed in the sensor head and situated to function like an open capacitor with air creating the insulator. There is little capacitance between the two plates without a target. As with the inductive sensor, the plates are linked to an oscillator, a Schmitt trigger, and an output amplifier. When a target is within sensing range the capacitance of the two plates increases causing oscillator amplitude to increase. When a threshold is exceeded the Schmitt trigger state changes creating an output signal. The inductive and capacitive proximity sensors differ in that with the inductive sensor the oscillator amplitude decreases when a target is present but with the capacitive design the amplitude increases. Since sensing requires charging the plates, capacitive sensors have a slower response time than inductive sensors with bandwidths ranging from 10 to 50 Hz. Sensing distance is 3 to 60 mm. As they detect most types of materials,

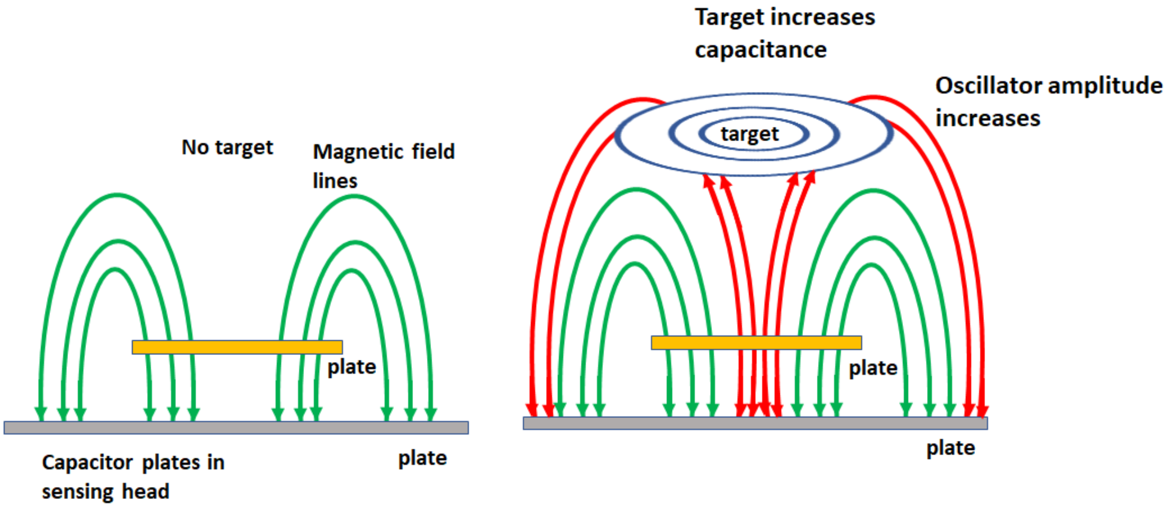


Figure 13.0 Capacitive Proximity Sensor [13]

capacitive sensors are more suspectable to false target detection. If the intended target is a ferrous material, an inductive sensor is probably better.

4.3.3 Ultrasonic proximity sensors [13]: These sensors are used in many automated production processes and employ sound waves to detect objects. A transmitter converts electrical energy into sound waves in the ultrasonic range, above 18 kHz. The echo returns are received and converted to electrical energy, measured and displayed. They are a good match for a many of applications that include long-range detection of clear glass and plastic, distance measurement, continuous fluid and level sensing and control. Sensor configurations are similar to those used in photoelectric sensing: described previously; through-beam, retroreflective, and diffuse versions. Ultrasonic diffuse proximity sensors use a sonic transducer to emit a sequence of sonic pulses that are reflected off a target object. The returns are detected at the receiver and once the reflected signal is received, the sensor signals an output to a control device. Figure 14.0 illustrates the principle of operation of the ultrasonic sensor, the transmitter sending out sound waves and the receiver sensing the echo returns detecting the presence, position, and movement of an object. Typical sensing range is a few meters. The standard diffuse ultrasonic sensors provide a simple object detection signal, while some produce analog signals with amplitude indicating distance. Ultrasonic retro-reflective sensors also detect objects within a specified sensing distance by measuring propagation time based upon the equation shown in Figure 14.0.

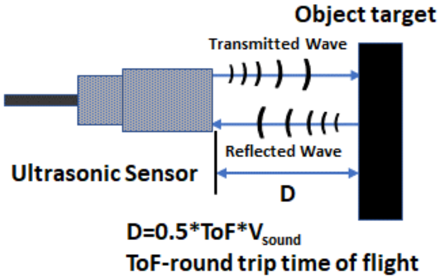


Fig. 14: Principle of Ultrasonic Sensor

Ultrasonic through-beam sensors have the emitter and receiver in separate housings as with through-beam optical sensors. When an object disrupts the sonic beam, the receiver triggers an output. The movement of ultrasonic waves vary with transmit media shape and type. They travel straight in a uniform medium. At the boundary between different media, however, they are reflected and transmitted back. A human body in air causes reflection that is easily detected. The transmission of ultrasonic waves is affected by multi-reflection that occurs when waves are reflected multiple times between the sensor and object. Most have an adjustable limit zone defined by a minimum and maximum sensing distance. A non-detection zone in front of the sensor defines the minimum distance that can be detected without interference due to multiple reflections prohibiting reliable detection.

4.4 Position Sensors

4.4.1 Linear Variable Differential Transformer: The Linear Variable Differential Transformer (LVDT*)* is an inductive position sensor that measures displacement motion. In principle, it functions similar to the AC transformer and provides very accurate linear displacement measurements. The LVDT output is proportional to the position of a moveable core within a hollow tube with three coils wound around the tube, one being the primary coil and the two coils to each side of the primary coil are the secondaries that are identical and connected electrically in series but 180° out of phase. The moveable ferromagnetic core connects to the object being measured. It slides within the tubular body of the LVDT. An AC reference excitation voltage, typically 2 – 20V rms @ 2 – 20kHz, is applied to the primary winding which induces an EMF signal into the two adjacent secondary windings. The LVDT concept is illustrated in Figure 15.0, showing the basic structure and electrical coil configuration. When the iron magnetic core is centered in the tube a null condition occurs since the induced emf's in the two secondary windings, being 180° out of phase, cancel each so the differential voltage between them is zero.

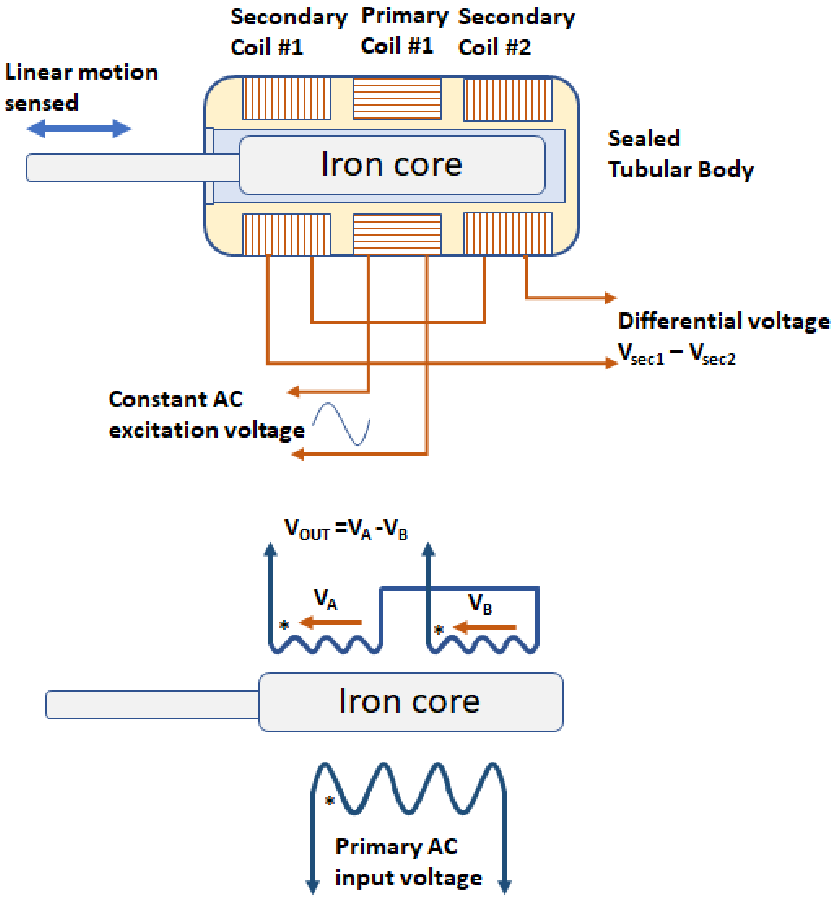


Figure 15.0 Linear Variable Differential Transformer

As the core is displaced from the null position, the induced voltage in one secondary becomes greater than that of the other secondary generating a differential output voltage. The output signal polarity depends on which side of the null position the moving core has been displaced. The output signal amplitude depends on how far from the null position it has been displaced; the greater the displacement the higher the amplitude of output signal. The differential voltage amplitude varies linearly with position displacement from null. Sliding the core from one end of the hollow tube to the other going through center results in the differential output voltage varying from maximum at one polarity to zero at null and back to maximum at the other polarity. As the LVDT output voltage is generated by an alternating current (AC) the change in polarity is actually a change in the phase angle by 180°. The advantages of the LVDT include high sensitivity, excellent accuracy and resolution, and linearity.

4.4.2 A Hall-effect sensor [14, 15]: These sensors measure the magnitude of an external magnetic field. They generate an output voltage proportional to the applied magnetic field strength. Applications include proximity sensing, positioning, speed detection, and current sensing. The sensor is fabricated from a thin rectangular metal strip of p-type semiconductor material such as gallium arsenide (GaAs), indium antimonide (InSb) or indium arsenide (InAs). A current or beam of charged particles is created flowing through the strip. In the presence of a magnetic field; as the beam passes through the field it is acted upon by forces that deflect the beam from a straight path. Electrons in the beam are deflected towards one edge of the strip which becomes negatively charged while the opposite edge of the strip becomes positively charged with holes. A voltage gradient is generated across the short side of the strip called the Hall voltage. A simple diagram of the concept is shown in Figure 16.0 with an external magnet providing the magnetic field.

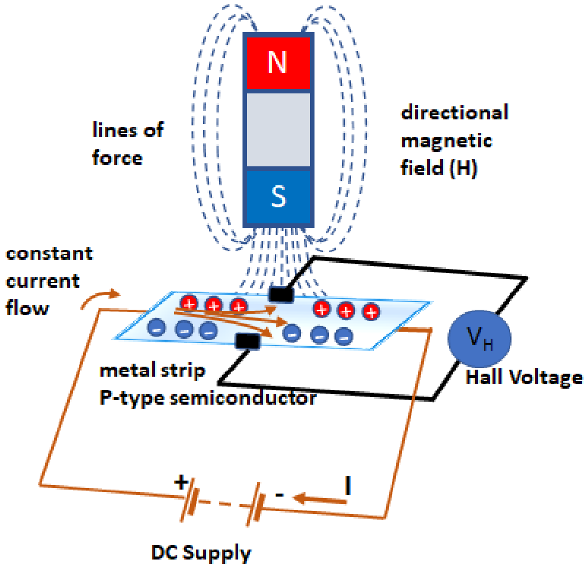


Figure 16.0 Hall Effect Sensor

Hall-effect sensors can detect static as well as the changing magnetics fields giving them an advantage over inductive sensors which can detect only changing magnetic fields. There are two kinds of Hall-effect sensors: linear with the output voltage linearly dependent on magnetic flux density; and threshold, in which there is a step decrease in output voltage at some magnetic flux density. The hall voltage is determined as [15]:

$$V_{\mathrm{H}} = R_H \cdot \left(\frac{I}{t} x B\right)$$

with V_H-Hall voltage (volts); I-current through sensor (amps); Rh=-Hall Effect coefficient; t-sensor thickness (mm); and B-magnetic flux density (telsas). The linear relationship between the Hall Voltage and magnetic flux density can be observed in the equation.

In many applications the sensor magnetic field is generated by a single permanent magnet connected to a rotating or sliding shaft or moving device. Motion geometries include; head-on, sideways, push-pull, or push-push. For any configuration, sensitivity peaks when the magnetic lines of flux are perpendicular to the Hall sensing area. Sensor linearity improves with greater magnetic field strength. Two common configurations with a single permanent magnet are head-on and sideways as illustrated in Figure.17.0.

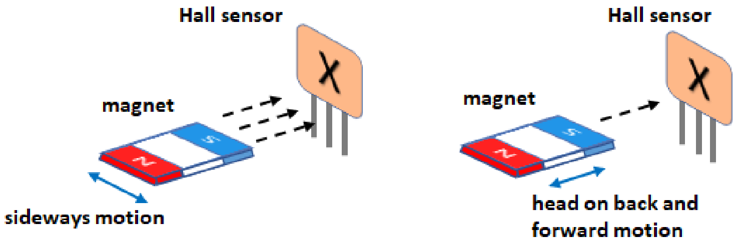


Figure 17.0 Sensing Configurations

With head-on motion the magnetic field must remain perpendicular to the hall sensing surface as the magnet and sensor translates towards and away from each other. This approach generates a linear output signal, V_H based on the strength of the magnetic flux density which varies as a function of distance between the magnate and sensor. The closer the magnet is to the sensor, the stronger the magnetic field and the higher the output voltage. With sideways motion the magnate moves across the sensing surface of the Hall effect sensor. This geometry is useful for detecting a magnetic field as it travels across the sensor face maintaining a fixed offset between the magnetic field source and the sensor, for example sensing motor rate or phase. Depending on the magnetic field position both a positive and a negative linear output voltage can be generated allowing for directional motion sensing. The Hall effect can determine magnetic pole polarity and magnetic field magnitude. As switches, the sensors are normally off when there is no magnetic field and turn on in the presence of a magnetic field. They are environmentally robust and have numerous applications especially in motion control.

4.4.3 Rotary Position Sensors [12, 16, 17]: These sensors determine the position or displacement relative to an instrument or device reference; often referred to as relative angle or position sensors. Angular position feedback is required for many motion control and robotic applications. Encoders, resolvers (or inductosyns), potentiometers, and rotary variable differential transformers (RVDT), the rotary version of the LVDT described previously, all generate angular position data. The most often used relative angle position sensors are resolvers and encoders capable of continuous 360° rotation. There are both rotary (angular measurement) and linear (linear displacement) versions of these devices. Each sensor type has inherent tradeoffs to be evaluated during the selection process. Resolver and encoder accuracy are normally specified in terms of arc-minutes or arc-seconds. There are 60 arc-minutes per degree. The component accuracy specification (maximum measurement error over full measurement range) depends on the configuration. With a completely assembled unit, accuracy includes both the electrical error as well as mechanical rotary alignment errors due to shaft and bore tolerances, bearing run-out, and winding perpendicularity. In many custom gimbal designs for precision pointing applications, a resolver or encoder is purchased as two pieces, for example for the resolver a rotor and stator, while with an encoder the rotating optical disc and a stationary read head. These are integrated by the user into the motor drive for each rotation axis. The accuracy quoted by a manufacturer addresses only the electrical and optical (encoder) error while the user mechanical design tolerances determine the mechanical error. Mechanical errors that are repeatable can often be reduced by measurement and calibration.

4.4.3.1 Resolvers: Resolvers provide absolute angular position with electrical errors ranging from 0.1° to as low as 10-20 arc-seconds with multi-speed resolver designs. Accuracy depends upon size and angular range. Some encoder manufacturers include AXSYS Technologies, Inland Kollmorgen, Admotec, MOOG, Dynapar, and Honeywell. Resolver construction similar to motor, rotor and stator coils produce sine/cosine signals from whose phase difference is proportional the absolute angle measurement. This construction makes them very robust in the presence of shock and vibration disturbances. They are also possibly the least sensitive to temperature variations when compared to encoders and potentiometers. Because of this they are often used in military applications. Small resolvers can be relatively inexpensive; however larger pancake resolvers with large inside diameters as used in many tactical gimbals can be quite expensive with long lead times. The resolver stator consists of two windings positioned at right angles to each-other. The rotor has a third winding that is energized with a sinusoidal signal and rotates relative to the stator. The signal in the rotor winding induces a signal in both stator windings whose magnitude varies as a function of the rotation angle; for a single speed, one sinewave per 360° rotation, an N-speed N sinewaves per 360° rotation, increasing accuracy. The voltage induced in one stator winding is in quadrature to the voltage in the other winding. Quadrature refers to the 90^o phase relationship between the signals of the two windings. The output voltage ratio of the two stator windings is proportional to the absolute angular position (arc-tangent of the ratio). A resolver to digital converter (RDC) is normally used to convert the analog resolver signals to a digital output read by a computer with the RDC having N-bits of quantization or resolution.

4.4.3.2 Encoders: Encoders are fabricated based upon both optical and magnetic designs. Some encoder manufacturers include Heidenhain, Renishaw, Netzer, Celera MicroE Systems, US Digital, and Gurley. The most basic optical encoder contains a bar type patterned rotating optical disk that interrupts transmission between an optical transmitter and a receiver. A pulse sequence is generated, similar to the detector described in section 4.2 on optical sensors, whose count is proportional to angle relative to a reference. The disk is normally glass substrate making them less desirable for high disturbance environments. With actual optical encoders, the optical source is typically a Light Emitting Diode (LED) and the receiver a photo detector. The disk has coded patterns, typically binary, grey code, or binary coded decimal (BCD), of transparent and opaque sectors that modulate the light measured by the read head photo detector. Encoder resolution is proportional to the number of pulses or counts per single revolution; also, being specified as lines per revolution, a reference to the number of opaque lines etched into the optical encoder glass disk radius. The number of counts (pulses) obtained as the disk rotates is a measure of the angular position, determined by a digital counter. The simplest and least expensive optical encoders are incremental and do not provide an absolute angular position. They have to be homed to an absolute reference position periodically to obtain as estimate of absolute position and will lose the position reference upon power reset. The most common type of incremental encoder has two output channels, channel A and channel B, that sense position. Two code tracks with sectors positioned 90° out of phase provide quadrature signals that can be used to detect position and direction of rotation. Counting both channels concurrently also enhances the encoder’s resolution. This configuration is termed a quadrature or sine wave encoder and uses two photo detectors offset from each other by 90^o in the read head, thereby producing sine and cosine signal outputs. Resolution of these devices is a function of the number of line segments, quadrature configuration, and interpolation algorithms used. Absolute encoders are normally larger and use more complex disk patterns containing 4-6 tracks. They generate a unique word (i.e., BCD or gray code) providing position and direction information for every angular position of the shaft. Each track has a photo detector to read a unique coded position value for each angle. The number of disk tracks corresponds to the binary bit-resolution of the encoder (i.e., 6 tracks equates to 6-bit resolution). The absolute encoder retains its position information and does not require a homing reference. Finally, in an encoder the quadrature error is the signal phase error between the two quadrature channels, commonly expressed as a percentage of the electrical cycle. For example, the distance between the leading edge of channel A and the leading edge of channel B is $L/4 \pm L/10$, where L is the cycle duration. The tolerance percent is essentially the divisor (i.e., 10). When encoder manufacturers don’t specify instrument error separately, this tolerance is typically the sum total of both the instrument and the quadrature errors. Instrument errors include quadrature error, sub-divisional error, and installation misalignments. Some alignment errors can be reduced using dual read heads

A magnetic encoder uses a magnet in place of the LED, a magnetic pickup sensor in the read head with the disk is replaced by a rotating gear made of ferrous metal. As the gear rotates, the teeth disturb the magnetic flux emitted by the permanent magnet causing the flux field to expand and collapse. These changes are again sensed as pulses by the magnetic pick-up detector. Magnetic encoders come in both absolute and incremental versions. They normally have less resolution than optical encoders but are more robust in a high disturbance environment. When the encoder is purchased as an assembled unit; shaft, bearing, enclosure, disk, and transceiver, the instrument error normally refers to the electrical and mechanical rotary error due to manufacturing tolerances including shaft and bore tolerances, bearing run-out, and alignment between the emitter, disk, and detector.

4.4.3.3 Potentiometers [12]: Potentiometers are normally used in low cost and low accuracy applications. Some potentiometer manufacturers include Bourns, Betatronix, Vishay, ETI Systems, and TT Electronics. They are effectively a variable resistor providing position information based upon an internal voltage divider. These devices provide an absolute measure of position with the measurement error being some percent of full scale and therefore proportional to the magnitude of the angle. They require an applied voltage but no other type of external excitation. The electrical angle specifies the potentiometer’s measurement range. A wiper contact connects to a rotating or sliding mechanical shaft. When the shaft moves the resistance between the wiper/slider and the end terminations varies generating a voltage proportional to angular or linear position. For single turn potentiometers, the electrical angle is slightly less than 360^o. Potentiometer linearity error is the mechanical and electrical rotary error inherent in the component design including shaft and bore tolerances, bearing run-out, wiper positioning, and element concentricity. It is typically given as a percentage of the output value. The device normally has three connections. A DC reference voltage across to the two outer terminals that connect to the resistor track end terminations determines the measurement voltage span. The center or middle terminal (T2) provides the output voltage signal, proportional to position. The signal voltage is derived from the divider circuit the slide/wiper terminal makes with the resistor track as illustrated below in Figure 18.0. The potentiometer output is an analog voltage. The potentiometer is low cost and easy to use, but wears, has limited frequency response, low accuracy and repeatability. The resistive track of most common potentiometers is fabricated from carbon film which is noisy, degrading further with wear. Smooth low friction electrically linear resistive tracks fabricated from a conductive plastic type polymer film, can now be obtained, that provide high accuracy low noise performance with long life and excellent resolution

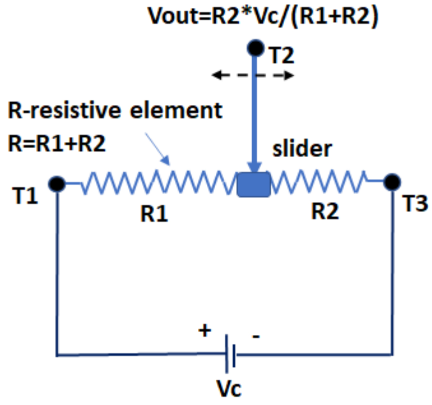


Figure 18.0 Potentiometer Circuit

4.4.3.4 Rotary Position Sensor Performance [17]: When determining angular position sensor requirements, there are several performance factors to evaluate. Since each component type can be mounted directly to the rotating axis, only the inherent component errors are considered. Typically, the resolver uses an RDC and the potentiometer an ADC to convert an analog signal to a digital signal for processing. A brief review of the formulas required to estimate component error quantities is provided.

Component Resolution Limitations: Many controllers available today have integrated ADC with resolutions of 10, 12, or 16 bits. When considering an analog controller for a resolver or potentiometer, resolution can be estimated in arc-minutes as:

$$R_{res} = \frac{360 \times 60}{CPR}$$

where the counts per revolution, CPR, for each sensor type is given in Table 3.0.

Table 3.0 Counts per Revolution for each Rotary Position Sensor Type

Rotary Position Sensor	Counts per Revolution
Resolver or Potentiometer	$CPR=2^{BITS}$
Incremental Encoder w/Quadrature	CPR=4*LPR

where $BITS$ A/D converter resolution or encoder resolution (bits)

LPR lines per revolution

Mechanical and Electrical Component Error: The assembled rotary position sensor has inherent error resulting from subcomponent and assembly tolerances. There is always some mounting angular and offset misalignment relative to the true rotation axis resulting in run-out error. But often this can be measured and the effect reduced with calibration. For potentiometers, this value reflects a worst-case condition occurring at one extreme of the electrical angle. It is not the potentiometer's linearity but the resistance tolerance on the peak value at the end of each revolution of rotation. Approximate formulas for different sensors are provided in Table 4.0. The encoder error does not assume any interpolation.

Table 4.0 Mechanical and Electrical Error for each Rotary Position Sensor Type

Rotary Position Sensor	Error
Resolver	$E_{me}=R_{acy}$
Potentiometer	$E_{me}=A_e*Ep_{LIN}*60$
Incremental Encoder w/Quadrature	$E_{me}=E_{instr}+E_{quad}$

where $$E_{\text{quad}} = \frac{K}{LPR} \cdot (60 \cdot 360)$$

and

E_{me} - Mechanical and electrical error
R_{acy}- Resolver accuracy (arc-minutes)
A_e - Electrical angle for potentiometer
Ep_{LIN} - Potentiometer linearity
E_{instr} - Instrument error (scale graduation, subdivisional error, alignment)
E_{quad} - Quadrature error
K - quadrature tolerance

The total rotary position error can be estimated as $E_{TOT} = R_{res} + E_{me}$.

4.5 Inertial Sensors [18]

Inertial sensors refer to sensors that provide an earth or gravitationally referenced output, since a body's inertia is a function of gravity. For example, the encoder described in the last section if attached to a motor rotor/stator provides the angular rotation rate of the rotor relative to the stator. With the motor off and locked one could pick up the total motor assembly, turn and rotate the assembly without the encoder measuring any motion since rotor/stator are fixed so there is no relative rate between rotor and stator. If an inertial rate sensor or gyroscope was attached to the rotor it will also measure the relative angular rate however when the motor is off, locked and the assembly picked up, turned and rotated the inertial rate sensor would measure that motion of the complete assembly. Any pointing system mounted on a moving platform that must remain pointed at a fixed inertial reference requires inertial sensors in the stabilization control system. The inertial sensor measures angular rate disturbances relative to an inertial referenced line of sight (LOS), using the measured rate as a negative feedback signal to the stabilization controller causing the disturbances to be attenuated. There are many types and configurations of inertial sensors. The performance grade of gyros can be categorized as rate, tactical and inertial. Rate used to refer to lower end performance applications, ranging from gaming to automotive or even relatively coarse

LOS stabilization, however there are now many rate sensors that are considered tactical grade with performance commensurate with low end fiber optic gyros (FOG) from manufacturers that include Gladiator Technologies, EMCORE Systron Donner, Analog Devices, and VectorNav. Tactical LOS stabilization applications normally include a track sensor providing periodic LOS updates; low noise is important but some drift is acceptable because it is compensated by the track sensor updates. As precision requirements increase, drift over a sample period becomes more of an issue and has to be evaluated. Inertial quality refers to applications where pointing is performed in absolute coordinates relative to true or magnetic north via an INS. Minimizing drift as well as angular random walk, and their dependence on temperature and acceleration are important. Table 5.0 provides rough estimates of typical gyro parameters associated with each sensor grade.

Table 5.0 Typical Performance Levels for Different Gyro Grades

Parameter	Units	Performance Grade				
		Rate		Tactical		Inertial
		min	max	min	max	(less than)
Angle Random Walk	Deg/sqrt hr	0.5		0.02	0.5	0.02
	dps/sqrt Hz	0.0083		0.0003	0.0083	0.0003
Bias Drift	deg/hr	10	1000	0.1	10	0.1
Scale Factor Accuracy	%	0.1	1	0.01	0.1	0.01
Full Range Scale	deg/sec	50	1000	50	1000	500

An inertial measurement unit (IMU) has three gyro or rate sensors mounted on three orthogonal axes measuring the inertial rate about each respective axis and three accelerometers aligned along the three orthogonal axes that measure acceleration along each axis. Many gyros have the option of an analog or digital output. Analog outputs can be more susceptible to noise than a digital output. The grounding/shielding interface design of a gyro with an analog output for a precision application is very important. Digital interfaces are not completely devoid of noise and come with inherent latency. A high bandwidth stabilization loop will require a high bandwidth gyro and high-baud data rates. There are many silicon MEMS (micro-electromechanical system) technologies available, including IMU as well as INS/GPS versions which are discussed in the second part of the course.

4.5.1 Inertial Rate Sensors: Manufacturers of high quality inertial rate sensors or gyro's include Gladiator Technologies (https://www.Gladiatortechnologies.com), EMCORE Systron Donner (https://www.emcore.com), Analog Devices (https://www.analog.com), and VectorNac (https://www.vectornav.com). These are MEMS sensors that utilize a vibrating structure to obtain linear translational motion and the Coriolis effect when the device is rotated to create a force orthogonal to the direction of linear motion that is proportional to an angular rate. Silicon is used to fabricate the vibrating structure which can be a beam, disc, wheel, cylindrical resonator, hemispherical resonator, or wine glass resonator. Capacitive or piezoelectric elements are often used to measure the Coriolis force produced. The IEEE has defined these as a Coriolis Vibratory Gyroscope. Some early designs as manufactured by Systron Donner used a quartz tuning fork to obtain the linear vibration and were called Quartz Rate Sensors (QRS); still being available today. These MEMS inertial sensors couple electrical and mechanical interactions with microchip fabrication methods that are described further in Part 2 of the course. Performance of the MEMS inertial sensors has improved substantially in the last 10 years to a point where they can be used for tactical applications providing the performance of a low-end fiber optic gyro at substantially less cost.

4.5.2 Magnetic Hydrodynamic (MHD): A manufacturer of this type of rate sensor is ATA Corporation (https://www.atacorp.com). This approach considers the behavior of a conducting fluid in the presence of a magnetic field. A permanent magnet is separated from a coiled electronic current loop in a conducting fluid. A static magnetic field is established between the coil and permanent magnet. When angular motion occurs about the sensitive axis of this type of sensor, there is a relative velocity difference between the fluid proof mass, which is very conductive, and the static magnetic field that moves with the case. This relative velocity difference between the conductive fluid and magnetic field generates an electric potential across the channel proportional to the velocity that can be sensed and measured. They are very accurate but do not have a DC response and are expensive.

4.5.3 Spinning Mass Gyros: For many years, the spinning mass gyro was the most prevalent gyro used for precision tactical applications and is still in use for precision pointing applications. One manufacturer of this type of gyro is Northrop Grumman (https://www.northropgrumman.com). It is normally configured as either a rate sensor or a rate integrating gyro (RIG). This type of gyro provides a good tradeoff between accuracy, size, and price. Several manufacturers provide the gyros with signal conditioning electronics that simplify integration. Central to the gyro design is a gimbaled mass or rotor spinning at a constant velocity. The gyro response is governed by Newton's Laws and conservation of angular momentum. The angular momentum vector of a spinning mass changes direction but not magnitude when a torque is applied about an axis perpendicular to the spin axis. Torques applied about one axis orthogonal to the spin axis result in a precession motion of the gimbaled mass about the other orthogonal axis due to a change in the direction of the angular momentum vector. Most gyros of this type sense motion about two orthogonal axes. The spinning mass is rotated by a motor at a high constant angular rate. High-speed operation is critical to gyro performance since the greater the angular momentum, the more resistance there is to perturbing torques. The basic configuration for this sensor type is as a RIG. If the precession of the rotor is spring restrained it can be used as a rate sensor. The RIG can also be configured as a rate sensor by closing a capture loop about it that acts as an electronic spring. Accuracy and drift rates adequate for many tactical applications can be achieved with this configuration. The dynamically tuned gyro (DTG) is a higher precision version of the spinning gyro. It uses flexure joints to suspend the spinning mass, and the gyroscopic effect to create balance torques proportional to the square of the spin speed. This effectively isolates the rotor at a particular speed, called the tuning speed, from internal friction disturbance torques. Normally significant conditioning electronics are required for operation including rotating the mass as well as capture and/or rebalance loops which ensures the gimbaled mass does not continuously impact the stops.

4.5.4 Ring Laser Gyro (RLG): A manufacturer of this type of gyro is Honeywell (https://www.honeywell.com). The RLG is a high precision gyro often used in navigation applications. It is normally large in size and costly. The RLG operation is based upon the Sagnac effect, an interferometric technique used for rotation rate detection. The Sagnac effect states the optical path lengths of a circuit will differ when it is rotated. RLGs operate with mirrors mounted in a laser cavity arranged in a triangular geometry, as shown in Figure 19.0, to set up a closed loop or ring for the laser path. The distance between mirrors or a triangle side length is about 10-15 cm and are tuned to create a perfect standing wave pattern. A precision laser source is divided into two beams propagating along the path in opposite directions. When the ring rotates, the path distance traveled by each beam increases for one direction and decreases for the other relative to a starting point. This is a relativistic property of light propagation in a rotating inertial frame. The path length difference corresponds to a phase difference, proportional to rotation rate, that is detected by the Sagnac interferometer. The RLG mirrors are dielectric mirrors, tuned to the laser wavelength, to form the closed loop path for the laser beams to propagate in both directions.

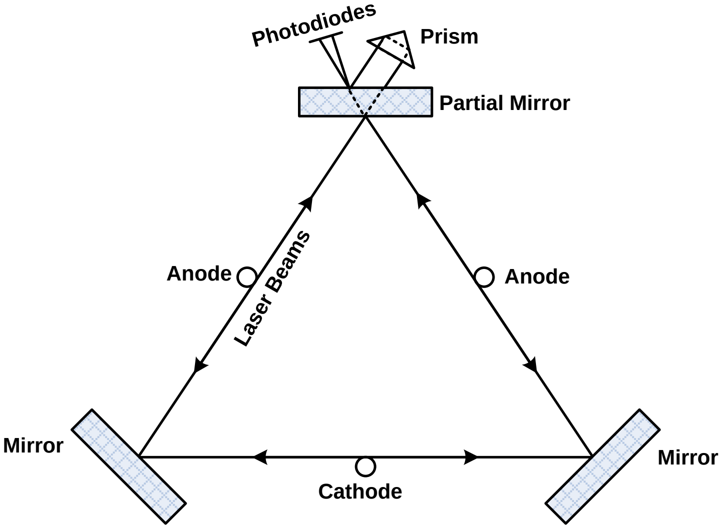


Figure 19.0 Ring Laser Gyro

A prism recombines the two beams creating an interference pattern at the photodiodes that is stationary if the gyro does not move. When the gyro rotates, the interference pattern moves across the photodiodes. The pattern minima or maxima are counted and then integrated to measure heading change. Resolution is roughly the path length divided by the laser operating wavelength (i.e. ~0.63 micrometers) and on the order of about two arc-seconds.

4.5.5 Fiber Optic Gyro (FOG): Manufacturers of this type of gyro include Northrop Grumman, Honeywell, and KVH (https://www.KVH.com). The FOG has high accuracy and is relatively insensitive to shock and linear vibration, but can also be large and expensive. The use of fiber optic gyros has increased significantly as fiber optic technology has improved, both in the cost and accuracy. FOG technology evolved from RLG systems, employing a similar principle, but without the need for a costly laser and laser cavity construction. FOG size and accuracy are inter-related since accuracy improves with fiber cable length but so does size. In a FOG, the optical path is a fiber optic cable coiled many times around a cylinder. The optical path length, L, with N loops is L=2Nπ. The beam is emitted from a laser diode or super luminescent diode (SLD), polarized and then divided by a 50/50 beam-splitter. The light going to one side of the coil passes immediately through a phase modulator. The light moving in the opposite direction does not pass through the phase modulator until almost completing the path and is about to be recombined by the beam-splitter. The recombined light is transmitted through the polarizer again and onto the photo-detector. The Sagnac phase shift is the difference in optical path lengths in the rotating circuit; as given by:

$$\Delta L \cong \frac{4\pi \cdot r^2 \cdot \Omega}{c}$$

where

r-effective circuit radius
Ω -rate of angular rotation
c-speed of light in the medium

A photo-detector measures the incident light intensity variations related to the Sagnac phase shift due to the angular rotation rate. Because the maximum sensitivity occurs when the phase difference is a multiple of 90°, to maximize sensitivity for small rotation rates a non-reciprocal stable 90° phase shift must be introduced between the counter-rotating beams often using an oscillator to sinusoidally modulate the phase difference. This creates an AC current versus rotation rate response sensitive to small rotation rates that is linear over a limited response region. A phase discriminator is used to detect the phase angle. This type of gyro is referred to as an open-loop gyro, and has limited range and high scale factor nonlinearity. A closed loop design provides feedback to the oscillator from the discriminator to drive the observed Sagnac effect to zero, a process similar to voltage-controlled oscillator control in a phase locked loop. Closed loop operation allows extremely high dynamics and good linearity. Resolutions on the order of 0.004 arc-seconds are possible using a FOG.

4.5.6 Inertial Measurement Unit (IMU): This is an integrated package that normally contains 3 gyros that measure angular rates about 3 orthogonal axes and 3 accelerometers that measure linear acceleration along these axes. They feature relatively small size, precise alignment between sensor axes, and the high bandwidths/data rates. Similar to gyro's, they come in different performance levels.

4.5.7 Inertial Navigation Systems (INS): This device is used when precision inertial pointing or location is required. The heart of an INS is a precision IMU. The INS integrate the rate and acceleration measurements to obtain the angular orientation and position of the platform relative to either true or magnetic north (depending upon the device). Usually Kalman Filter based navigation algorithms, discussed in Part 2 as an example of sensor fusion, are used to provide the state estimates of orientation and position. Unaided position estimates can be in the 10's of meters, but when integrated with a GPS, very accurate (<< 1 meter) position information can be obtained.

4.5.8 Gyro and INS Specifications: Gyro specifications vary depending up the sensor type. This final section provides a brief overview of typical gyro specifications.

Gyro Drift/Bias: The gyro bias or drift rate indicates the steady increase in angular error with time (in deg/h or deg/s). A constant drift bias can be compensated to within the stability of the drift. Drift is dependent upon both temperature and acceleration.

Gyro Scale Factor Error: This is a measure of the instantaneous angular measurement error that occurs during rotation due to scale factor variation with environmental conditions and inherent non-linearities.

Gyro Excitation Noise: Gyro output noise can include significant components at the gyro spin with a spinning mass gyro, or the excitation tuning fork frequency with a Quartz Rate Sensor.

Misalignment: between the gyro axes causes cross-coupling between the rate measurement axes with rate error proportional to the alignment error. With inertial pointing applications, the gyro or IMU must be aligned with the platform INS using a procedure referred to as transfer alignment.

Gyro Random Walk: This quantity, given in °/√(hr.), is a measure of the inherent gyro noise. Within a FOG, it is caused by shot noise, relative intensity noise, and detector noise. It is the integrated value of the gyro's white noise whose variance increases with time. Higher values indicate more noise associated with the angular rate measurements. Some companies specify it in °/hr./√(Hz), dividing by 60 this is equivalent to °/√(hr.).

INS Related Errors: INS performance is normally specified in terms of aided and unaided. Aided positioning means another sensor system, usually a GPS, is providing periodic updates to the INS in terms of vehicle position or velocity. Even with a GPS, however, INS performance could be unaided under conditions of a GPS outage.

5.0 Summary Part 1

Part 1 of this coarse examined sensor types and characteristics, developing a model of the sensing system. Performance characteristics were defined and output processing techniques to improve performance described. Finally, descriptions of several types of sensor were provided including temperature, optical, proximity, position, and inertial sensors. MEMS sensors, critical in today's environment are discussed in part 2 with networking and sensor fusion. A brief review of key points from each section follow.

Section 1,2 Sensor Overview

- Sensors are a system: sensing element, processing and conditioning, output display and or communication link, power supply
- Sensors normally provide some measure of an electrical, mechanical, thermal, chemical, light intensity, biological or magnetic energy quantity
- Typical measurements include temperature, pressure, force, flow, or level
- Sensors can operate in a passive or active mode
- Silicon is a material that has numerous micro-electro-mechanical systems (MEMS) applications

Section 3 Model and Performance

- Performance defined by accuracy, is a function of sensitivity, resolution, repeatability, dynamic range, response time, signal to noise ratio (SNR)
- Many noise sources impact measurement accuracy performance
- Sensor performance can be improved with averaging, amplification and filtering, or more sophisticated filtering algorithms such as the Kalman Filter

Section 4 Sensors

- Thermostats are a bimetallic switch using the difference in the thermal expansion coefficients of two metals to create a bending motion to make or break a switch contact
- A Resistive Temperature Device (RTD) measures the change in resistance of a conducting metal and require an excitation current to output a voltage measurement proportional to temperature
- A Thermistor measures the change in resistance of a semiconductor material due to temperature and requires an excitation current to output a voltage measurement proportional to the temperature change.
- A Thermocouple uses Seebeck Effect; a temperature difference across two dissimilar conductive metals, in isothermal contact at one end, generates a voltage proportional to the temperature difference

- IC temperature sensors, the potential across a semiconductor junction varies with temperature inducing a change in current proportional to temperature
- Photoconductive, photovoltaic, and photo resistive are the most common types of optical detectors
- A camera array is either a charge coupled device (CCD) or complimentary metal-oxide semiconductor (CMOS) based
- Inductive proximity sensors use the principles of Faraday's law for detection
- Inductive sensors work with only metallic objects
- Capacitive proximity sensors are used with metallic and non-metallic objects sensing a change in capacitance for detection.
- Resolver, encoders, and potentiometers, LVDT, and Hall Effect are all position sensors
- Encoders use a rotating optical disk to generate a pulse train whose count is proportional to position
- Encoders can be relative to a device reference or absolute
- Resolvers use windings similar to motors with the phase between the excitation winding and the rotated winding providing the angular position, they are absolute devices
- Potentiometers low cost and not very accurate, they are effectively a variable resistor created by a voltage divider with a variable center junction pickoff.
- A gyro measures inertial rate and many different techniques are used; spinning mass, Coriolis rate sensors, Magnetic-Hydrodynamic, fiber optic gyro, and ring laser gyro.
- Ring Laser Gyros use the Sagnac Effect to measure inertial angle and rate
- Fiber Optic Gyros also use the Sagnac Effect
- An inertial measurement unit (IMU) includes 3 gyros plus 3 accelerometers measuring inertial rates about three axes and acceleration along 3 axes.
- An inertial navigation system is a precision IMU integrated with a Kalman filter-based navigation model and GPS.

References

(1) https://en.wikipedia.org/wiki/Sensor

(2) https://www.electronics-tutorials.ws/io/io_1.html

(3) https://www.ni.com/en-us/innovations/white-papers/13/sensor-terminology.html

(4) Bernstein, Dennis S., "Sensor Performance Specifications" IEEE Control Systems Magazine 0272-1708/01, August 2001

(5) Andersen, Brian D. O., Moore, John B., Optimal Filtering Prentice-Hall Inc., Englewood Cliffs, NJ 1979 ISBN 0-13-638122-7

(6) Sheingold, Daniel H. (editor), Transducer Interfacing Handbook A Guide to Analog Signal Conditioning, Analog Devices Inc, Norwood Mass. 1972

(7) https://www.electronics-tutorials.ws/io/io_3.html
(8) https://pdfserv.maximintegrated.com/en/an/Temp-Sensor-Tutorial.pdf
(9) https://www.digikey.com/en/blog/types-of-temperature-sensors
(10) https://www.electronics-tutorials.ws/io/io_4.html
(11) Kennedy, P. “Sightline Control Basics for Geo-Pointing and Locating Part 3” SunCam Course 360, 2019, www.SunCam.com
(12) https://www.electronics-tutorials.ws/io/io_2.html
(13) Kinney, Thomas, “Proximity Sensors Compared: Inductive, Capacitive, Photoelectric, and Ultrasonic”, Machine Design-3331, 9/2001
(14) https://en.wikipedia.org/wiki/Hall-effect_sensor
(15) https://www.electronics-tutorials.ws/electromagnetism/hall-effect.html
(16) Eitel, Elizabeth, “Basics of Rotary Encoders: Overview and New Technologies Machine Design-8029, 2014
(17) Renishaw White Paper, “Accuracy of Rotary Encoders” Renishaw plc 2019
(18) Kennedy, P., Kennedy, R., Stabilizing the Line of Sight, ISBN: 978-0-9988539-7-0 0; 978-1-7320721-0-7, Laurin Publishing/Photonics Media Press, April 6 2018