Crystal Clear Electronics

The development of the Crystal Clear Electronics curriculum was supported by the European Commission in the framework of the Erasmus + programme in connection with the “Developing an innovative electronics curriculum for school education” project under “2018-1-HU01-KA201-047718” project number.

The project was implemented by an international partnership of the following 5 institutions:

- Xtalin Engineering Ltd. – Budapest
- ELTE Bolyai János Practice Primary and Secondary Grammar School – Szombathely
- Bolyai Farkas High School – Târgu Mureș
- Selye János High School – Komárno
- Pro Ratio Foundation working in cooperation with Madách Imre High School – Šamorín

Contact http://crystalclearelectronics.eu/en/
info@kristalytisztaelektronika.hu

23 – Additional Useful Sensors

Written by Tamás Pepó

English Translation by Xtalin Engineering Ltd.

Revised by Gergely Lágler, Ádám Szabó

By the end of this chapter you will have learned about the most commonly used electrical sensors and will be able to solve more comprehensive, complex problems with the help of these devices.

Introduction

In the previous chapters of the curriculum you have learned about different sensors which are useful for gathering data about the outside world. We are using sensors to translate a specific part of reality into our systems, such as the temperature, the presence of a magnetic field, the intensity of light, or the pressure of a liquid. Generally speaking, sensors are signal transducers which convert some kind of physical quantity (light intensity, temperature, etc.) into an electric signal. The converted signal then gets passed to a system which processes it the way we want.

Are human sensory systems sensors? While the answer is not a simple yes or no, they meet the definition since they create an electric signal inside our brain according to an external stimulus. Our eyes provide light sensing and converts the image it sees to a series of electrical impulses the brain can understand. You can see from this the sensors and human sensory organs are very similar. By taking the analogy further we can see that sensors are as important to our electric systems as sensory organs are to humans if we want to have any information about the outside world.

General Structure of Sensors

What does a sensor look like? Let's think together!

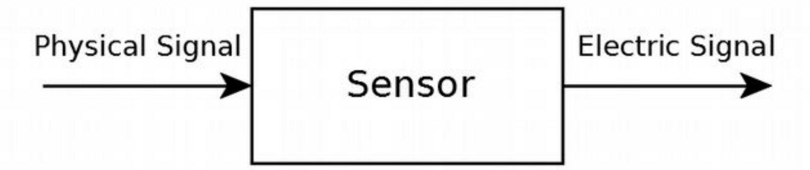

Figure 1 – General model of sensors

On the input side we need some mechanism that can measure a non-electric signal, and on the output side we need an electric signal that our embedded system will be able to process.

In practice sensors usually turn the input signal into another intermediate non-electric signal which is easier to convert to an electric signal and the conversion is more robust. The figure below shows the operating principle of different basic sensors.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

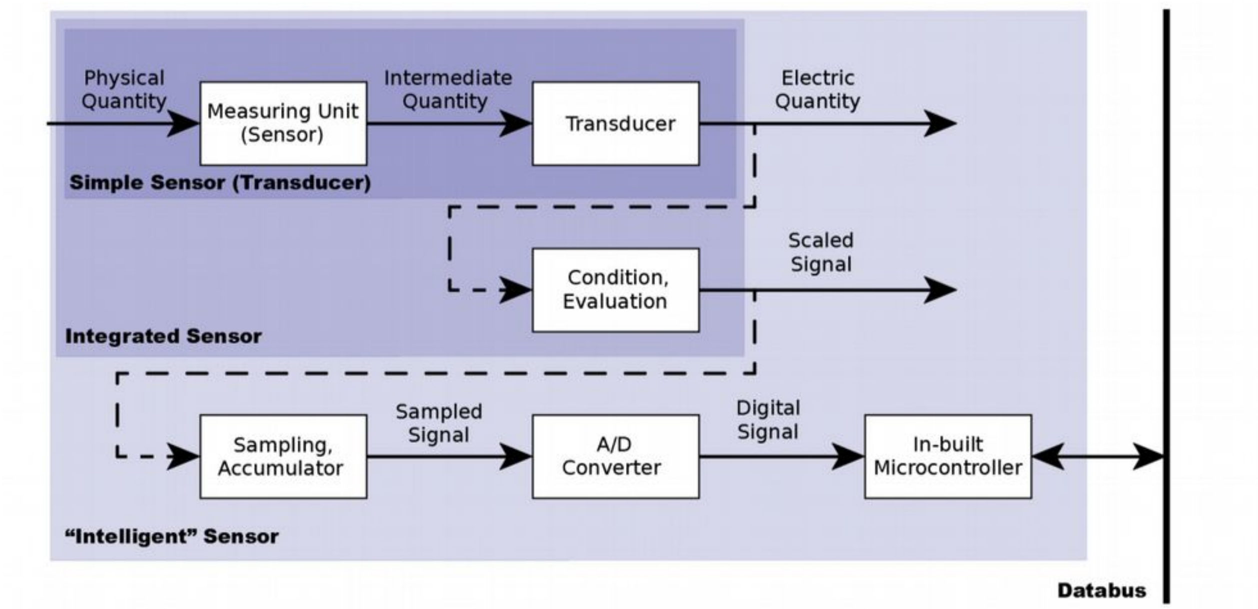

Figure 2 – General structure of sensors

A *simple sensor* only has the two components we have already talked about: the sensor unit, which measures the physical quantity and converts it into some intermediate quantity, and the transducer which converts the intermediate quantity further into an electric signal. This signal consists of the input signal, and an ever-present noise.

For easier understanding let's see an example: If we want to measure the rotation of an axle with a potentiometer, then the contacts inside the potentiometer rotate on a resistor-strip. The position is an intermediate quantity, and the resistance that depends on this position is the output electrical quantity. In this case the sensor unit is the axle of the potentiometer and the contact attached to it, the transducer is the resistor-strip and the moving contact.

We can not only use simple sensors by themselves, we can use *integrated sensors* as well. These sensors have more simple sensors inside and some circuitry that collects and evaluates the measurements of all the simple sensors. A very good example for an integrated sensor is an ultrasound sensor, which can only detect sounds coming from a single direction by itself. To cover a bigger section of space, we need to use multiple simple sensors, which we can purchase built into a single integrated sensor. Another advantage of integrated sensors is that they usually do signal level matching as well, which makes the output voltage range of the sensor match the usual analog voltage ranges frequently used and can be easily processed with microcontrollers.

We call sensors that don't simply output the measured signal as analog voltage, but also process it *intelligent sensors*. The signal processing can be in the simplest case analog-digital conversion, which creates a digital signal from the analog voltage. Usually these sensors contain a built-in microcontroller which evaluates the measured signal and processes it according to some settings. We can communicate with these intelligent sensors via the usual communication protocols commonly found in microcontrollers such as UART or SPI, and we can alter the signal processing behaviour of the sensor with various settings.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

COMMUNICATION

If the output of a sensor is not a simple binary signal, but a series of bits then that has to be forwarded properly on a communication channel. We have two options: serial and parallel communication. The number representing the measured quantity that has to be outputted can have many bits. Think about that a measured range divided into 10 000 equal parts (which is not a really high resolution) requires 14 bits to represent, and it is not unusual to see sensors with a 64-bit output such as encoders.

With *parallel communication* we need one signal wire for each bit. It is easy to see that this is really fast, as all digits reach the microcontroller at the same time, so the data can be read in one step. The biggest disadvantage is the number of wires required, in which not only the data bits have to be considered but all the wires required for synchronizing the transmitter and the receiver. Using it is only feasible over very short distances, usually inside a device, for example for the memory buses in a computer. Another important disadvantage is that the wires use up a lot of pins for both the transmitter and the receiver microcontroller, which is often not permittable as we would have to use a higher pin count device, and the size of the circuit would also increase considerably. On top of that connecting such a high number of wires when designing the probability of making a mistake is higher. If the speed requirements don't justify it, we usually don't use parallel communication.

In *serial communication* only one wire is used to transfer data in one direction, on which the bits are sent one at a time, after each other. The most important task of the receiver and the transmitter is to stay in sync.

Synchronization can be achieved in multiple ways, for example by transmitting a signal that represents the timing of the data (clock). If the clock and the data lines are out of sync, then we may read faster or slower than necessary, which creates errors.

The advantages of serial communication include the low number of wires required, which also means a simple hardware, but the disadvantage is that communication is slower compared to the parallel case, and keeping the two ends in sync is difficult over long distances. Today the slower speed is barely noticeable since the speed of microcontrollers is increasing rapidly, increasing the communication speed as well.

Sensor with digital outputs are almost always use serial communication to integrate into the system. An advantage of this method is that swapping the sensor to a higher resolution one does change the design, as the same number of wires are required for communication.

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Noise

Let’s talk more about measuring signals with noise. We often experience that the measured signal has a weird, abnormal value which is very different than the expected. After examining the measurement setup we might think that the protection against noise is inadequate. Generally, a measured signal consists of the real value of the physical quantity we would like to measure, and some noise which is superimposed on the signal. At any point in time we can only measure the sum of the two signals together.

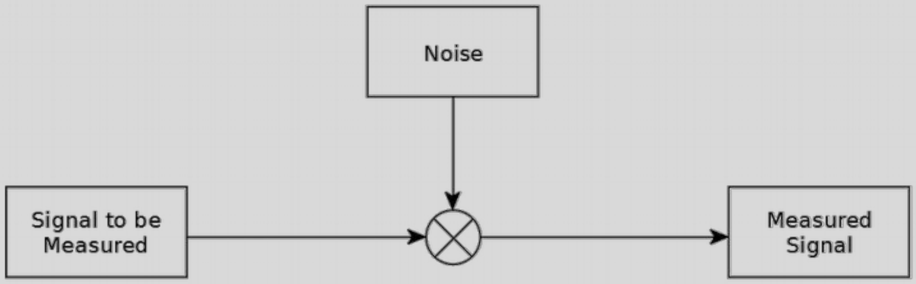

Figure 3 – Noise superimposed on the signal to be measured

We can use a simple trick to eliminate the noise without significant data loss (from an information theory standpoint any deformation of the useful signal reduces the available information, thus leading to data loss). Since the noise superimposed to our useful signal is usually a white noise, meaning that it is a random value with a zero mean, we can use an averaging filter to filter it out. We can only use averaging in a window that is small enough compared to the original signal so that the averaging does not distort the underlying signal. Think about if you would average a sinus wave over a longer period of time you would get zero, which is not very useful and does not tell a whole lot about the signal. The algorithm of an averaging filter is pretty simple: replace every point of data with the average of its immediate surroundings. With a formula:

$$y(n) = \frac{1}{N}\sum_{i=0}^{N-1}(x_{n-i})$$

Signal distortion occurs with the averaging filter too the following way: When averaging, you lose information about fast changes of the signal, which is sometimes very important. Let’s have a look at a practical example, a window opening sensor. When the window is opened the sensor sends a quick pulse on the output signal, but if we have a long averaging window instead of a sudden change we only see a slow transition, since the surroundings of the pulse all contain the previous state. This slows down the detection mechanism, which may or may not be critical depending on the application.

This is why averaging is usually not used on signals that can change suddenly, since the white noise superimposed on these signals is a lot smaller then the actual change we need to detect, so with a simple threshold level is enough to reliably detect the change. The example below shows a signal where we want to detect a short 5 V pulse and the noise has a 1 V amplitude. The signal is 0 V before the pulse, and the 5 V pulse is high enough that we can detect it without averaging. You can see the two thresholds in the

Erasmus+
This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

picture below, the output is 0 if the voltage is below 1.5 V, and 1 if it is above 3.5 V. This way we can reliably detect an input change of a noisy signal without averaging.

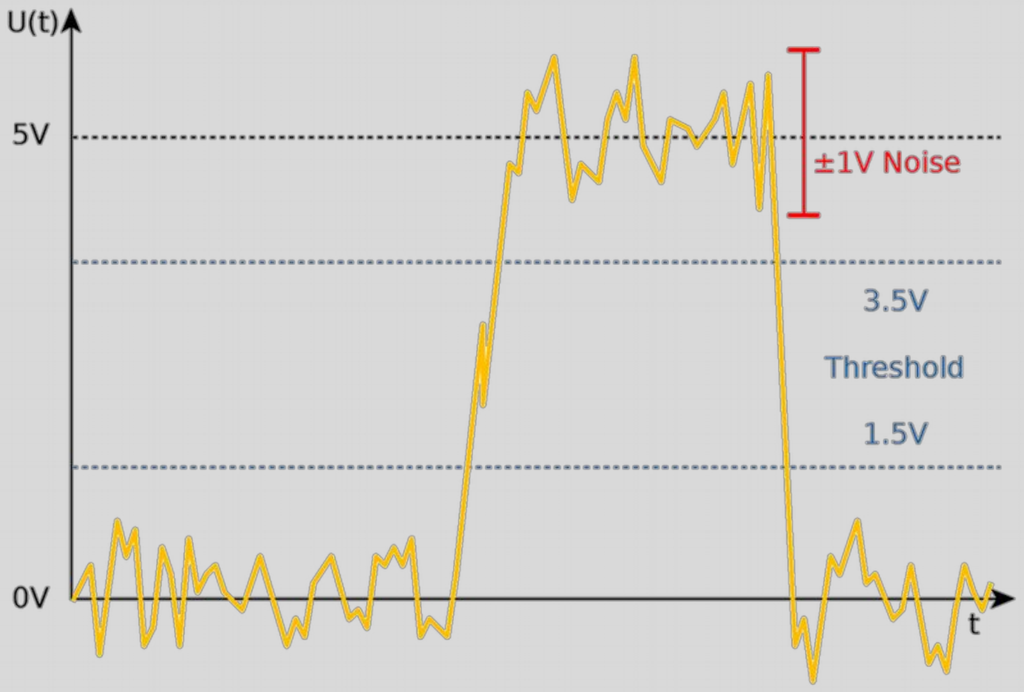

Figure 4 – Detection of short signals with noise

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Classification of Sensors

After the short introduction, let's look at in detail, what type of sensor you will meet later on if you deal with electronics!

Classification by operating principle

Operating principle	Examples
Analogue	force- and pressure sensors, flowmeters, temperature meters, odometer, length sensor, rotation sensor, etc.
Digital	position sensors, proximity sensors, pressure switches, switching thermometers, etc.

Classification by physical quantity

Quantity to be measured	Examples
Mechanical	position, displacement, acceleration, force, speed, rate of flow, sound waves, pressure, etc.
Thermodynamic	temperature, heat, heat flux, etc.
Electric and magnetic	electric field, magnetic field, charge, voltage, current, resistance, etc.
Radiation	electromagnetic radiation (radio wave, microwave, x-ray) and particle radiation (alpha radiation, beta radiation, ionizing radiation, neutrons)
Chemical	flow, concentration, activity of neutral and charged particles in a certain medium
Biological	specific features of living organisms

Classification by energy requirement

Energy source	Explanation
Active (Generator type)	They do not require separate energy source; they get the energy for the operation from the observed medium
Passive (Modulator type)	They require external energy source for their operation

Basic Terms

In this section I will talk about the basic terminology commonly used when talking about sensors so you can better understand the upcoming sections about specific sensor types, and so later when reading about a sensor you can understand its operating principle better.

The main thing that describe a sensor is the *static characteristic curve*. This is the relation between the input signal and the output signal of the sensor. If we want to make any type of measurement, we cannot ignore this function. Maybe the simplest way of illustrating the concept is by the example of a light sensor. Imagine a cartesian coordinate system, where the input values are on the x axis, and the output values are on the y axis. The units for the two axes are not the same, in the case of the light sensor the input axis is in Lumens, while the output axis is in Volts. If you map each input to each output in the sensors range onto this coordinate system, you get a function $y = f(x)$ which is the static characteristic curve.

The static characteristic curve is linear, if $f(x)$ is a line, which means that the input and output values are directly proportional.

The *offset* (or *offset error*) is the value the sensors outputs when the input is zero, or $f(0)$ on the static characteristic curve. *Sensitivity* (E) is the steepness of the static characteristic. If the characteristic is linear, the sensitivity is a constant, since the curve has the same steepness at every point. If the characteristic is

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

nonlinear, the sensitivity itself is a function of the input. In the figure below you can see a static characteristic of an imaginary sensor, which illustrates the concepts above. The yellow line shows the ideal static characteristic of the sensor, and you can also see the sensitivity.

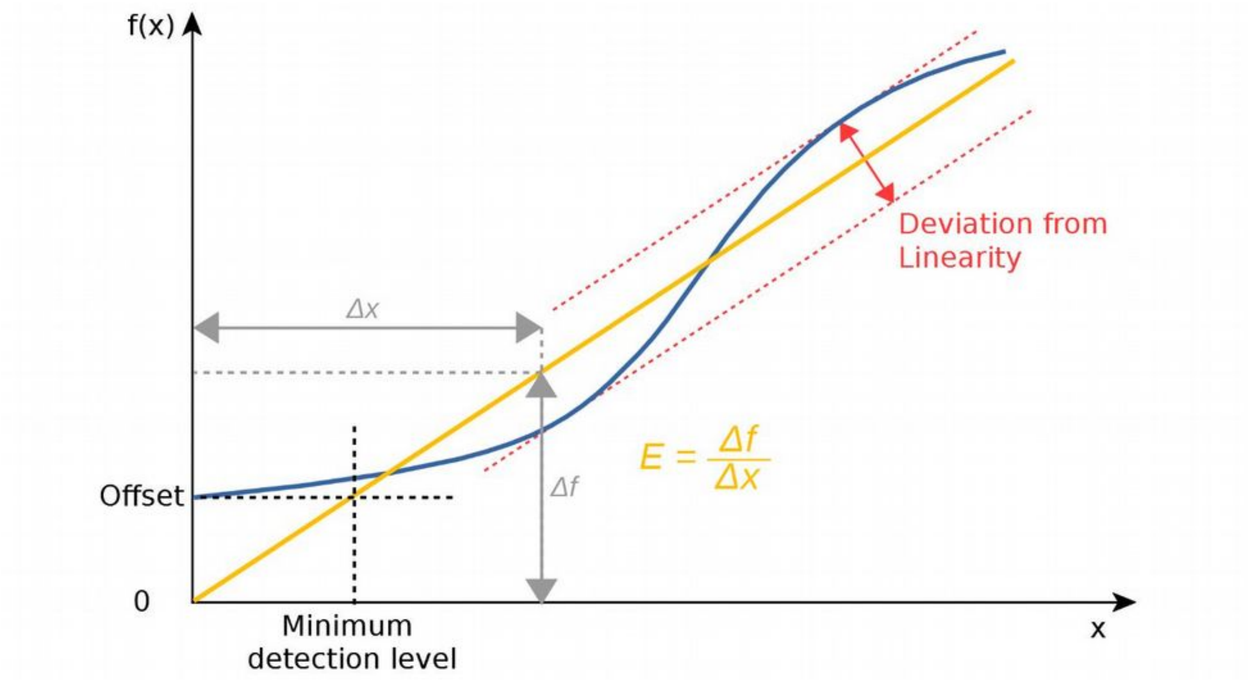

Figure 5 – Static characteristic

The *detection threshold* seen in the figure above is the minimum input value which has a valid output. The sensor cannot distinguish lower values (the curve flattens, almost becomes horizontal).

In the next section we will summarize the parameters of an ideal sensor, which of course does not exist in the real world, but we can see where sensor technology is going, what would be ideal from an electronics standpoint.

The *static characteristic* of an ideal sensor is linear throughout the whole range, the *sensitivity* does not depend on the input. There is no *offset error*. The *response time* is zero, which means all input changes get reflected on the output immediately, without delay, so at any point in time the output value represents the input value according to the static characteristic. If the response time is zero, then the *bandwidth* of the sensor is infinite, which means its working properly at any frequency, the output can keep up with the input changes, no matter how fast they may be. Its *detection threshold* is infinitely small, so any small input is detected, and the *upper limit* is a calibrated maximum value.

It's important to remember that in a real sensor some of these properties are very hard, or impossible to implement, so in the real world we always have compromise if we want an accurate, easy-to-use, affordable sensor. Despite this with the required accuracy and input range in practical measurements these compromises usually do not pose a problem. A good example is that we usually don't want to measure arbitrarily large weights (such as a car, or a person) with microgram precision, 1 kg or ¼ kg precision can be more than enough.

Erasmus+
This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

SPECIAL SENSORS

PROXIMITY SWITCHES

In the curriculum we talk about the following proximity switches:

1) Magnetic
2) Inductive
3) Optical
4) Capacitive
5) Ultrasonic

Generally, a job of a proximity switch is to detect the presence of a close object by utilizing some property of the object for detection. This property of objects is the way we can categorize the different types of proximity detectors available. In beginner electronics projects the most usual types are magnetic and optical sensors because of their simplicity. We will talk about these two types in more detail.

Magnetic Proximity Switch

Magnetic proximity switches (also known as REED-relays) can detect the presence or the lack of a magnetic field, may it be the magnetic field of a permanent magnet or an electromagnet. They are commonly used as limit switches for example in cylinders made out of plastic or aluminium which does not affect the detection. In these cases, usually the piston or the seals are magnetic which is detected by a magnetic proximity switch outside the cylinder. This arrangement has the advantage of the sensor being separated from the insides of the cylinder, and this way it does not have to be as durable as a sensor inside would need to be, and it can be easily replaced without replacing the cylinder. They also can be used for detecting window, door, or lids of electronic devices openings

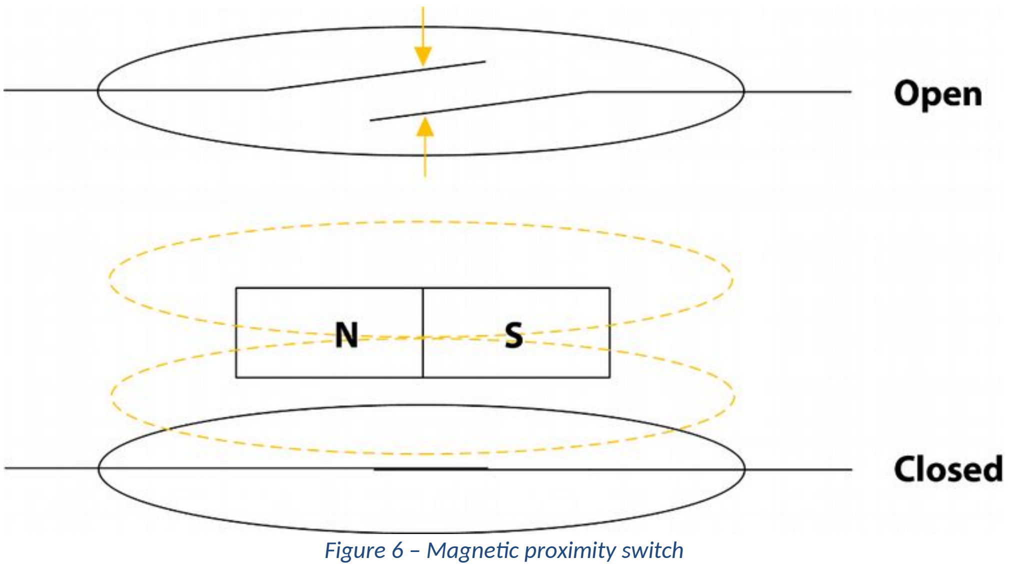

Figure 6 – Magnetic proximity switch

The structure of a REED-relay can be seen in the picture above. When a magnetic field is close, the two contacts that are only a few hundred micrometers apart, touch each other, and the circuit is closed. These two are called open and closed state. We can buy switches that work reversed too, which are called normally closed type. The two contacts are usually in a glass case filled with inert gas.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Inductive Proximity Switch

Inductive proximity switcher are active devices, they generate their own magnetic field, and detect ferromagnetic materials that enter this field. Usually they are used in special circumstances, such as where waterproof devices are needed, or the environment is heavily contaminated.

Its operation is much more complicated compared to the magnetic proximity switch mentioned above. Without going into detail, it is worth to know that it generates an oscillating magnetic field. The ferromagnetic materials that enter this field slow down this oscillation and induce a voltage in the circuit generating the magnetic field. Remember, some sensors convert the physical quantity to an intermediate quantity first, in this case this is the frequency of the oscillation. The oscillator circuit is located in a magnetically sealed box, only the coils are exposed. The opening in the magnetically sealed case ensures that we can detect objects in that specific direction.

How can a varying frequency signal become a binary output (object present or not present)? We use a threshold level for this, and the binary signal is the frequency compared to the threshold level. We don't want to go into any more detail here, because it is rare for a hobbyist to use this sensor type. If you come across them during your studies later, you will learn a lot more about their interesting properties depending on your application field.

Optical Proximity Switch

Optical proximity switches are commonly used in packaging machines, or to detect people or objects (such as cars) passing through a gate. These are more commonly called photocell gates. Their operating principle is quite simple: pulses or infrared light coming from the light emitters travel to an object or a reflector, and back to the light sensor. They can work at much longer distances compared to magnetic or inductive sensors, but the light emitting, and sensing parts must be kept clean, so in a dirty environment their usage is not recommended.

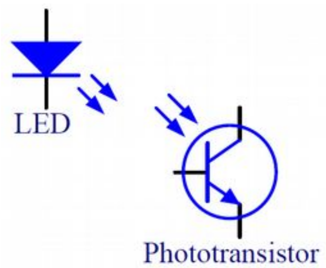

Figure 7 – Phototransistor as light detector, and photodiode as light source

The light emitting component is a photodiode (LED), the light detector subassembly is the phototransistor, which opens when light is shining on the base.

Capacitive Proximity Switch

Capacitive proximity switches have an active surface which measures capacitive coupling. If we move an object close to this surface, the capacitive coupling between the object and the surface changes, similar to how the capacitance of a capacitor changes when the distance between the two plates changes. When using them in industrial applications it is often an advantage that they work without direct contact, for example in power tools. Besides industrial applications they are often found in aquariums, terrariums, or close to plants.

Erasmus+
This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Ultrasonic Proximity Switch

These are very similar to photodetectors as both of them work by wave propagation. These sensors have a transmitter and a receiver too, but instead of light they use ultrasound, which is reflected from the object to be detected. Ultrasound has a higher frequency than sound we can hear (that is from 20 Hz to 20 000 Hz), usually between 30 kHz and 300 kHz.

Soundwave propagation is greatly influenced by the temperature and humidity of air (or other medium). This effect does not cause any problems if the error is negligible with the precision requirements of the measurement, for example if we only want to detect whether an object is present or not. Sound waves are hard to focus in a certain direction, so the smallest detectable object size depends on how much the waves spread.

Hall-Effect Sensor

A Hall-effect sensor is a sensor based on a physical magnetic phenomenon called the Hall-effect. We can utilize it in many ways, such as for measuring current or magnetic field strength.

Hall-Effect

The Hall-effect has been discovered by Edwin Hall in 1879. If we consider a rectangular slab of semiconductor material in which current is flowing between two opposing sides, and a uniform magnetic field is present between two other opposing sides, perpendicular to the current flow, then a voltage is induced between the remaining two sides perpendicular to both the magnetic and the electric field (current).

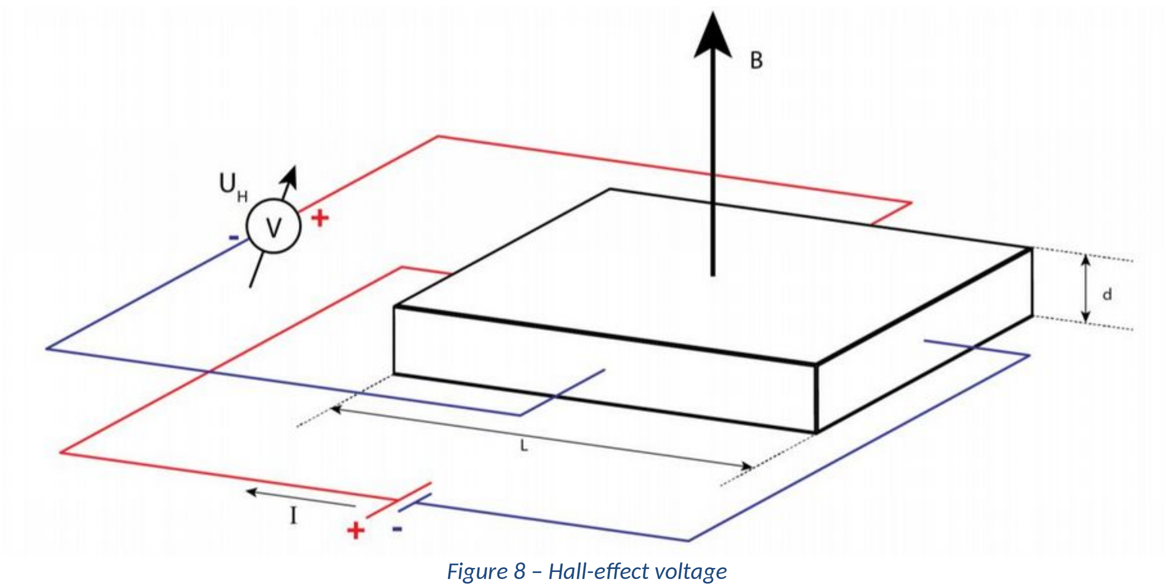

Figure 8 – Hall-effect voltage

In the picture above the magnetic induction (B) is pointing upwards, and the current (I) is flowing from left to right in the region. Between the front and back sides of the slab the Hall-voltage (U_H) can be measured. The voltage is induced because of the Lorentz-force acting on the moving charges in a magnetic field causes the distribution of charges to change, creating a potential difference.

Ignoring the lengthy proof, we state that the Hall-voltage U_H can be calculated using the following formula

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

$$U_H = R_H \bullet B \frac{I}{d}$$

where R_H is a constant called the *Hall-constant*, B is the magnetic induction, I is the current, and d is the thickness of the semiconductor slab as seen in the figure above. As you can see from the equation, if the magnetic induction is known, we can calculate the current by measuring the Hall-voltage; or we can calculate the strength of an unknown magnetic field by using a known current and measuring the Hall-voltage.

Application

Hall-sensors can be used in many applications, because it is easy to measure both magnetic induction and current with them. There are four main areas of application:

- **Linear output:**
  - *Current measurement* – We provide a constant magnetic induction, and we measure the Hall-voltage, which is directly proportional to the current flowing.
  - *Magnetic field strength measurement* – The measured Hall-voltage is directly proportional to the magnetic induction, if the current is kept constant.
- **Binary output**
  - *Unipolar* – Using the Hall-sensor as a proximity switch, where a magnet in proximity keeps the output on
  - *Bipolar* – The proximity of one magnetic pole switches the output on, and the proximity of the other magnetic pole switches it off.

To give a better understanding what Hall-sensors can be used for, here are a few common real-world examples:

- As a binary output detector, they are commonly used on rotating axles to measure rotation. If a magnet is attached to the axle, and a Hall-sensor is put beside it, the sensor will detect each rotation when the magnet is passing by. This is the most basic type of application. Of course, more complicated detection can be implemented with more magnets placed symmetrically on a disk attached to the axle, which can provide information on rotation speed as well. Remember this, when we are talking about encoders in a later section of the chapter.
- As a linear output sensor, they are commonly used in accelerator pedals of cars to measure the position of the pedal. When pressing the pedal, a magnetized arc rotates and changes the magnetic induction of a Hall-sensor proportional to the pedal position, which gives a proportional voltage output.
- As an active proximity switch, which switches its output when a magnet is removed from the proximity of a Hall-sensor. The main issue here is that unlike REED-relays, an external circuit is needed to provide the constant current flowing through the sensor.
- Automotive applications:
  - Measuring position of pistons in an engine
  - Ignition switch
  - Measuring position of valves
  - Detecting blocking of wheels in the brake drum
  - Fuel-level sensor (with a magnet attached to a buoy)

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Why do we use Hall-sensors in so many places? Because they are usually very tolerant to the environment, so in cars they can withstand cold and ice and still work reliably. An even more important property is that they don't have any moving parts, and as the mechanical engineers like to say, "What is moving, breaks". You can remember this saying when buying computer parts and deciding between an HDD and an SSD.

Their most important property is that they always produce a valid output, even at high frequencies. These sensors can work up to a few hundred kHz with great accuracy, but they are also good for static measurements.

MEMS Gyroscopes and Accelerometers

Gyroscopes and accelerometers are used where changes in position or orientation is needed to be tracked, without using an external device, or in other words we want to detect these with sensors on the moving object itself. So, we are not using a stationary sensor placed on the ground to detect the movement or rotation (orientation change) of an object.

It's important to emphasize, that even though we want to get information about the position of an object most of the time, we cannot use these sensors to get absolute values, only the relative changes compared to an initial state. To complicate things further, if an object is moving in the same direction with a constant speed, sensors attached to the object without any connection to the outside world cannot detect the speed of motion, or even tell if the object is moving or stationary. This is why when talking about gyroscopes and accelerometers we talk about measuring changes in motion.

Simply gyroscopes are being used in space and marine technology since long ago. These were complicated mechanical devices, with a price tag way too high for any commercial product. This changed with the introduction of MEMS (Micro Electro-Mechanical System) technology, with which we can create small, accurate, and cheap solutions so people can afford to use these sensors even in hobby projects. The drastic decrease in price and increase in accuracy of the devices brought a revolution in robotics as well, where it can be used to approximate the position and orientation of mobile robots.

Figure 9 – KUKA mobile robot
Author: KUKA Laboratories GmbH [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)]
https://commons.wikimedia.org/wiki/File:KUKA_youBot.jpg

Erasmus+
This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

We can mention lots of examples in our everyday lives where we are using MEMS accelerometers and gyroscopes. Most modern smartphones and tablets contain such a sensor, which can tell when to rotate the screen. Sometimes a device uses both solutions (accelerometer and gyroscope) for greater precision, but most of the times they use a simple accelerometer which can measure gravity. This is why the phone screen does not rotate when the phone is rotated while placed flat on a table. Another example is the controllers for gaming consoles which can read the motion performed by the player using MEMS sensors.

I would like to take a little detour to the field of robotics. I'm sure that most of you are familiar with drones, which have become a popular hobby. Basically, a drone is a remote-controlled quadrocopter (helicopter with four rotors), but six-rotor hexacopters and eight-rotor octacopters also exist. Autonomously controlling such a robot is not an easy task. Unlike a self-parking car, where the controller has to control the motion on one plane, in this case it has to do it in every direction of space, and on top of that to remain stable it has to control rotation on every axes too. You will learn more about control in the last chapter, it is enough to know that to keep the drone at a certain position with a certain orientation the controller has to constantly read sensors and adjust the speed of the four rotors separately (this has the effect of increasing or decreasing lift on any sides, which can alter position and orientation too).

In the following you will learn about two MEMS sensors in detail.

Accelerometers

Accelerometers can tell us how much a certain object is accelerating, which is the result of the sum of forces acting upon the object. What happens if the object is not accelerating, but moving with a constant velocity? Well, then the accelerometer measures zero, so we have no way of telling whether the object is stationary or moving at a constant velocity.

We can use an accelerometer to measure absolute position, but only with certain restrictions. We cannot measure absolute position as we have already discussed, only the relative position change from the initial state by using the well-known equations of acceleration, time, and distance. By knowing the time and the acceleration we can calculate the velocity, which can determine the distance travelled when combined with the time again. This method works, but it has a great drawback, that is it need accurate time measurement.

Without going into detailed mathematics, we state that in continuous time addition becomes integration. Multiplication is a kind of addition too, and when we want to calculate velocity, we are multiplying by time. Hidden in the formula $v = v_0 + a \cdot t$ we can see that velocity is an integral of acceleration by time.

We had to mention this so we can see the problems in our method of calculating position. The main problem is that we are measuring relative position (or velocity in our case) so at any given point of time the measured value depends on all the measurements before it.

This can be proven, since our current velocity depends on our initial velocity, and all the previous accelerations since the beginning of the measurement. If we have an error while measuring the initial velocity v_0 (which is always the case thanks to noise) then that affects the current measurement and through it all future measurements too.

Because we have still just calculated velocity, we need another integration to derive the position ($s = s_0 + v \cdot t$). This brings another set of errors into our measurement, which only accumulates over time.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

You can see from that that accelerometers are not used to determine absolute position alone, usually they are combined with other sensors providing absolute position information (such as GPS) to improve precision. Or they are used to calculate the forces acting upon the object at a given time, or interval of time (such as in a car cornering or having an accident).

Accelerometers are usually able to measure acceleration on only one axis, in other words are one-dimensional devices. In practice we usually use three-dimensional accelerometers that have three separate one-dimensional accelerometers integrated. This way we can measure acceleration in any direction of space and derive their sum.

Basically, we know sensors with two different operation methods:

1. ***Capacitive***

We use a complex spring-mass system in order to sense acceleration. You can imagine this like putting two capacitors back-to-back and connecting one of their electrodes together. We hang this common electrode in such a way that its plane is perpendicular to the axis where we want to measure acceleration. When accelerated hung electrode behaves like a mass on a spring, bending in the opposite direction to the force causing the acceleration. Because of the bend, the distances between the inner and the two outer electrodes change, which results in a capacitance change in the two capacitors. You can see the structure in the picture below:

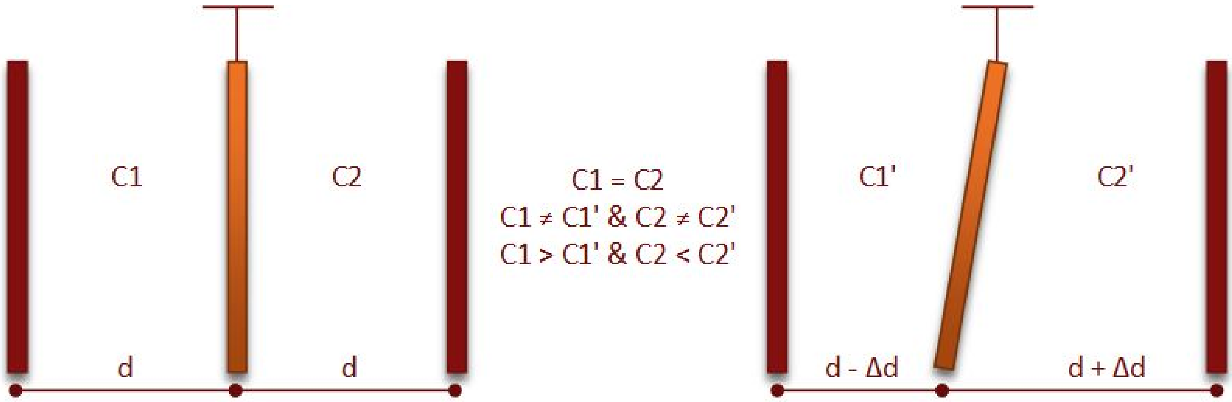

Figure 10 – Schematic structure of capacitive acceleration sensor

2. ***Piezoresistive***

Piezoresistive sensors utilize the fact that acceleration causes elastic deformation is certain materials, which is linear to an extent (to the elastic limit). The elastic limit is the maximum force that does not cause permanent deformation in a material.

We are measuring this effect by using resistors most of the time, based on two physical phenomena: resistance changes if the shape of the material changes, and specific resistance changes when the material undergoes elastic deformation (this is called the piezoresistive effect, and it is used in many areas, for example sound detection). The resistance change caused by the two effects results in an easily measurable electric signal. If we hang a mass (similar to the previous example with the common electrode), and attach strain-gauges (special resistor used to measure mechanical strain) to the attachment points, the amount of elastic deformation can be calculated from the resistance change of the gauges. From this we can easily

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

calculate the force that caused the deformation, and by that, the acceleration. This can be seen in the following figure:

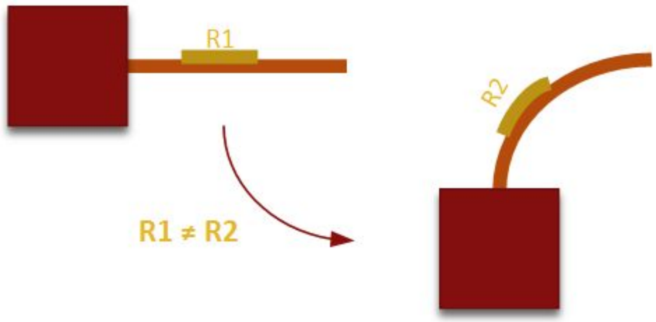

Figure 11 – Schematic structure of piezoresistive acceleration sensor

Gyroscopes

By attaching a gyroscope to a vehicle, we can measure the parameters of their motion through arcs, such as the rotation of the vehicle, or the angular velocity that caused the rotation. Gyroscopes are much more complicated devices than accelerometers, where we only relied on elastic deformation measured in two different ways.

Before discussing MEMS gyroscopes let's have a look at mechanical gyroscopes too! The first gyroscope has been invented in the mid-1800s, by Focault (you can remember him by his pendulum). Using one of the basic laws of physics, the conservation of angular momentum.

Angular momentum

Angular momentum is a physical quantity, which is defined by the product of a rigid body's moment of inertia and angular velocity. Its symbol is N. We also know that the net torque on a body (M) can be calculated by multiplying the moment of inertia (Θ) with the angular acceleration (β). From this, the following equation can easily be derived:

The angular momentum is the product of the rigid body's moment of inertia and the angular velocity, its symbol is an *N*. We also know that the resulting torque (*Me*) of the body can be calculated as the product of the body's moment of inertia and angular velocity. From these, the equation above can be easily deducted, which is the theorem of angular momentum:

$$M = \Theta \bullet \beta = \Theta \bullet (\Delta\omega / \Delta t) = \Delta N / \Delta t$$

We can see that the resulting torques causes a change in angular momentum if it is not zero. This also means that if the net torque on an object is zero, its angular momentum must be constant. This is the law of conservation of angular momentum. Because one side of the equation is a vector quantity (the torque), the angular momentum must also be a vector quantity. This way not only the magnitude of the angular momentum is conserved, but also its direction. In the next picture you can see these physical quantities, the vectors and their corresponding notation are depicted with identical color.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

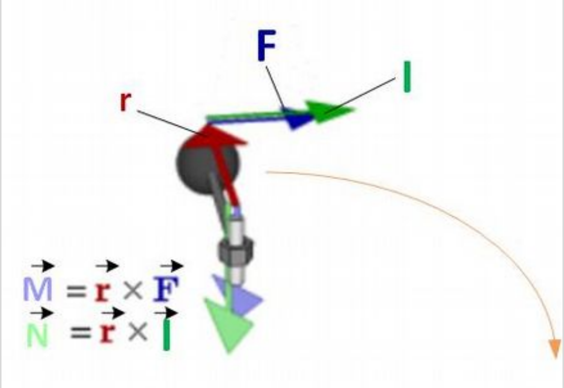

Figure 12 – Force, torque, momentum, angular momentum

In the figure above you can see a rotating ball, which is rotating in the direction shown by the orange arrow. The force (F) causing its movement exerts a torque on the ball with a magnitude of $F \bullet r$ where r is the distance between the axis of rotation and the point where F is exerted. The angular momentum of the ball (N) can easily be calculated in this case by multiplying the impulse of the object (I) with r ($r \bullet I$). Note that in the picture a cross is placed instead of a dot denoting multiplication. We won't go into the details here, but this cross means vector product. If the two vectors are perpendicular to each other their vector product equals to the product of their lengths, pointing in a direction perpendicular to both of them, according to the right-hand rule. In all other cases the direction of the product vector is still the same, but the magnitude is the product of the lengths of the two vectors, multiplied by $sin(\alpha)$, where α is the angle between them.

The mechanical gyroscope is based on a teetotum, on which we can apply the law of conservation of angular momentum, so it will always try to retain the direction of its rotation axis. Many factors can hinder it, but in a special case our teetotum can rotate freely along any axis. To achieve this we will have to fulfil some conditions:

- its speed and mass should be as big as it can be
- our teetotum should not slow down (i.e. it has to hold its speed) – despite the fact that the loss, generated by friction, decreases its angular velocity

If we can retain its speed, then the freely rotating teetotum can provide information in any direction of space, so it can be used as a sensor. These sensors are the three degrees of freedom gyroscopes.

Three decrees of freedom gyroscopes are commonly used as mechanical gyroscopes in moving vehicles to provide a measurable signal related to the spatial orientation of the vehicle. We can also say that if we know the direction in which the gyroscope was started, it will keep that direction, so we can compare any later orientation of the vehicle to it. We can see that this is again a relative measurement technique, because we can only measure orientation relative to a certain direction.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

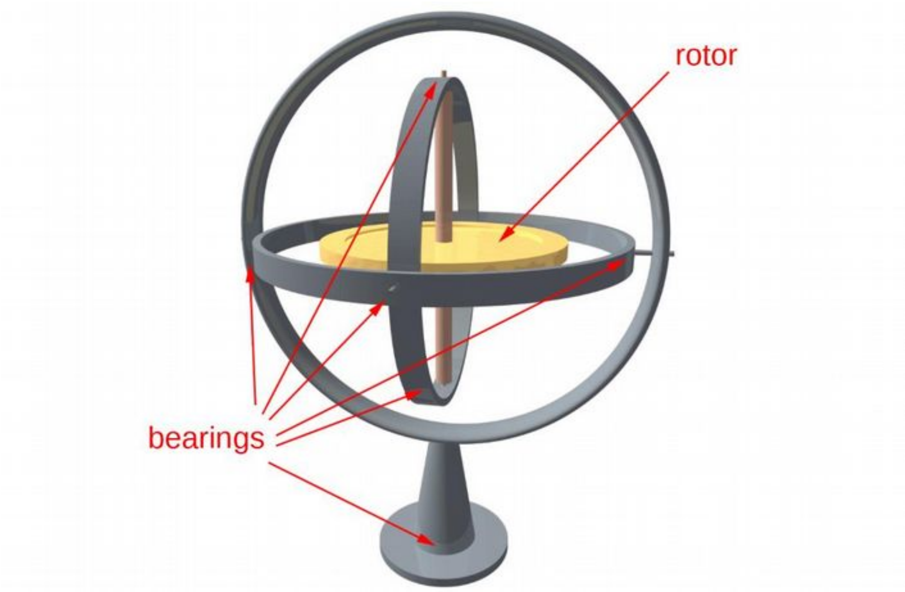

Figure 13 – Classic mechanical gyroscope

The gyroscope also contains electronic circuits which fulfil the physical requirements mentioned above. In many cases to maintain angular velocity, the spinning part of the gyroscope is a rotor of an electric motor, so it is trivial to maintain its speed. When using mechanical gyroscopes, the main problem usually comes from the large speed, large size, and short lifetime caused by the first two. Thankfully these problems don't always have to be solved.

A big mass puts a great stress on the gimbal frame holding the spinning disk, especially when the vehicle is subject to high acceleration. While the increased mass increases the precision of the measurement, it also subjects the device to changes in gravity. As you probably know gravity on Earth is not uniform, the gravitational constant is only approximately $10\ [\mathrm{m/s^2}]$, it changes depending on location. This is the reason that in vehicles travelling long distances (container ships, rockets) these changes in gravity cannot be ignored during design, or evaluation of measurement data.

The other important requirement was the high rotational speed. When increasing speed, the friction also increases, and the effect of friction reduces the lifespan of the sensor. Typically, in rockets the speed has to be very high, even with a large mass, but we don't care if the sensor lasts long or not, since no rocket has ever stayed up. These high-precision gyroscopes are not designed to last long, their expected lifetime is only a few tens of minutes tops. This is why usually these sensors are not as expensive as long-lasting sensors of airplanes or ships, which are very reliable, can withstand large forces, last long, and exceptionally accurate. If all these conditions are true it is not uncommon for such a sensor to cost 100 – 200 thousand dollars.

After learning about mechanical gyroscopes, we can talk about MEMS gyroscopes! You have seen that a mechanical gyroscope has a complex inner structure, which is partly the reason for their high price. Instead, MEMS technology brought simplicity, cheaper mechanical construction, and increased accuracy while also integrating the signal conditioning electronics into the sensor.

MEMS gyroscopes utilize the Coriolis-effect. By definition, the Coriolis-force is a force acting upon an object in a rotating coordinate system. It was first noted by the French mathematician Gaspard Coriolis in

Erasmus+
This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

1835. Calculating the effect of the force is complicated, because it is proportional to the angular velocity of the rotating coordinate system, the velocity of the object moving inside the rotating coordinate system, the sinus of the angle between the axis of rotation and the objects velocity vector, and the mass of the object. Its direction is perpendicular to the velocity vector of the object.

In MEMS gyroscopes a small resonating mass on a silicon substrate is used to measure the Coriolis-force. The mass can move in a certain direction, and if we rotate the substrate a force is exerted proportional to the angular velocity, and perpendicular the resonating direction. We can imagine this easily if we think of the substrate rotating from under the mass while it is resonating between the two endpoints, moving the mass out of its original direction. The amount of movement can be detected with one of the methods we have already talked about when talking about proximity sensors or accelerometer: measuring change of capacitance.

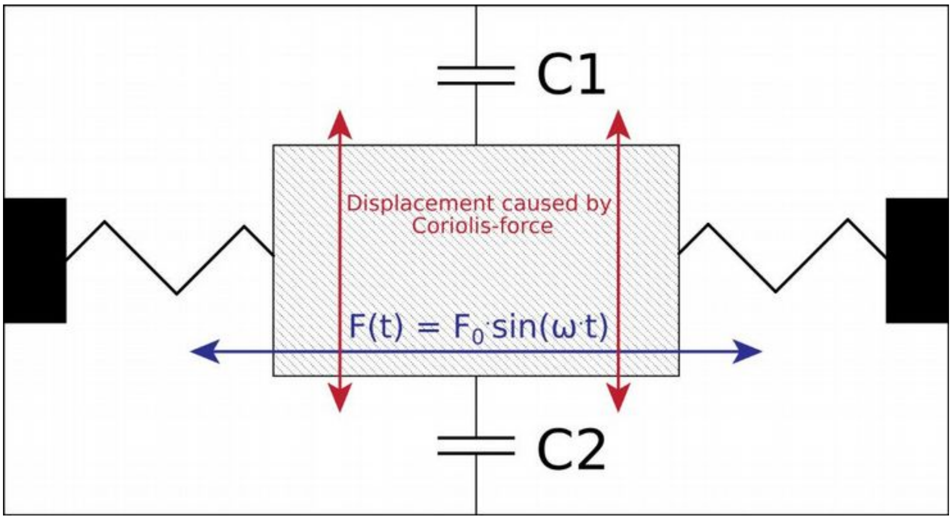

Figure 14 – Structure of the gyroscope which measures with a vibrating mass

The Coriolis-force creates a change in capacitance between the resonating mass and the side of the substrate. Without the Coriolis-effect the mass would not rotate and the capacitances C_1 and C_2 in the figure above would be the same. The picture below illustrates the size of such a MEMS gyroscope.

Figure 15 – Modern MEMS gyroscope
Author: SparkFun [CC BY 2.0 (https://creativecommons.org/licenses/by/2.0)]
https://cdn.sparkfun.com/assets/0/1/5/c/f/5112d377ce395ffd27000002.jpg

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

The is another important thing about MEMS gyroscopes and accelerometers that should be mentioned. These two sensors are frequently integrated to one sensor, which gives as an integrated sensor with six degrees of freedom. Degrees of freedom might be unclear for you right now, so we will talk about what it means through two examples.

The first example shall be the position of a car in a parking space. What do we have to do to know where the car is, and what is its orientation? First, attach a cartesian coordinate system to one corner of the parking space. We can easily tell the position of the car in this coordinate system, with x and y components.
I am free to place the car anywhere in the parking space where its dimensions allow it, I am free to choose along both axes. But by knowing the coordinates of the car's center of mass, I have not yet fully described its position, because I need to know the orientation too, which can be described by an angle of rotation.

I can also choose this angle freely, only the boundaries of the parking space are limiting factors. If the parking space was infinitely big, I could choose any angle as there would not by any walls. The conclusion is that I can place my car on a place by choosing 3 parameters, and I can do that independently and freely.

The independent quantities that describe the state of a system are called degrees of freedom (DOF). In the parking space example above, the system has three degrees of freedom (3DOF), because I can describe the state with three independent variables.

The second example is based on the first example but expanding it to three-dimensional space. Let's assume that we have cube of space and we have placed a helicopter (or a drone) into it.

Again, we are facing the problem, that the drone cannot be placed in any orientation in the corners of the cube, but let's ignore that, we could be talking about an infinite space too.

If we want to describe the position of the drone, we need a cartesian coordinate system, with the origin placed on one of the corners of the cube. For simplicity, assume that the axes are parallel to the sides of the cube (of course this does not matter, but it is easier to imagine this way).

In this coordinate system the x, y, and z coordinates can be used to describe the position of the drone. But this does not say anything about its orientation.

The drone can be rotated along any axis, independently to each other. So, it is easy to see that spatial position recognition can be done with six degrees of freedom (6DOF).

Plains reduce the space in such a way that planar motion only has three degrees of freedom.

I would like to apologize for the short mathematical intervention, let's go back to integrated MEMS accelerometers and gyroscopes with six degrees of freedom. The size of these is so small, that it is comparable to a smaller cherry, so they can be built into a lot of things making spatial position control easy. These sensors are called inertial sensors (IMU – Inertial Measurement Unit), and they consist of three gyroscopes and three accelerometers.

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Figure 16 – Six degrees of freedom inertial sensor
Author: SparkFun [CC BY 2.0 (https://creativecommons.org/licenses/by/2.0)]
https://commons.wikimedia.org/wiki/File:SparkFun_6DoF-IMU-Digital-Combo-Board_ITG3200%2BADXL345_10121-01d.jpg

A Little Robotics

For the end of the chapter we will take a little more interesting detour. At the time of writing this chapter I am working for a German multinational robotics company, so robotics is obviously close to my heart. In the following section you can read about the most important sensors in robots, the ones that measure the speed and position of joints and forward this information to the robot controller.

General Structure of a Robot Arm

When talking about robot arms, we mean electromechanical devices that mimic the human arm, can move along a pre-defined path, and can complete task along the way. Such a robot arm can be seen in the picture below. The job of a robot arm is generally to assume certain positions in the space it can reach and do some task there. To understand this, we need the definition of degrees of freedom again. If we want to be able to position and orient a robot arm freely in a confined space, we need at least six degrees of freedom (for the sake of simplicity, let's ignore for now that a robot arm has a finite length, so it cannot reach all points in space).

Can it have more? Of course it can, but it will require more complicated control while making our robot arm redundant (in theory at least, in reality the length of the robot arm poses constraints again), which means it can reach every point in space in multiple orientations. This is because there may be multiple combinations of robot joint positions that result in the same position and orientation or the end of the arm.

Erasmus+
This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Figure 17 – Picture of a real robot arm
Author: KUKA Systems GmbH [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)]
https://commons.wikimedia.org/wiki/File:Application_field_aerospace.jpg

The degrees of freedom of the robot arm is provided by independently controllable prismatic and revolute joints. Prismatic joints provide linear motion in a single direction and revolute joints provide rotational motion over a single axis. You can see from the above picture that if you move a joint, every other part of the robot arm that comes after it also move – this is just one of the many problems we need to face when controlling robots.

One of the most computationally expensive tasks of robot control is to calculate the joint positions to any spatial position and orientation of the robot arm. If we want the end of the arm to exert a force or torque in said position too, that requires another calculation requiring even more computational power. To completely scare anyone away from robotics I just want to note that robots are complex, non-linear system, whose precise control requires dynamic, time-varying mathematical models. This means that at every moment we have to solve a different set of equations, and these usually only give an approximate result. This model comes from the momentary state of the robot as a physical system.

Now you can see that if we put some computational uncertainty into this system by not sensing joint positions accurately, the whole control becomes unreliable, and inaccurate. Each joint has rotary encoders on them, whose job is to accurately sense the joint position and report it to the control system. Of course, encoders are not only used on robots, but also anywhere where a rotational motion's position is important.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

INCREMENTAL ENCODERS

With incremental encoders we can measure the relative position and velocity of revolute joints. Outside robotics they can be used for example to measure the position of a conveyor belt. Incremental encoders measure relative rotation, while absolute encoders measure absolute position. Both sensors can be found in applications where high precision is a must.

Generally speaking, incremental encoders output a signal proportional to the amount of rotation. While detecting the rotation is usually easy, the sensor has complex electronics built inside to evaluate the data, so it is an integrated sensor. The simple sensor inside outputs a pulse-train according to the rotation, and the integrated circuity evaluates and transmits the data over a serial port.

Sensing

First let's talk about the pulse train providing the information. These sensors usually consist of a disk attached to an axle, with periodic holes along the edge (or black marks on a transparent disk). A light source is placed on one side of the disk, while a light sensor is placed on the other (remember the proximity switches!). When the disk is rotating with a constant velocity, we get a square wave with constant frequency, where the frequency of impulses is proportional to the rotation speed of the disk.
This behavior can not only be achieved with optical devices, but there are also cogged disks, and magneto resistive encoders.

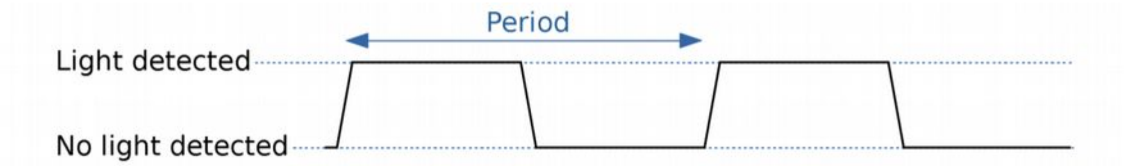

Figure 18 – Pulse train describing speed

The main disadvantage of the measurement is that we cannot determine the direction if rotation as the pulse-train looks exactly the same no matter in which direction the disk is rotating. To detect direction, more pulse-trains are generated by using more light sources and detectors (or a single light source with multiple detectors) in such a way that the two pulse-trains are skewed by one-quarter of a period. This way when the disk is rotating, we can determine the direction of the rotation by comparing the two pulse-trains to each other. This is illustrated in the figure below with pulse-trains A and B.

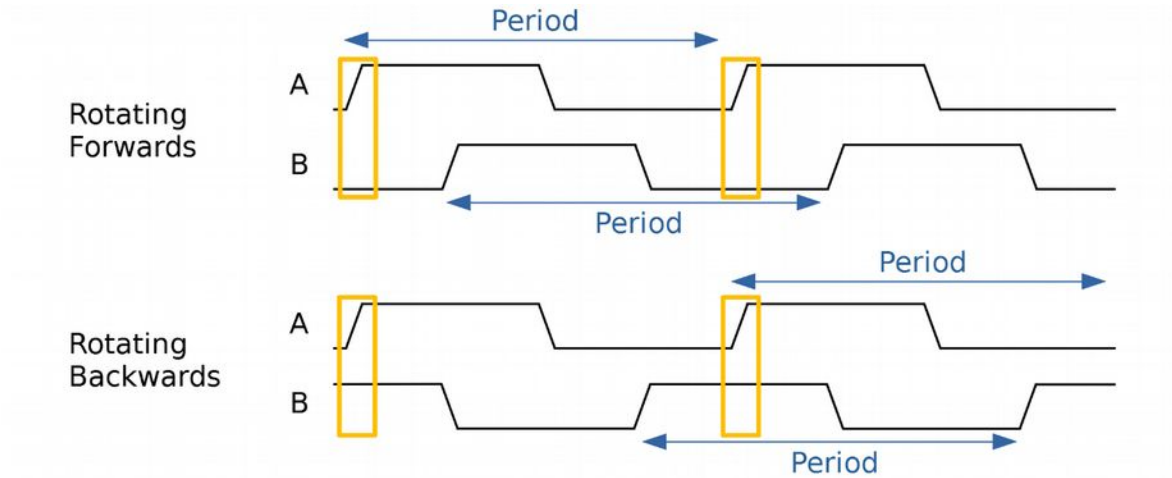

Figure 19 – Detecting rotation direction with incremental encoder

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

What happens when the disk is accelerating? Then, the frequency of the pulse-train is constantly changing, even during a period. Therefore, are not measuring the quarter-period skew between the two signals, but instead we measure the logic level of one signal when the other one has a rising edge. You can observe this highlighted with orange in the figure above.

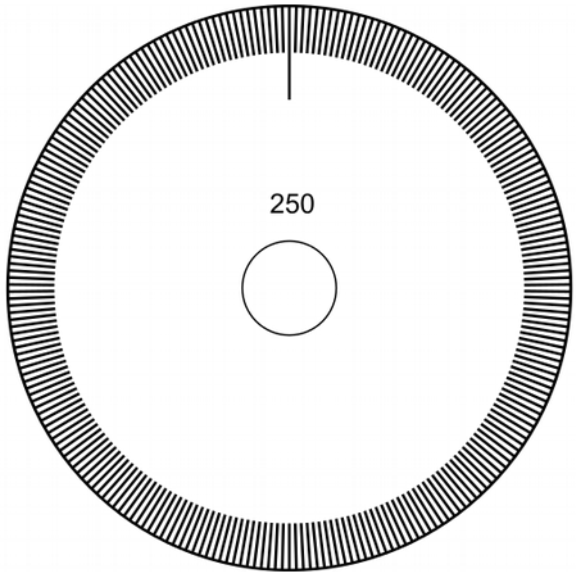

Figure 20 – Glass rotating disk with opaque metal marks
Author: Saure [CC BY-SA 2.0 de (https://creativecommons.org/licenses/by-sa/2.0/de/deed.en)] https://commons.wikimedia.org/wiki/File:DMT_Winkelkod-inkr.svg)

The simple sensor that outputs the pulse-trains is usually sold in a closed metal housing that protects the disk and the sensors inside from the environment. This is exceptionally important with optical sensors where contamination or external light reaching the sensors can affect the measurement results. Since an enclosure is easy to manufacture, and the optical solution provides the highest resolution (we can create very thin markings on a disk by metal evaporation), this is the most common. In many cases the separate signals are not created by two separate set of markings on the disk, but by one set of longer markings and the quarter-period skew is achieved by positioning the sensors accordingly.

As I have already mentioned, incremental encoders are integrated sensors containing a simple sensor and evaluation circuitry. The latter solves the signal level matching towards our system, and processes the signals coming from the disk. Unfortunately, the signals coming from the disk in the real world are not such perfect square waves as we have shown in the figures, they look more like sinusoidal signals. These are evaluated by the circuitry inside, and the output becomes a nice square pulse-train. These types of incremental encoders are called impulse encoders. There are other types as well, where the signal is not conditioned, only level-matched to the voltage level of our system, providing with a lot more information, these are called analog encoders. We can use an analog-digital converter to analyze this signal in the resolution we want. (You can probably see that the two types only differ in that impulse encoders connect the sinusoidal signals coming from the disk to a comparator and thus have a two-bit output).

With all the information mentioned above we can determine the angle of rotation along a single axis. We can imagine this even if we haven't discussed it in detail, the angle is proportional to the number of impulses received. Before talking about the methods used for precise measurement, let's talk about how to circumvent the lack of absolute position information.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

Using the impulses, we can measure the relative rotation of an axle. In other words, we have no information about the current position of the axle. To illustrate the problem with a real-world example, think about having a robot-arm in a certain position, for example it just inserted a key into a lock. The lock has been opened, but before removing the key the system has shut down. After restarting the robot, it knows where in the program code the shutdown occurred, so the next step is rotating the key back and removing it. But it doesn't know the current position of the revolute joint holding the key, it doesn't know how many degrees the key was rotated back before the system shut down, so it doesn't know whether the key can be pulled out or not. In general, when controlling robots absolute position information is necessary.

To solve this problem, we can use external devices providing absolute information, such as potentiometers. The main issue with these is that their accuracy is orders of magnitude worse than the accuracy of encoders.

This is why this solution is not used in practice. Instead a third signal is attached to the incremental encoder, the null impulse. This is a single marking on the disk, which emits one impulse after every rotation, with the length of a quarter of the period of signals A and B.

The null impulse is only active once every rotation for a duration of a quarter-period of A and B. The precision with which the position in one revolution can be determined is the same as the relative precision of the measurement.

With this solution we still don't know how many degrees the axle has rotated in total, but we can use a potentiometer for that. It does not have the accuracy for absolute position measurement over the whole range, but it is accurate enough to detect in which rotation the axle is in currently.

By which revolution we mean that the axle with the encoder can rotate fully multiple times, but if we use a potentiometer to determine how many times it has fully rotated, we can use the encoder to determine the precise position inside that revolution. The two solutions together gives us an absolute position information.

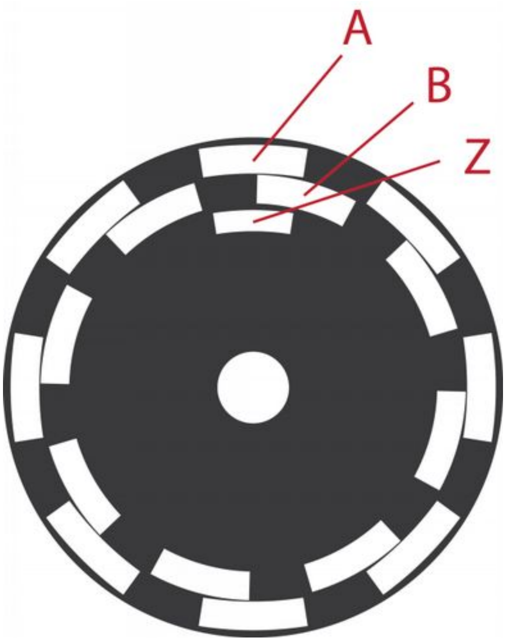

Figure 21 – Null impulse (Z) and rotating transmitter signals (A,B) on a disk

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

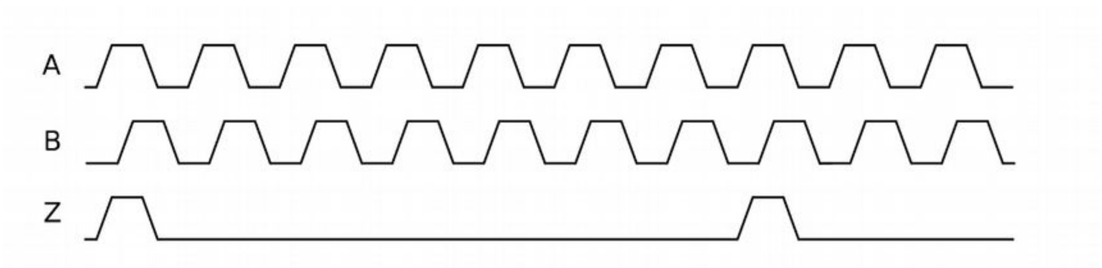

Figure 22 – Null impulse (Z) and rotating encoder signals (A,B)

Evaluation

Evaluating the signal can be a complex task, especially if we want to get as much information out of them as possible. The basic idea is to count the impulses of A and B signals, we can know the direction because of B, and we can know the revolution because of the null impulse. This may all be true, but the signal contains a lot more information than that. If during evaluation of one signal we don't take into account the changes of the other signals we are talking about single evaluation.

Double evaluation means when we are counting both the rising and the falling edges of A, so we increment our counter variable twice for every impulse. This obviously increases precision since we are counting two marks inside a single period.

The best accuracy can be achieved by quadruple evaluation. In this case we are counting both the rising and falling edges of both A and B. If we remember, the null impulse is only active for a quarter-period, and by quadruple evaluation we can detect rotations with the accuracy of relative position measurement.

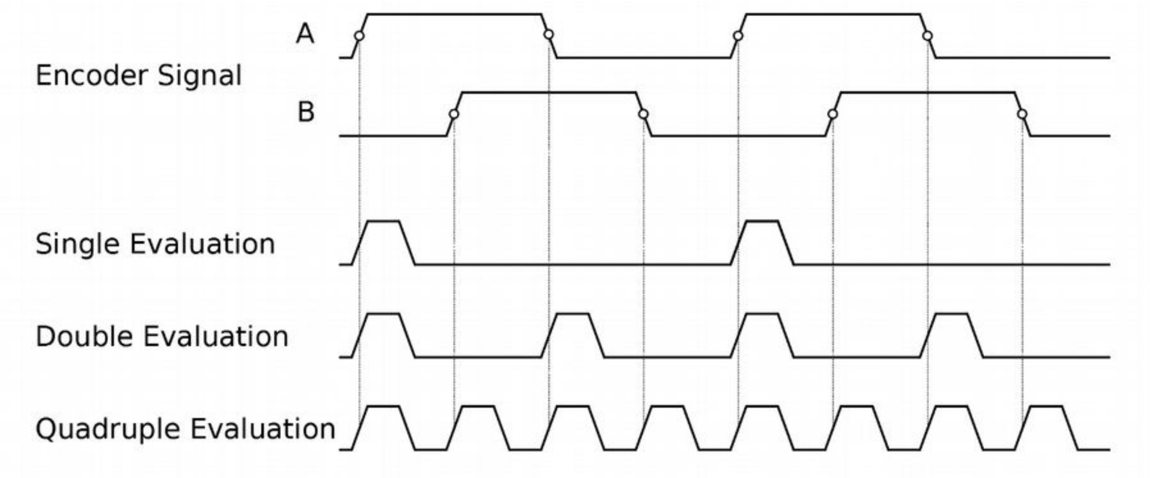

Figure 23 – Evaluation methods

It's important to know that the number of impulses generated by the signals, and the increments counting them are not necessarily equal. During an impulse we can watch for the rising and falling edges, or its timing compared to another signal.

The evaluation methods we have discussed so far have the side effect that the impulses are very high speed. Trying to process them in a microcontroller would be impossible. The constant interrupts would load the processor, it would "get stuck" in the interrupt handler routine and it wouldn't have time to do anything else. To avoid this, microcontrollers can have a dedicated peripheral specially designed to process signals coming from incremental encoders, this is called the direction detection unit.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).

The direction detection unit can be described as a peripheral which takes the signals of an incremental encoder connected to its inputs, determines the direction of rotation, counts the number of pulses, and detects the null impulse. To do all this it utilizes a counter, which can be increased and decreased according to the incoming pulses and set to a value when the null impulse is received. The peripheral can generate an interrupt toward the microcontroller, but it also makes asynchronous, non-interrupt related reads possible. The processor can write the counter variable too which can be useful when loading stored positions. The peripheral can include error detection which validates the value of the counter based on the incoming impulses and raises an error if an anomaly is detected (e.g. invalid counter value when the null impulse is received).

Advantages and Disadvantages

Let's talk about the advantages of the sensor first. The main advantage is high precision, which can sometimes mean 50-100 thousand pulses in each revolution with quadruple evaluation. Every position can be perfectly reproduced, and the measurement is reliable because of the enclosed housing. Integrating the sensor is easy thanks to the integrated signal level matching, and it can be easily used with a dedicated peripheral.

The solution has disadvantages too of course. The main disadvantage is the lack of absolute position, and the need for evaluation circuitry to get absolute position information. Related to this we can mention that after turning the system on, there is no information about the current position, we need to rotate the axle to get any kind of data. When designing the dedicated peripheral, we must be aware that losing an impulse can cause serious damage, it may make any later position information inaccurate.

ABSOLUTE ENCODERS

Absolute encoders are similar to incremental encoders, but they provide absolute position information inside one revolution at any moment in time. This means that position information is available without rotating the axle. This already solves two problems we were having with incremental encoders. The figure below shows the disk of an absolute encoder (left) compared to an incremental encoder (right).

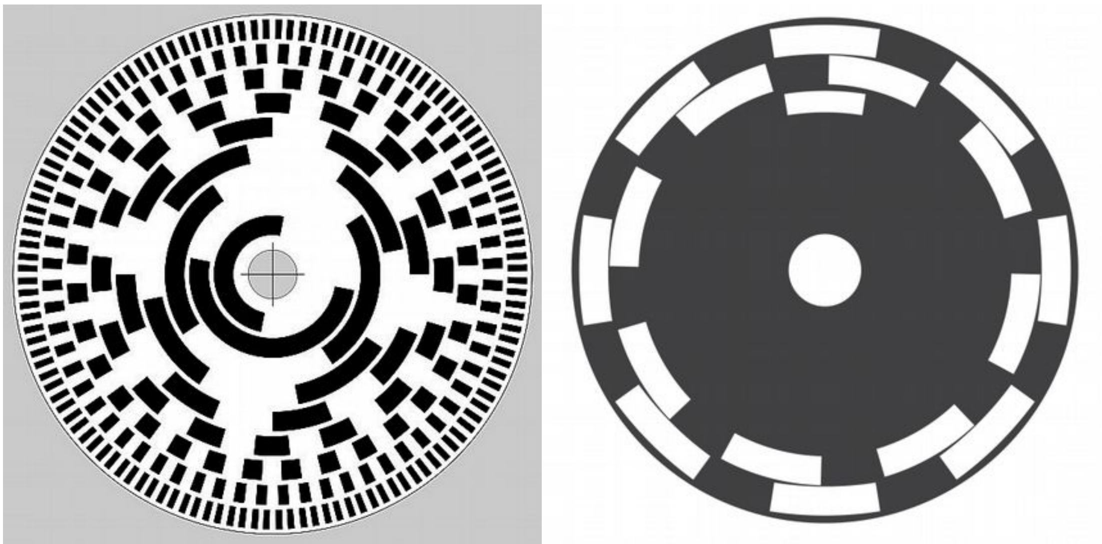

Figure 24 – Disks of absolute and incremental encoders
Author (left): W.Rebel [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)]
https://commons.wikimedia.org/wiki/File:Gray_disc.png

The difference between absolute and incremental encoders can be easily seen when looking at the disks. The absolute encoder reads a binary number at any position.

To do this it needs as many impulse sources as the number of bits in the position information. In the figure above you can see that along a single radius line of the disk the markings give a binary number according to the sensors detecting a signal or not (i.e. with optical sensing, the light is blocked or let trough)

The absolute encoder seen in the picture above only has markings along 8 arcs, which means it gives an 8-bit number even when the axle is not rotating. The outermost ring represents the lowest bit (LSB), and the innermost ring represents the highest bit (MSB), which means that the smallest detectable rotation is when the outermost ring has a signal level change, while the inner ring will only change after half a turn.

Since we are not talking about only counting pulses and increasing a counter, but detecting absolute position information, interfacing the sensor is not that easy. We are not talking about transmitting impulses but transmitting 8-bit (in this case) numbers. To do this we usually need some kind of high-speed serial communication. Most modern robots use absolute encoders to detect joint position.

Closing Words

We have finally reached the end of the long chapter. I hope, we hope with my co-authors, that you have found this chapter useful despite it being a more theoretical and less practical introduction to the world of sensors. You will need this knowledge of sensors to understand the next chapter since no closed-loop control can be realized without measuring the output with the required precision and feedback. And to those who understand the methods of sensing physical quantities, and the complex control systems built around them, there is no impossible task.

Erasmus+

This project was supported by the European Commission. The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the author(s).