Imperial College London

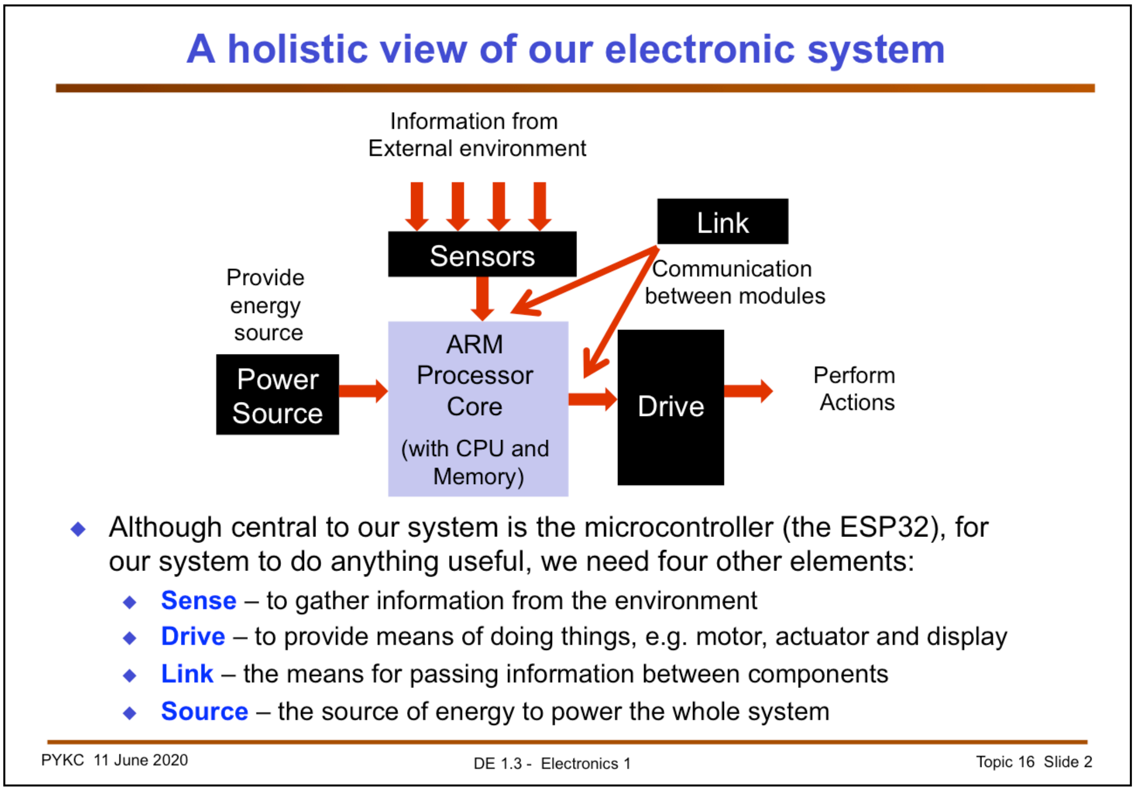
Topic 16

Sense

Prof Peter YK Cheung
Dyson School of Design Engineering
Imperial College London

URL: www.ee.ic.ac.uk/pcheung/teaching/DE1_EE/
E-mail: p.cheung@imperial.ac.uk

In this lecture, we will look at how physical parameters such as temperature, position, sound etc. are converted into electrical signals.



In this lecture we will examine how a microprocessor based system interact with its surrounding. This high level view of an intelligent electronic system consists of four main components:

1. **Sense** – How are external physical parameters such as temperature be converted into electrical signals? This is concerned with information INPUT to the system.
2. **Drive** – How can we produce motion, light and other actions? This is concerned with OUTPUT of the system.
3. **Link** – How can information be transferred between different electronic modules? This is concerned with information COMMUNICATION of the system.
4. **Source** – How is power be provided to the system and how can the power be used efficiently? This is concerned with the ENERGY supplied to the system.

Sensors

- To be useful, systems must interact with their environment. To do this they use sensors and actuators, which are examples of transducers.
- A transducer is a device that converts one physical quantity into another.

- The important parameters of senor performance are:
  - **Range** – maximum and minimum values that can be measured
  - **Resolution** – smallest discernible change in the measured value
  - **Error** – difference between the measured and actual values, which can be random errors or systematic errors
  - **Accuracy** – accuracy is a measure of the maximum expected error
  - **Precision** – a measure of the lack of random error (scattering)
  - **Linearity** – maximum deviation from a ‘straight-line’ response, normally expressed as a percentage of the full-scale value
  - **Sensitivity** – a measure of the change produced at the output for a given change in the quantity being measured

PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 3

Almost any physical property of a material that changes in response to some excitation can be used to produce a sensor

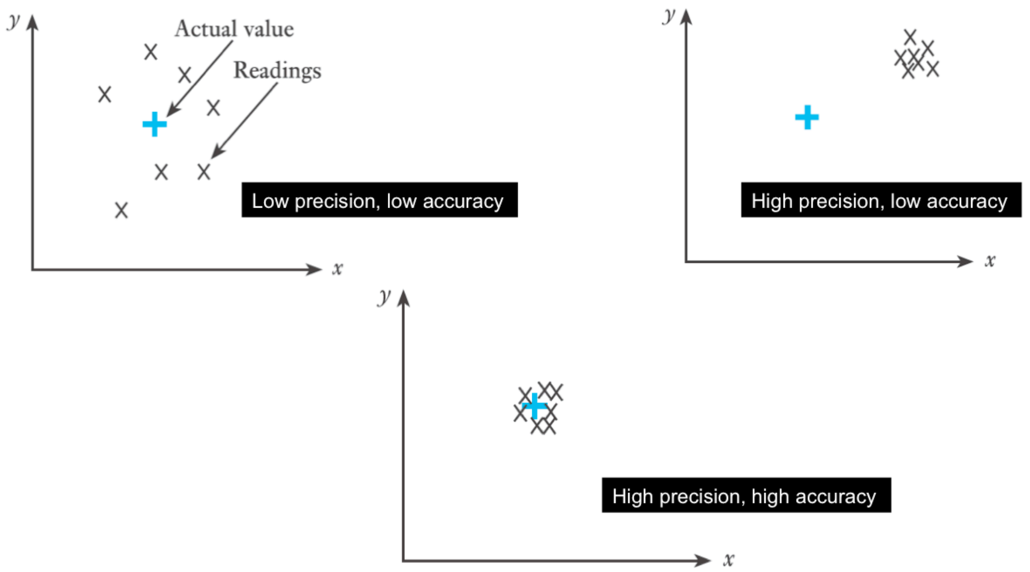
Widely used sensors include those that are:

- Resistive – change in resistance by the environment
- Inductive – change in inductance by the environment
- Capacitive – change in capacitance by the environment
- Piezoelectric – generate electrical signal by change in pressure
- Photoresistive – generate electrical signal by light, affecting resistance
- Elastic – gnerate electrical signal by elongation
- Thermal – generate electrical signal by change in temperature

Before we consider the different type of sensors, it is important that you understand the different performance measures related to any sensors. These are listed above. The most important are range, resolution and accuracy.

The Difference between precision and accuracy

- Precision and Accuracy describe two very different properties as illustrated in the graphs here:



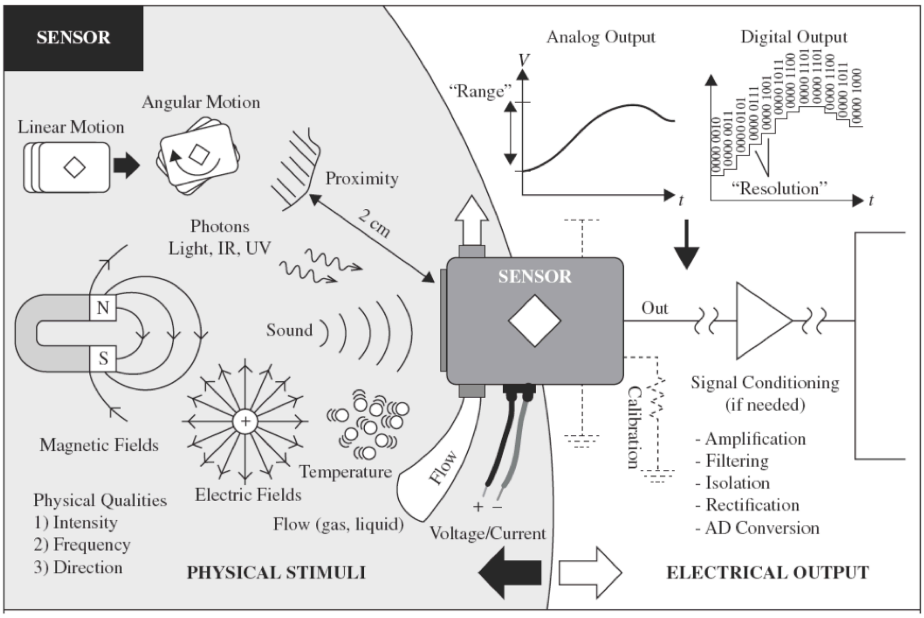
PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 4

Beware of the difference between precision and accuracy. These three plots illustrate their differences clearly.

A sensor with high precision would provide consistent readings (clustered), but this may not give you the correct answer (centred away from the true value).

For sensor whose converted to a digital reading, you must also understand the meaning of resolution (12 bit digital data has a resolution of 1 in 2^{12}).

Overview of sensor and its interface



There are many different types of sensors as summarized above. The output of the sensor is an electrical signal that may be analogue or digital in nature.

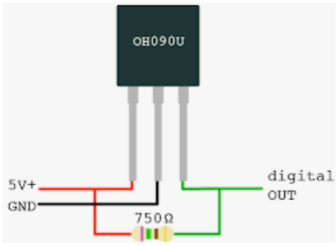
For analogue signals, they have to be converted to digital form using a analogue to digital converter (ADC) before these values can be used for computation and decision making.

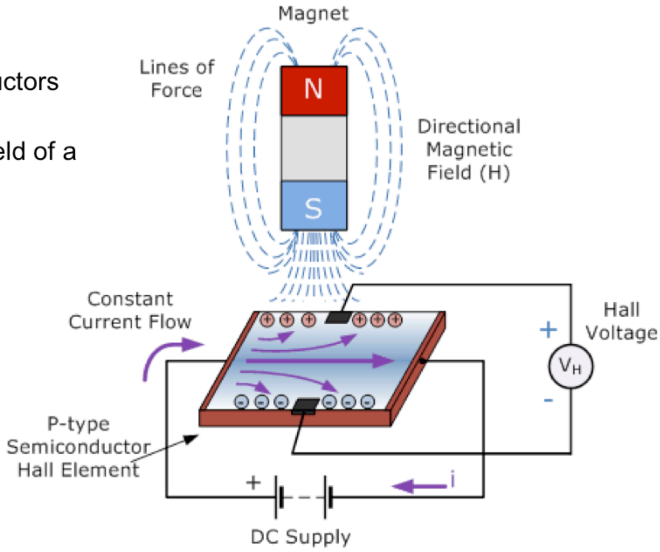
The ESP32 hastwo 12-bit ADC to converter an external sensor signal that has a voltage range of 0V to 3.3V. However, its accuracy is significantly worse than 12-bit!

Magnetic Field sensor – Hall Effect

Hall effect sensor

- Manufactured similar to transistors in semiconductors
- Create a voltage in the presence of magnetic field of a certain pole
- Simple, cheap, reliable





You have already been using a Hall Effect sensor to detect the presence of a magnetic field. The device is made from semiconductor material with a technology that is similar that of making integrated circuit components. The two surfaces of the sensor is sensitive to the two separate poles of a magnet as you have experimented in the Lab.

Hall effect sensors are very low cost and is widely used in detecting displacement or rotational speed of a motor, as will be seen later.

A video about Hall Effect sensor

This video can be found on this weblink:
https://www.youtube.com/watch?v=Scpi91e1JKc

Displacement Sensing – Resistive, Inductive, Switch

Potentiometers

- Resistive potentiometers are one of the most widely used forms of position sensor
- Can be angular or linear
- Consists of a length of resistive material with a sliding contact onto the resistive track
- When used as a position transducer a potential is placed across the two end terminals, the voltage on the sliding contact is then proportional to its position an inexpensive and easy to use sensor

Inductive proximity sensors

- Coil inductance is greatly affected by the presence of ferromagnetic materials
- The proximity of a ferromagnetic plate is determined by measuring the inductance of a coil
- Inductance changes resonant frequency of a LC tuned circuit – hence easy to detect if something is near.

Digital displacement senor

- Fancy name for a switch!
- Needs contact
- Easy to understand and implement - cheap

We often need to measure displacement. For example, in Lab 4, the outputs from the joystick controller is a displacement measurement device based on potentiometers (i.e. variable resistors). The two ends of the potentiometers are connected between a voltage source, and a carbon contact is made somewhere between the two ends. Since the resistor divides the supply voltage linearly (or this can also be in log scale), the displacement of the carbon contact is directly translated into electrical voltage (which is DC, but can also be AC).

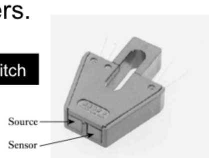
Another type of displacement sensor is based in inductance. This has the nice property that there does not require physical connection between what is displaced and the measurement device. Moving an object in and out of the centre of a coil changes the inductance. This could be measured in directly as the change in the resonant frequency of a LC tuned circuit.

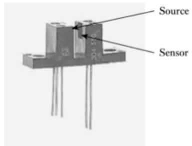
Perhaps the most common displacement sensors are based on optical means and are optical.

Displacement Sensing – Optical

Optical switches

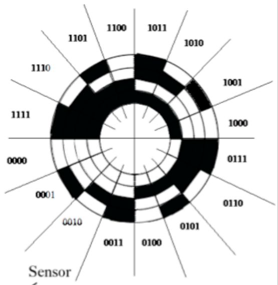
- Consist of a light source and a light sensor within a single unit.
- Typically using light emitting diode (LED) and light sensitive diodes (or photodiodes) as transducers.

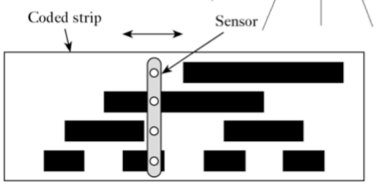




Absolute position encoders

- A pattern of light and dark strips is printed on to a strip and is detected by a sensor that moves along it.
- The pattern takes the form of a series of lines as shown here.
- Or as a disk with black/white pattern in grey code (neighbouring code only change by 1 bit).
- The combination is unique at each location.
- Sensor is an array of photodiodes.





Optical displacement sensors can be very simple – an light emitting diode, whose light is reflected (or not) and detected by a photo-sensitive diode or transistor. This normally gives a digital output, but it could also be arranged to measure the reflected light intensity.

Far more common is to use such devices in conjunction with a black/white coded strip or disk. This may also be implemented using slot cut-outs (black) or no cut-out (white).

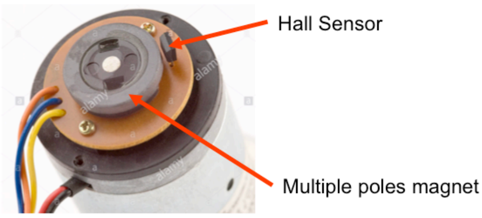
Displacement Sensing – with and without direction

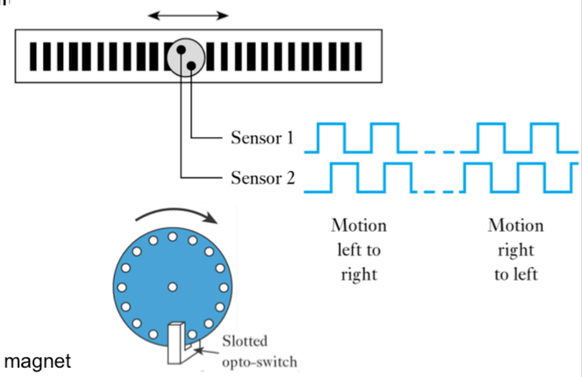
Incremental position encoder

- Uses a single line or a circular disk with alternating black/white bars (or slots)
- Can use simple slotted optical switch with a disk with slot and counter pulses – no direction.
- To know the direction, use two slightly offset sensors produce outputs as shown below. This detects motion in either direction, pulses are counted to determine absolute position (which must be initially reset)

Hall effect sensor

- We also use Hall Effect sensors to detect rotational displacement as shown here.



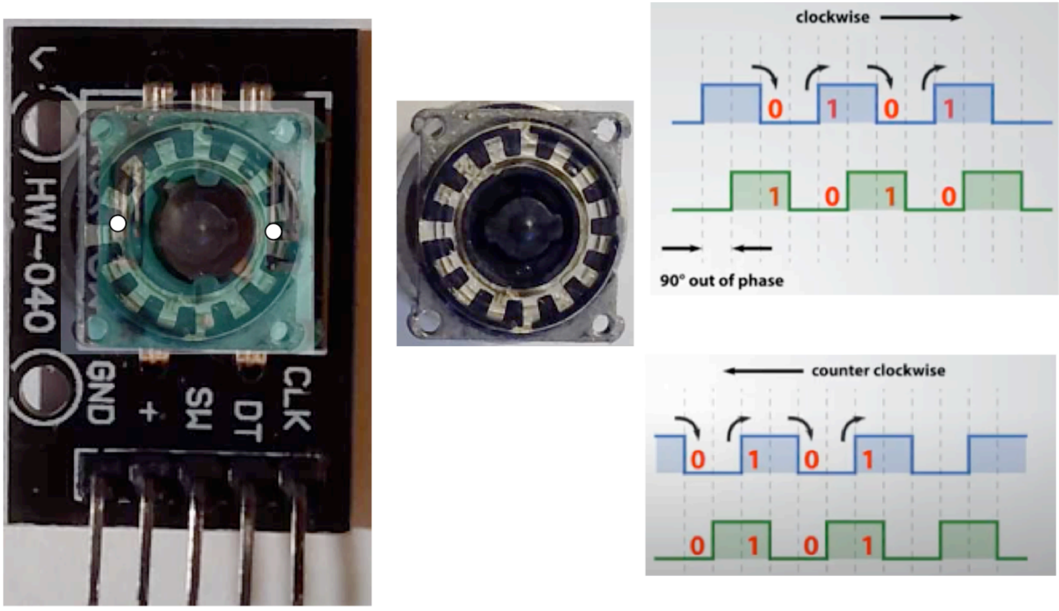


PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 10

With slots or black/white pattern, you can use a photodiode or a phototransistor to detect presence of absence of light in order to detect displacement. Shown here on the right is a sensor with TWO diodes slightly displaced from each other. The black-and-white strip is moving in a linear direction. The two diode sensors will produce two signals offset in phase as shown in the waveform. This arrangement allows you to determine the direction of travel of the linear motion based on whether sensor 1 signal leads or lags sensor 2 signal.

Another popular alternative to using light is to use Hall Effect sensors. You often find motors equipped with this method for rotational speed detection. Attached to the axle of the motor is a multi-pole circular magnet. Close to the magnet are two Hall sensors, again slightly displaced. As the motor rotates, the sensors emit pulses similar to that of the optical method.

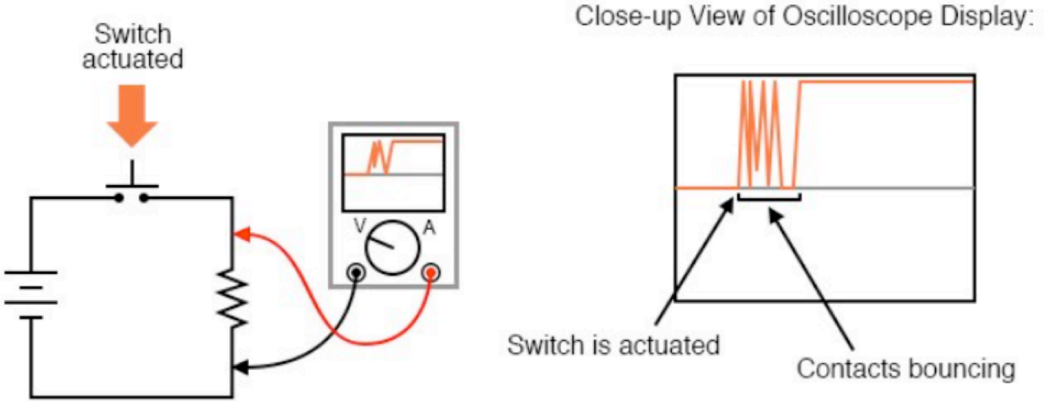
KY-040 Rotary Switch



This is the rotary switch that we use for our Home Lab Kit. You can see that it consists of a disk on the spindle with alternating metal and insulation slots. The base part has two contacts at different positions such that their contact with the metal and insulation part of the rotating disk is out of phase.

The timing diagrams shows that the two signals (CLK and DT) are indeed out of phase by 90 degrees (1/4 of a cycle).

Contact Bound in switches



PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 12

The rotary switch has the rotational knob, and also has a switch that one can activate by pressing down. This action short circuits the SW pin to the GND pin.

Any contact switches (as suppose to optical switches_ suffer from a phenomenon known as “contact bounce”. Immediately after a switching action, the contact will bounce up and down for a short duration (could be as much as 10’s of msec) before settling to the final position. This results in a signal with transient as shown.

Debouncing a switch can be achieved by two methods:

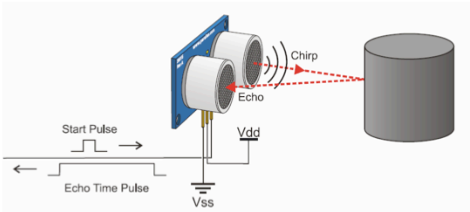
1. Adding a suitably sized capacitor – capacitor voltage cannot change suddenly. Therefore it forces the signal to rise towards its final value slowly and avoid the bounce.
2. When detect the first change (say from 0 to 1), delay for a period of time longer than the bounce period before reading the next value.

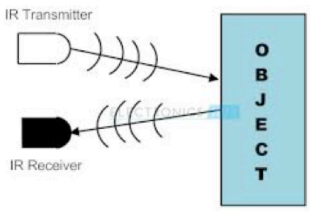
You will experiment with contact bounce in Lab 4B.

Distance Sensing – Echo location

Time-of-Flight Sensors

- Almost all distance sensors are based on time-of-flight principle.
- A source signal is sent as a burst of pulses, and the echo is detected.
- Distance is derived using the delay time between the source signal and the detection of the echo signal.
- Ultrasound is often used as a cheap and low accuracy time-of-flight sensor. Its effectiveness depends on the object surface property and orientation. Good for robot cars, but not for industrial applications.
- Instead of ultrasound, one could use infra red sources. Usually good for short distance (a few cm).
- Laser sensors are commonly used for industrial applications, due to their robustness, accuracy and low sensitivity to surface reflectivity and orientation.
- An alternative is to use infrared transmitter/receiver as you did in Lab 4.





Sensing distance can be achieved by echo location with ultrasound, not dissimilar to what a bat uses to sense its environment. This method requires a transmitting device sending either a pulse or a continuous signal, and a detecting device that senses the echo signal.

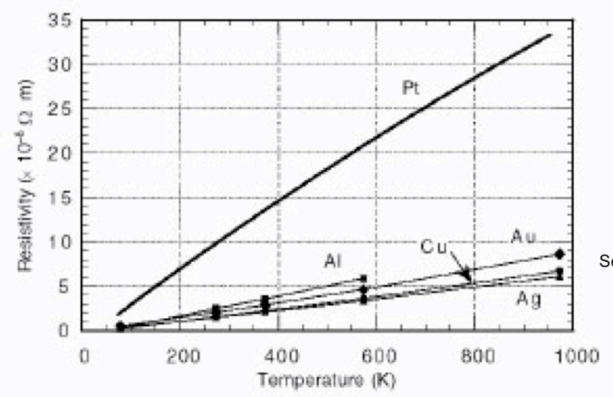
With the pulse method, a pulse burst (i.e. multiple pulses) are usually used. Then the time it takes for the echo to return is proportional to the distance of the reflecting surface.

Time-of-flight sensor can use ultrasonic transmitter and detector, or can use light of a laser diode. You can find low cost sensors of this type from many sources.

In Lab 4, we use infrared sensors, and only detect the presence or absence of an echo. The strength of the echo signal is related to the distance from the obstacle. An analogue comparator circuit (with an adjustable threshold) is used to determine if there is an echo or not.

Temperature Sensing - PRT

- Platinum Resistive Thermometers (PRT) - devices using platinum wire whose resistance changes with temperature
- Shown here is a plot of resistivity vs temperature characteristics for five different metals. It shows platinum has the highest sensitivity.
- PRT has good *linearity* but has poor *sensitivity* when compared to other types of temperature sensors.
- It works up to high temperature.



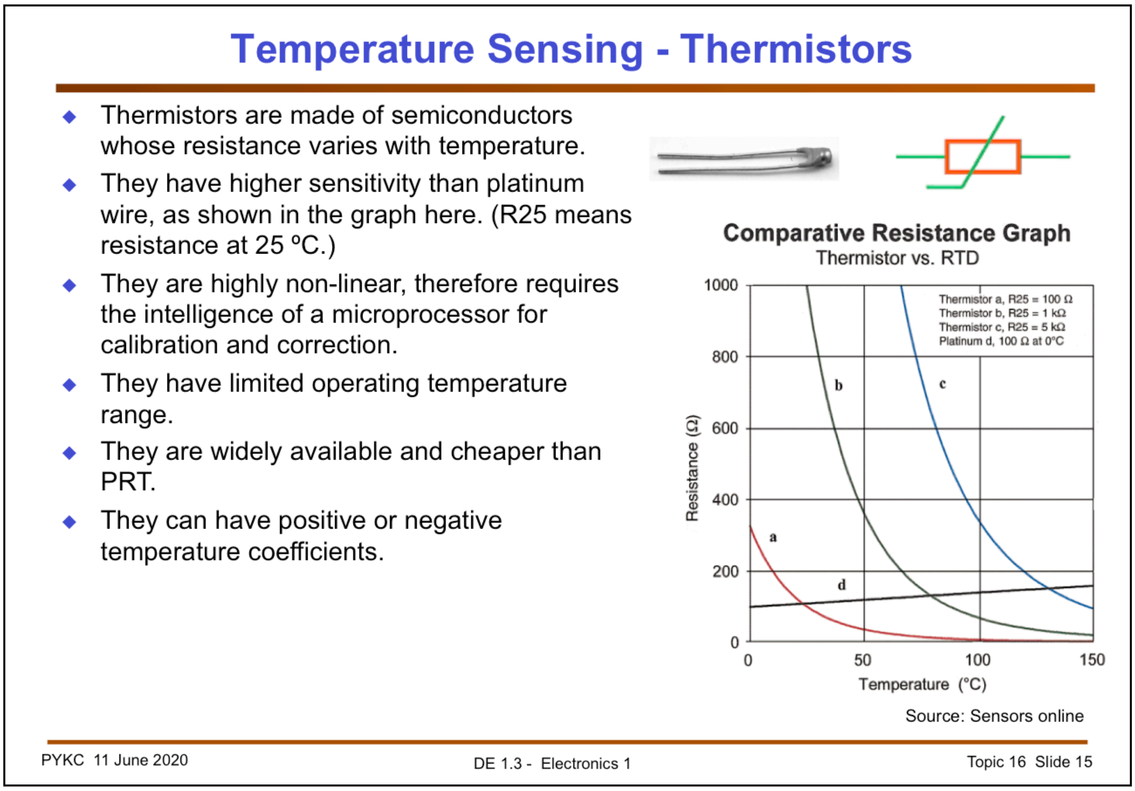
Source: NPL

The electrical resistance of many metals increases approximately linearly with absolute temperature. Therefore we can measure temperature by passing a current (AC or DC) through a metallic wire and measure the voltage potential across it. This type of sensors are collectively known as Resistive Temperature Detectors (RTDs).

Plotted above are the resistivity vs temperature characteristics for five metals: **platinum**, aluminum, copper, silver and gold. Platinum is the best among them with the highest sensitivity (i.e. highest dR/dT), not too soft and can be made into wires with uniform cross section. Such transducer made of Platinum wire is known as Platinum Resistive Thermometers (PRT).

The length and diameter of the platinum wire used in a thermometer are often chosen so that the resistance of the device at around 0 ºC is 100 ohms. Such a sensor is a called a PT100 sensor, and its resistance changes by approximately 0.4 ohms per degree Celsius. Using a typical 1 mA measuring current, at around 0 ºC a PT100 sensor would have a voltage drop of around 100 mV across its terminals and this would change by approximately 0.4 mV per degree Celsius. This makes measuring temperature really easy.

PRT has the advantage of working to a very high temperature (1000 ºC), and very linear. But it is also expensive and not very sensitive when compared to other types of transducers.



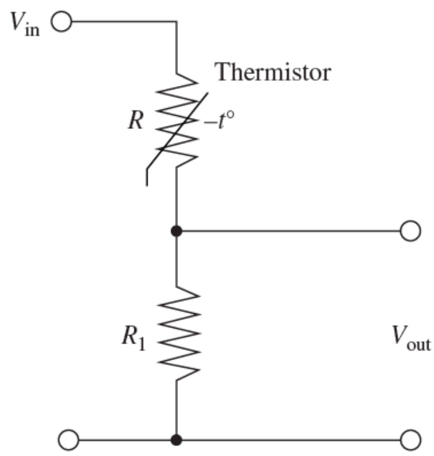
Thermistors are also resistance-based transducers made of semiconductor materials. They are typically composed of ceramic or polymer materials with high temperature sensitivity. Therefore they can be used in precision applications. However, the resistance versus temperature characteristics is generally highly non-linear. Therefore one would need to use a microprocessor to perform correction.

Thermistors are generally cheaper than PRT and are widely used in automotive, domestic, medical and food processing applications.

Thermistors have a relatively narrow working temperature range, limiting their applications as compared to PRT and other resistive devices.

Thermistors come in two main types: those with positive temperature coefficient (PTC) and those with negative temperature coefficient (NTC).

Here is a simple circuit that uses a thermistor (NTC) as a thermometer:



Temperature Sensing – pn junction

- pn junction diode is made of silicon semiconductor materials
- A diode only conducts current in one direction (when a positive end known as anode to negative end known as cathode), when V_D exceeds some threshold.
- When a diode is conducting, it is being forward biased.
- The diode voltage V_D changes by around -2mV/ºC – hence we can use this to measure temperature.

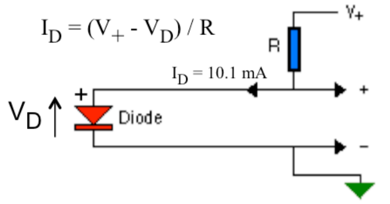
Advantage:

- Cheap or free – already available inside chips

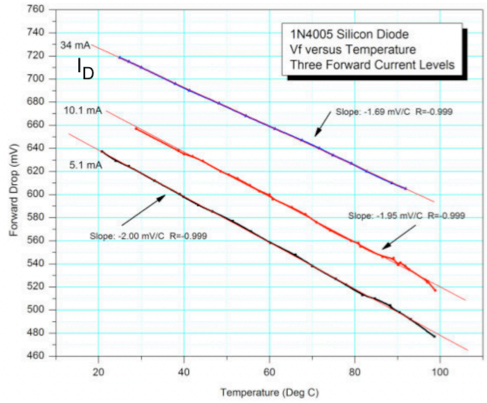
Disadvantages:

- Limited operating range
- V_D varies with current through diode, and from device to device – difficult to do accurate absolute measurements

- Generally useful to detect overheating – found in almost all semiconductor chips now (e.g. Pentium or ARM processors)



Diode as a Temperature Sensor



Semiconductor devices can also be used to measure temperature.

At a constant current bias, the voltage drop across a silicon P-N diode junction shows roughly a –2 mV/ °C temperature coefficient. Because the P-N junction is the basic building block of diodes, transistors, and ICs, temperature sensing can be incorporated in many devices at low cost. This technique is used in the onboard temperature sensors of microprocessors (e.g., Intel's Pentium) and for the thermal-shutdown circuits of power-supply chips.

The main advantage of pn junction as a thermometer is therefore is low cost. However there are many disadvantages:

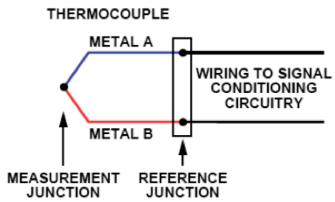
- Limited operating range – say from -20°C to 125°C
- Not robust – it is only semiconductor!
- V_D depends on current through the diode (i.e. its operating condition)
- Sensitivity and V_D value changes from chip to chip

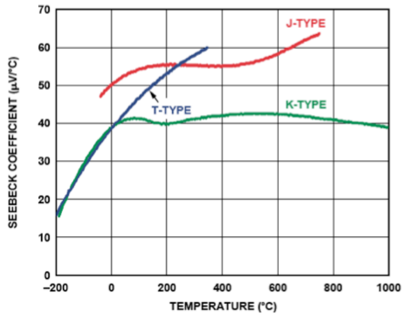
In general, we only use silicon diodes as a temperature sensor for the purpose of catastrophe avoidance. For example, all reasonably expensive chips are now equipped with pn junction temperature measurement circuits to detect overheating, and shut down the power supply voltage if a certain threshold is reached.

On microprocessor chips, (such as Intel processors), this measurement is also used adjust the supply voltage level AND clock frequency automatically if the chip runs too hot.

Temperature Sensing – Thermocouples

- Thermocouples are made with joining two types of metals.
- A voltage is developed at the junction of the two metals, and the voltage is temperature dependent.
- This coefficient (dV/dT) is known as Seebeck coefficient (the person who discovered this property).
- Advantages are:
  - High operating range: -200ºC to +2500ºC
  - Robust: just two wires wielded together!
  - Rapid response: small, low heat capacity, msec
  - No self-heating: passive device, not energised
- Disadvantages are:
  - Produces very small voltage – hence expensive signal conditioning
  - Highly non-linear, needs calibration/correction
  - Can corrode
  - Low accuracy – around ±1 to 2 ºC





Source: Analog Devices

PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 17

A thermocouple consists of two wires of dissimilar metals joined together at one end, called the *measurement* ("hot") junction. The other end is usually made of copper wires and is called the *reference* ("cold") junction.

Advantages of Thermocouples:

- High Temperature range: –200°C to +2500°C
- Robust - Thermocouples are rugged devices that are immune to shocks and vibrations
- Rapid response: Small therefore have low thermal capacity and fast respond time typically within a few hundred milliseconds
- No self heating: Require no excitation power, therefore not prone to self heating and are intrinsically safe

Disadvantages:

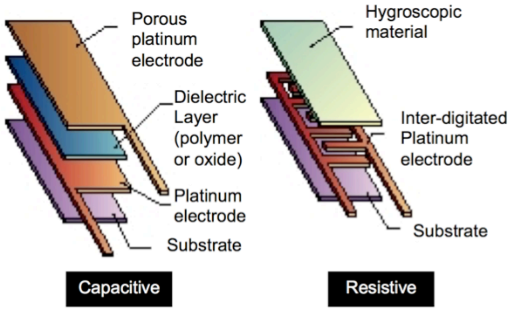
- Complex signal conditioning: Non-trivial to convert the thermocouple voltage into a usable temperature reading
- Accuracy: Generally only accurate to within 1°C to 2°C.
- Susceptibility to corrosion: Using of two dissimilar metals, hence more susceptible to environment
- Susceptibility to noise: Produce small change in signal in micro-voltage region, sensitive to unwanted noise signals such as main power frequency (50 or 60Hz)

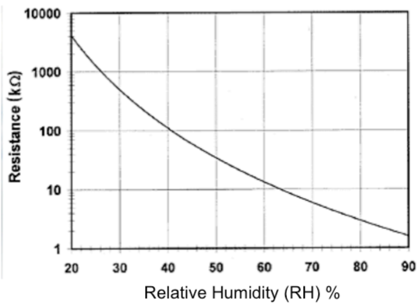
The most common thermocouple types are J, K, and T. At room temperature, their voltage varies at 52 µV/°C, 41 µV/°C, and 41 µV/°C, respectively. Type K has a relatively constant coefficient over the operating temperature (see plot).

A typical K-type is made of **Chromel** (90% nickel and 10% chromium) and **Alumel** (95% nickel, 2% manganese, 2% aluminum and 1% silicon).

Humidity Sensing – Capacitive and Resistive

- **Capacitive Humidity Sensors** – measured relative humidity through change in capacitances.
- Transducer is made of thin film of polymer or metal oxide deposited between two conducting electrodes.
- Sensing surface protected against contamination with porous metallic material.
- Change in capacitance is typically 0.2 to 0.5 pF/ºC and quite linear.
- Has low response time, typically in 10's of seconds.
- **Resistive Humidity Sensors** – based on interdigitated electrode with a deposition of a hydroscopic polymer coating on top.
- Resistance changes as an inverse exponential with humidity.
- Response time is also in 10's of seconds.





PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 18

Humidity sensors can be based on capacitance or resistance.

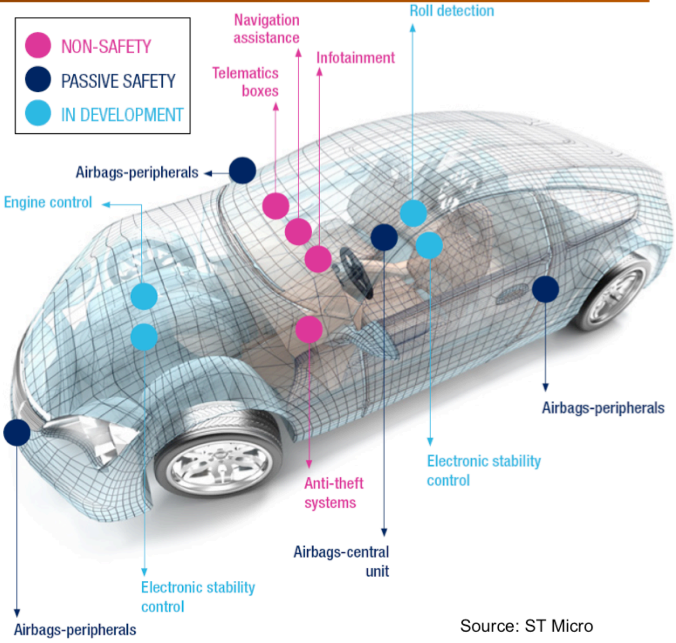
Capacitive sensors have a porous metallic electrode (often platinum) on top of a dielectric layer of polymer or oxide, sandwiched by anther metallic electrode to form a capacitor. Vapour content of the environment permeate through the porous top surface to affect the capacitance, which is measured to produce a electrical signal that is correlated to humidity.

In a resistive sensor, the electrodes are inter-digitated as shown in the diagram. (This means that there are two sets of comb-link fingers that interleave each other). The electrode has a hydroscopic polymer coating. This makes the resistance varies negatively with humidity – more humid, lower resistance (see graph above).

Both types of sensors have low response time of tens of seconds.

Introduction to Micro-ElectroMechanical Systems

- Combine electronics with mechanical functions on an integrated circuit.
- Often use same silicon process as making chips.
- Grew since the 80's, and now dominate the sensor area.
- Cars now have MANY sensors made from MEMS.



Source: ST Micro

PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 19

MEMS stands for **Microelectromechanical Systems**. Its definition (Sci Am Sept 95) is: "MEMS is the name given to the practice of making and combining miniaturized mechanical and electrical components."

It is also called: Micromachines in Japan and Microsystems technology in Europe.

MEMS devices use existing IC-based fabrication techniques (but now extend to other non IC techniques). Therefore it is potentially economical to make through batch fabrication. Thousands of MEMS devices (scale from ~ 0.2 μm to 1 mm) could be made simultaneously on a single silicon wafer.

Distinctive features of MEMS are:

Miniaturization

micromachines (sensors and actuators) can handle microobjects and move freely in small spaces

Multiplicity

cooperative work from many small micromachines may be best way to perform a large task

inexpensive to make many machines in parallel

Microelectronics

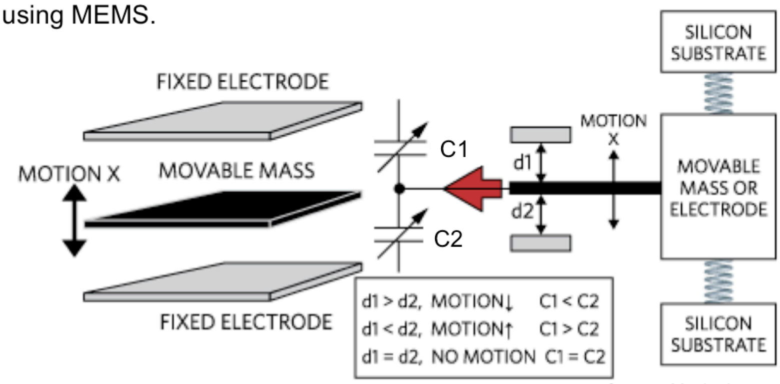
integrate microelectronic control devices with sensors and actuators

Modern car has many MEMS sensors, the most important and common is the accelerometer that triggers the deployment of the airbag on impact.

Motion Sensing – Accelerometer

Basic Principle

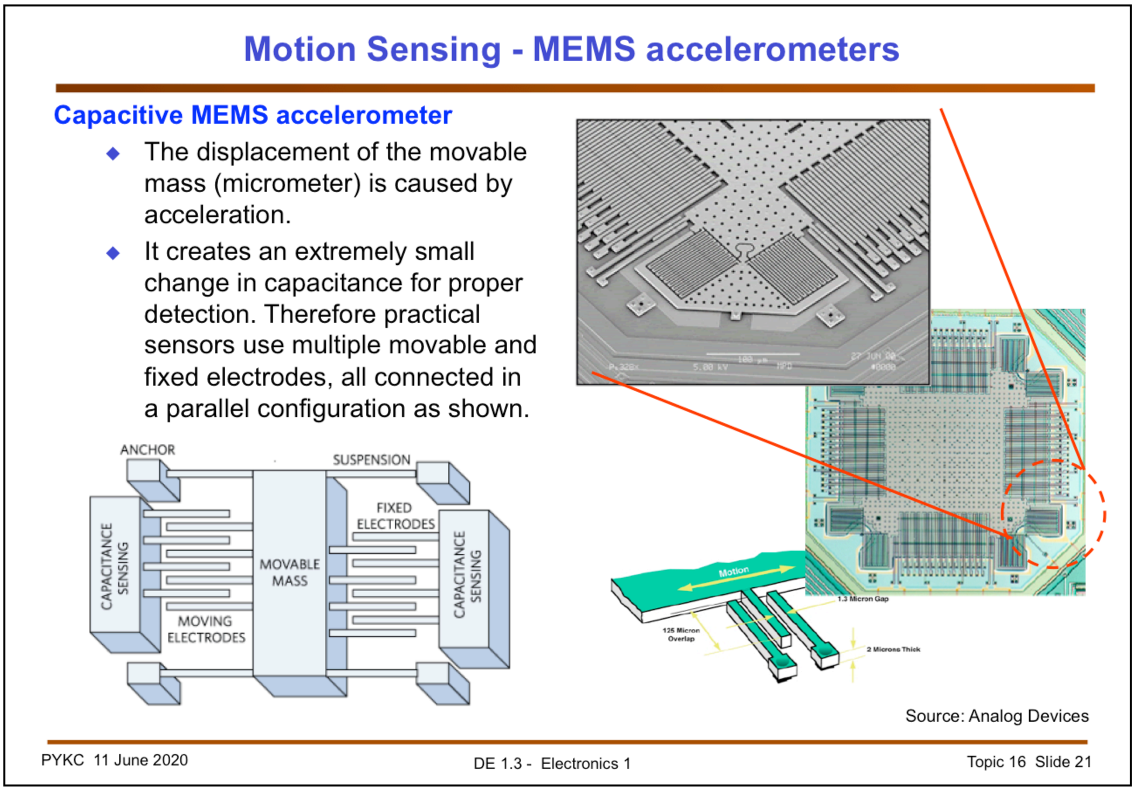
- Newton's 2nd Law of motion: F = mass x acceleration.
- Sense acceleration is really sensing the force on a mass.
- Use capacitive sensing with MEMS.
- Acceleration causes mass to move.
- Mass pivoted on springs anchored one side as shown.
- Implemented using MEMS.



Source: Maxim Integrated

PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 20

A common sensing approach used in accelerometers is capacitance sensing in which acceleration is related to change in the capacitance of a moving mass. This sensing technique is known for its high accuracy, stability, low power dissipation, and simple structure to build. It is not prone to noise and variation with temperature. Bandwidth for a capacitive accelerometer is only a few hundred Hertz because of their physical geometry (spring) and the air trapped inside the IC that acts as a damper.



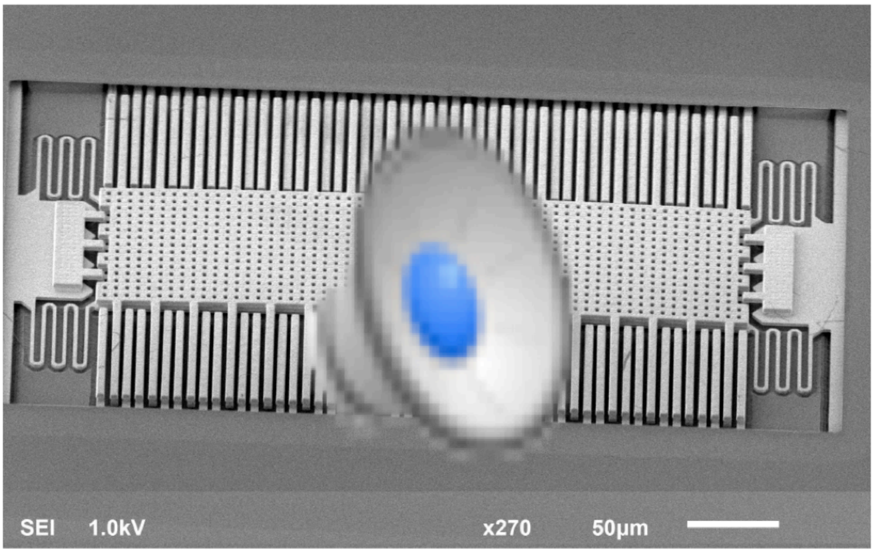
The displacement of the movable mass (micrometer) is caused by acceleration, and it creates an extremely small change in capacitance for proper detection. Therefore when implement such accelerometer on MEMS, many parallel capacitors are created as shown in the chip photo here.
The configuration enables a greater change in capacitance, which can both be detected more accurately, and ultimately makes capacitance sensing a more feasible technique.

Capacitive MEMS accelerometer

- High precision dual axis accelerometer with signal conditioned voltage outputs, all on a single monolithic IC
- Sensitivity from 20 to 1000 mV/g
- High accuracy
- High temperature stability
- Low power (less than 700 uA typical)
- 5 mm x 5 mm x 2 mm LCC package
- Low cost ($5 ~ $14/pc. in Yr. 2004)

MEMS based accelerometer chips now come with 3-axes, with three separate sensors whose masses move in three orthogonal directions: X, Y and Z.
Makers of accelerometers include: Analog Devices, Bosch, ST Micro, Texas Instruments.

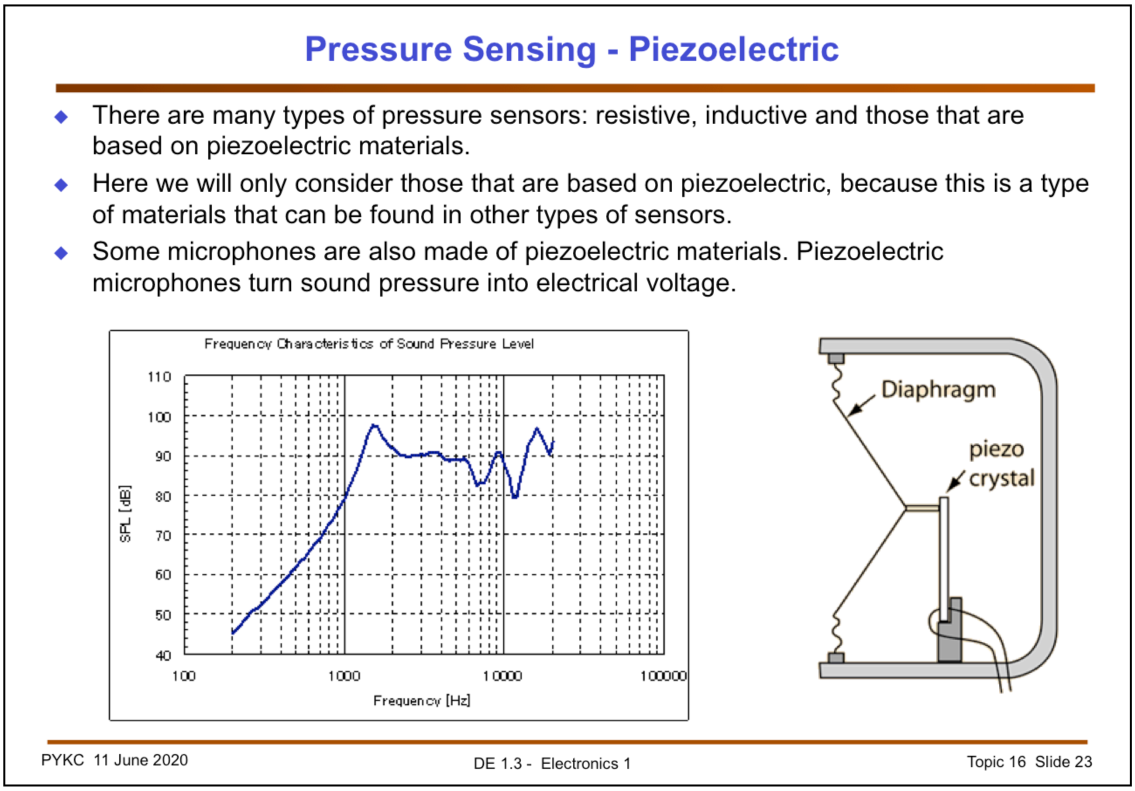
A short video on "MEMS Accelerometer"



A scanning electron microscope photo of a lateral accelerometer
Piotr Michalik et al, IEEE Senors, Nov 2015

This video can be found on:

https://www.youtube.com/watch?v=i2U49usFo10



Piezoelectric transducers are bidirectional – they are capable of converting stress into an electric potential and vice versa.

They consist of metallized quartz or ceramic materials. These transducers provide an output signal only when the input is changing. This means that these sensors can be used only for varying pressures. In that way, a piezoelectric transducer is like a capacitor – it blocks dc and passes (produces) ac.

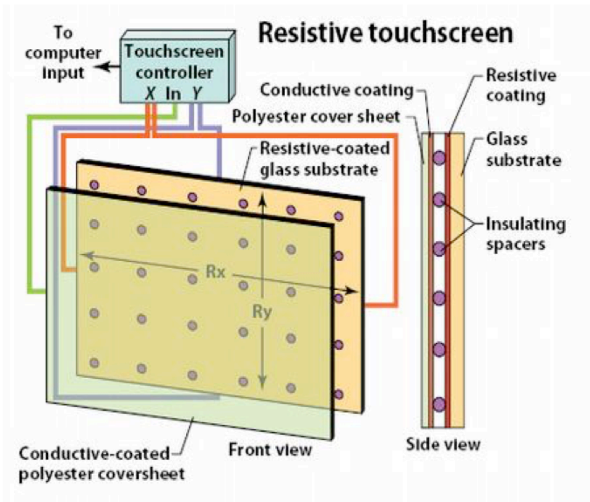
The piezoelectric element has a high output impedance output therefore the signal it produced must be carefully buffered (e.g. with a high impedance op-amp circuit.) Some piezoelectric pressure sensors include an internal amplifier to provide an easy electrical interface to other circuits.

Here is a video about an experiment that shows a home-made piezoelectric crystal produces an electrical voltage when squeezed:

https://www.youtube.com/watch?v=K3G2QM5a-9U

Touch Sensing - Resistive

- Composed of multiple layers separated by thin spaces
- Using indium tin oxide (ITO) layers – optically transparent, electrically conductive
- Contact made when pressed.
- Uniform voltage on first screen for X and second screen for Y
- Resistive screen works well with and without a stylus
- Low cost and rugged
- Generally cannot detect more than one touch point

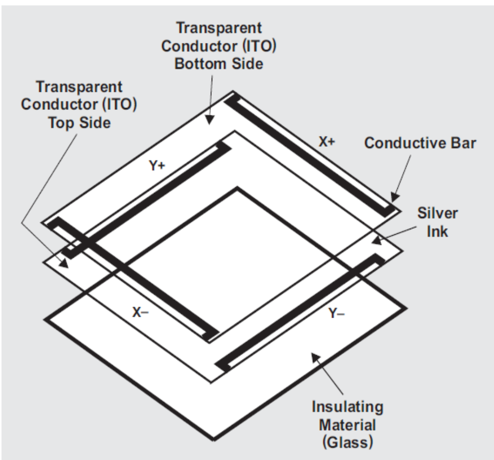


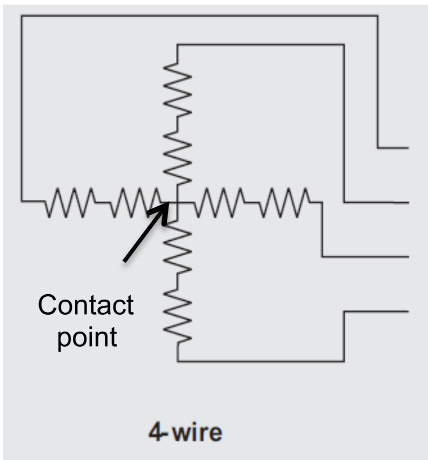
PYKC 11 June 2020 DE 1.3 - Electronics 1 Topic 16 Slide 24

Resistive touch screen has multiple layers of glass separated by a resistive and a conductive coating. Usually there are two separate layers for X and Y directions.

The ITO conductive coating is pretty transparent, but the resistive coating can block out some light. So resistive touch screen can reduce brightness of screen.

One common design for resistive touch screen is the 4-wire design as shown below:





When contact is made, a resistor divider circuit shown on the right is formed. The voltages at the terminals is dependent on the X, Y location of the touch.

Touch Sensing - Capacitive

- Becoming popular is capacitive touch sensing.
- Capacitor is formed with conductive coating (ITO) and insulator layer (glass or air).
- There are two types of capacitive touch sensors: surfaced and projected.

Surface type

- Only one side of the glass is coated
- Electrodes are at the edges
- Capacitor is formed ONLY after touch with finger – you are the earth terminal, completing the circuit
- Single touch only and limited resolution
- Contact location determine X, Y coordinate

Multi-touch Sensing – Projected Capacitive

- Has two parallel ITO layers and two sheets of glass
- Capacitor array distributed on the surface at many locations
- Touching changes the electrostatic field at the location of touch, changing many capacitances through field **projection** through glass
- Measure capacitance distribution, and can therefore work out **all** touch locations