Australian National University

I once thought I had

Physical Interfacing

for an entire year.

It turned out I was just really bored.

ENGN8537

Embedded Systems

Overview

- Interfacing Requirements
- Sampling and Domain Conversion
- Temperature Sensors
- Position Sensors

Interfacing Requirements

We have argued that Embedded Systems are physical instantiations of embodied phenomena. In short: Embedded Systems can't be defined outside of their interactions with their environment.

Given that, the Embedded System certainly must be able to sense environmental state around them.

Interfacing Requirements

Remember, the environment includes the user and other connected systems – one particular set of hardware may not have any explicit sensors*, but it still will sense and interact with its environment.

We assert that effectively interfacing with the user is very important, however we won't look at this in any depth. This lecture will focus on the physical aspects of interfacing.

* but it might have implicit sensors. More on this later

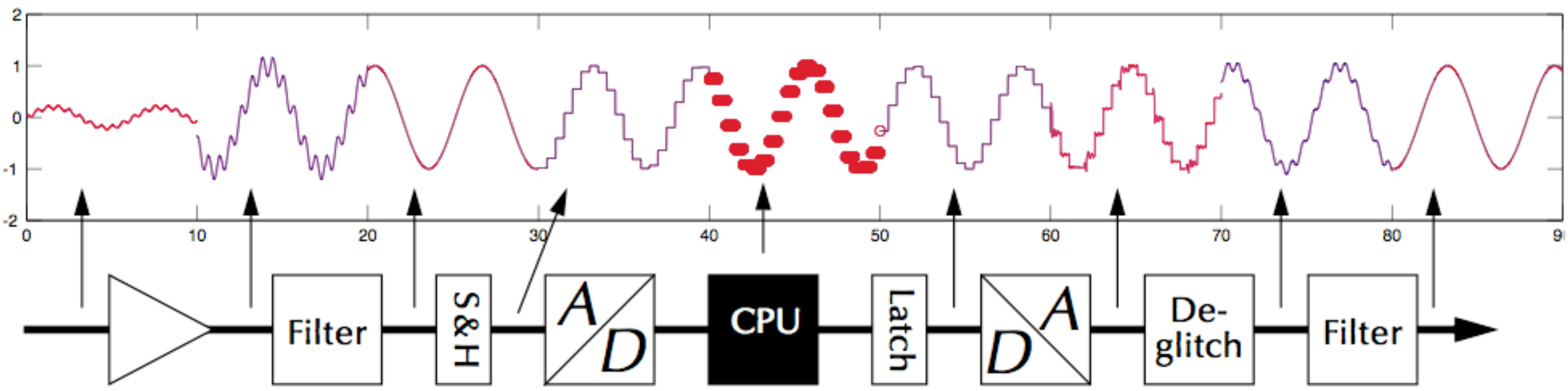
Interfacing Requirements

Fundamentally, processing is performed in a digital domain however the inputs are typically analogue.

- Temperature
- Distance
- Depth
- etc.

Sampling

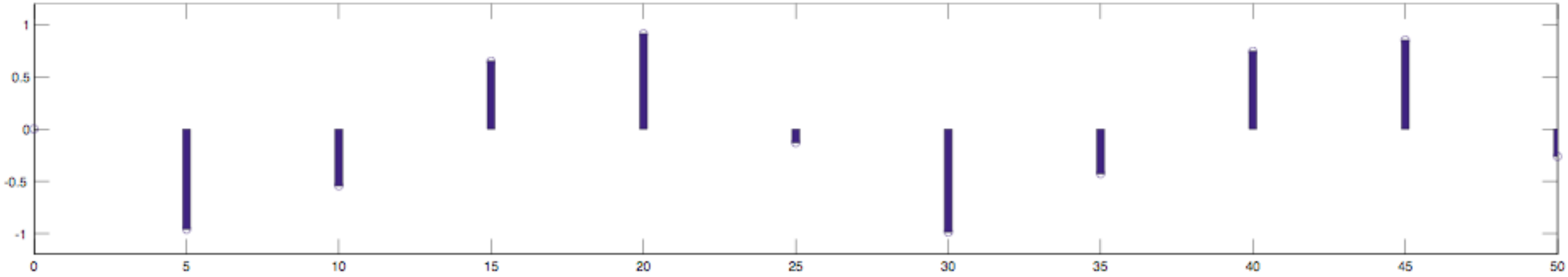
The analogue signal chain can be summarized as



Thanks again to Dr Uwe Zimmer!

Sampling

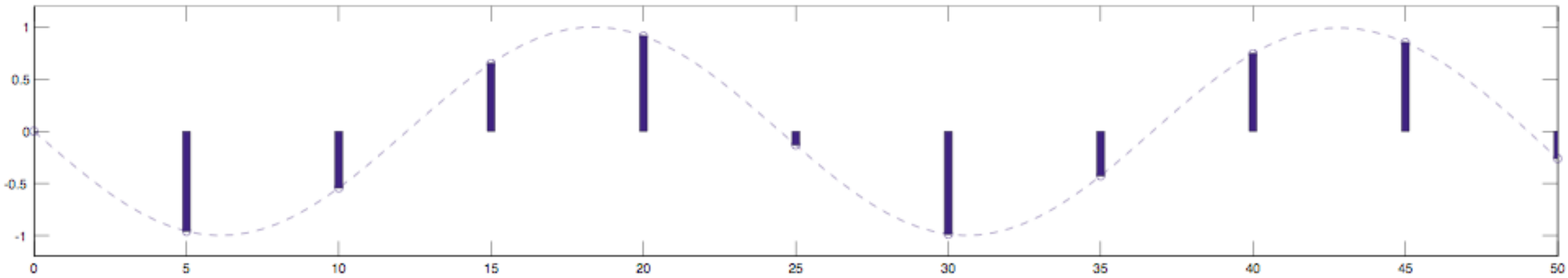
Sampling an analog signal at 5-second intervals



Thanks again to Dr Uwe Zimmer!

Sampling

Interpolation would suggest something like a sine wave

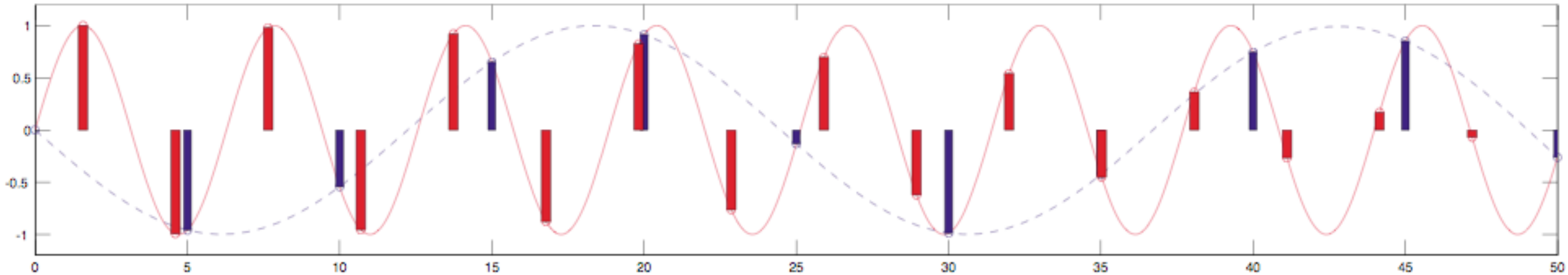


Thanks again to Dr Uwe Zimmer!

Sampling

Interpolation would suggest something like a sine wave.

But if we sample at a higher frequency, we might gain a different idea of the original signal



Thanks again to Dr Uwe Zimmer!

Sampling

This phenomenon is called ‘aliasing’. Sampling at a lower frequency has given us the wrong picture of the original signal

With sampling frequency F_s and signal Bandwidth B, Nyquist’s criterion states

$$f_s > 2B$$

Also sometimes called the Nyquist Limit. The sampling frequency must be at least twice the analog signal bandwidth in order that no information be lost

Sampling

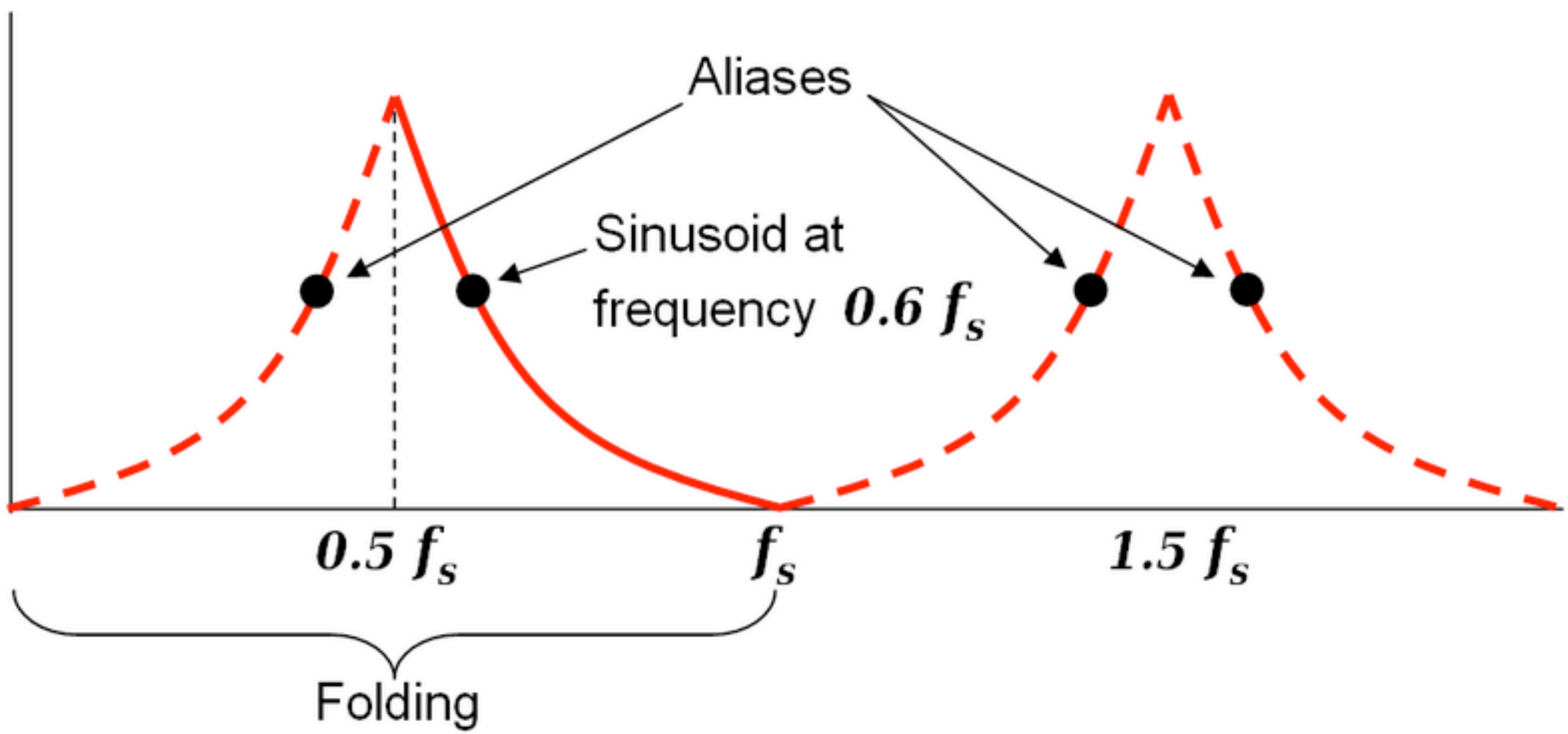
An extreme example of aliasing would be $f_s = B$.

That is, for example, 1 Hz Sine wave sampled once per second. In this case, the sample would occur on exactly the same point in each cycle, giving the impression of a flat line.

The 1Hz sine wave would have aliased to DC.

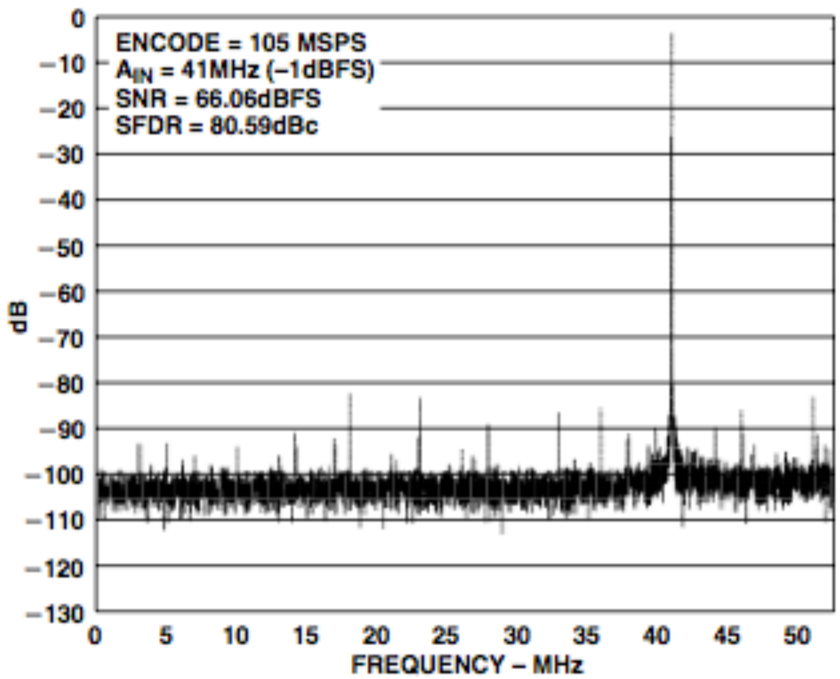
Sampling

In general, the aliased signals are folded around the sample frequency.



Sampling

An example from the Analog Devices AD10200 characterization:

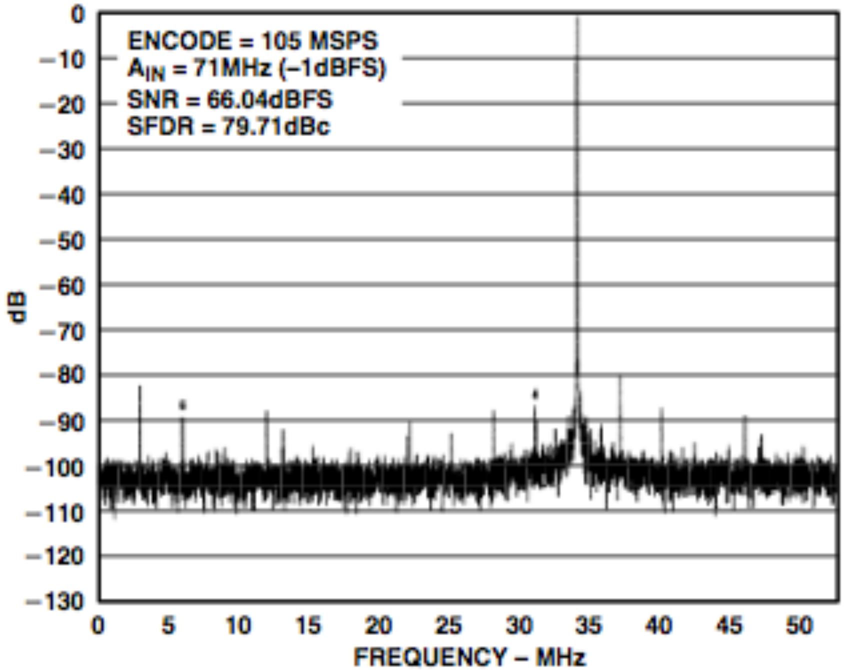


105MSPS
f_s / 2 = 52.5MHz

41MHz isn’t aliased

Sampling

An example from the Analog Devices AD10200 characterization:



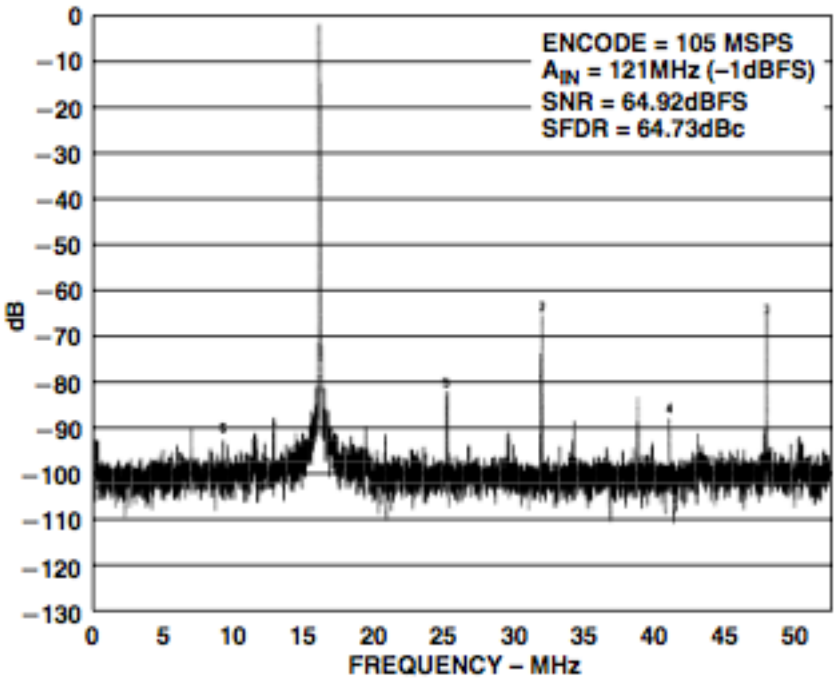
105MSPS
f_s / 2 = 52.5MHz

71MHz is aliased. The alias is folded around 52.5MHz

71 – 52.5 = 18.5
52.5 – 18.5 = 34MHz

Sampling

An example from the Analog Devices AD10200 characterization:



121MHz aliases to around 16MHz. Note though that in this case, higher order harmonics of the 16MHz aliased signal are within bandwidth. Note the extra peaks at 32MHz, 48MHz.

Sampling

A sampled signal is quantized.

Quantization is a noise source! It's a non-linear and signal dependent noise source. For a sine wave input, the signal to noise ratio due to quantization is given by:

$$SNR_{ideal} = 20N\log 2 + 10\log\frac{3}{2}$$

N is the number of bits

Sampling

If you put in the actual SNR and rearrange for N, we get the effective number of bits (ENOB).

$$ENOB = \frac{SNR_{actual} - 10\log\frac{3}{2}}{20\log 2}$$

Sampling

This represents the fact that a high resolution converter is only more useful than a low resolution one if the noise is sufficiently low that the extra bits contain actual information.

This may seem intuitive, however it's quite common for manufacturers to claim an ADC has performance figures like "4MSPS 12 bit" but if you read closely, not at the same time. For example, at 4MSPS the noise might be higher therefore the ENOB might be lower; the ENOB might only be 12 bits at lower frequencies.

Sampling

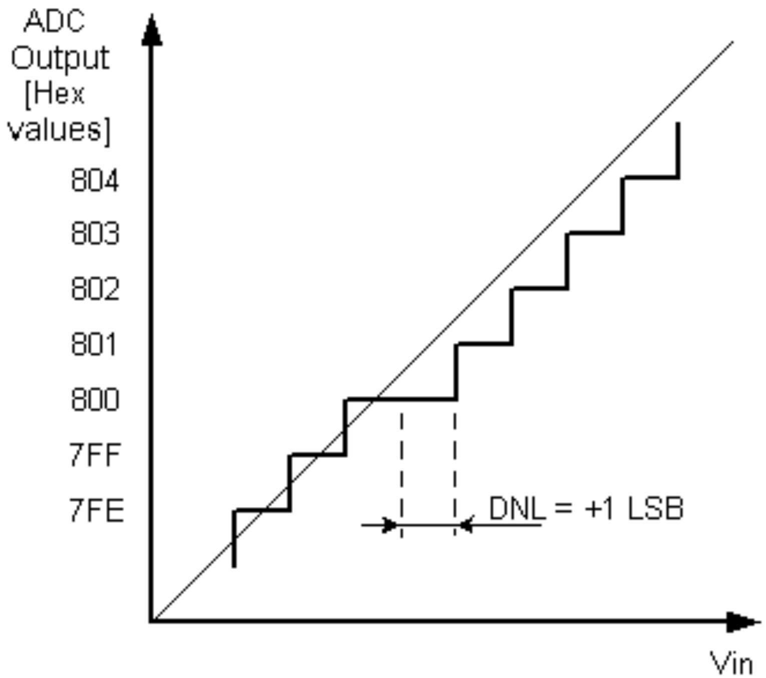
Ideally the digital output would vary linearly with the input voltage, however actual parts aren't ideal!

There are two typical types of non-linearity for converters:

- Differential Non-Linearity (DNL)
- Integral Non-Linearity (INL)

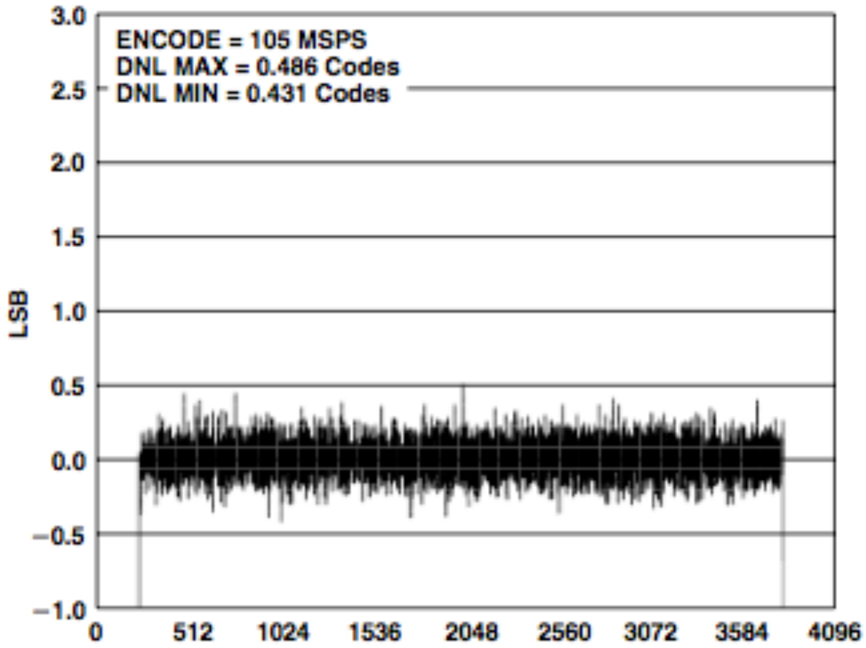
Sampling

DNL describes an error between successive codes. A DNL less than 1 means there are no missed or repeated codes.



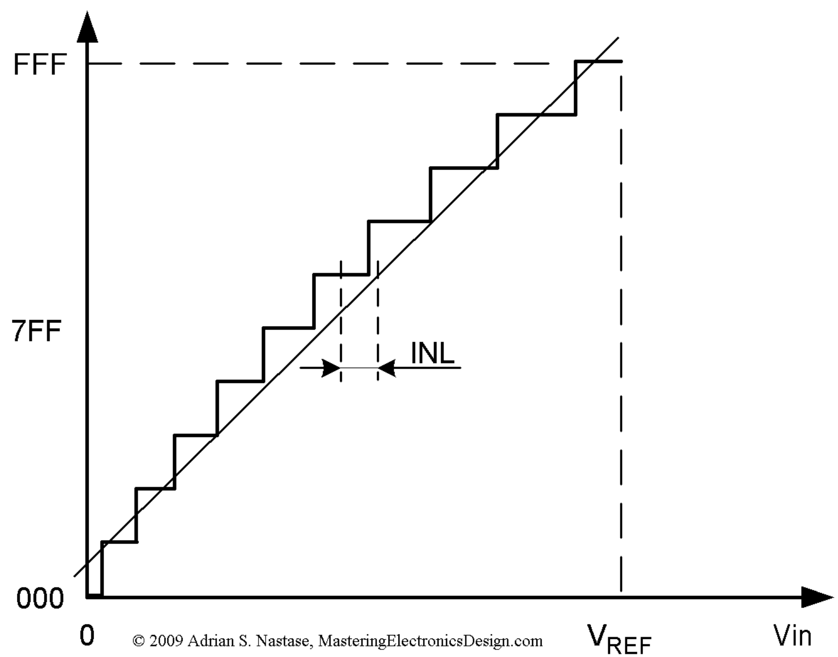
Sampling

A typical high-end ADC has a DNL less than about 0.5 meaning that all 'steps' in the previous diagram are within 50% of the same width as each other.



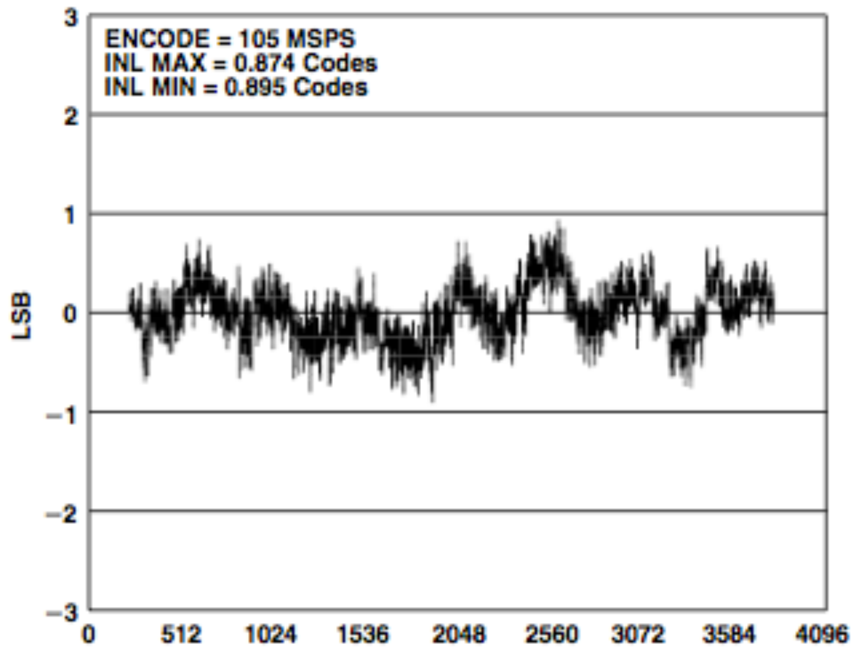
Sampling

INL describes the deviation between the middle of a code and the voltage that code represents



Sampling

An INL always less than 1 means that no input voltage in the middle of a code can trigger an adjacent code.



Converters

There are different types of ADC, each of which make a trade off between

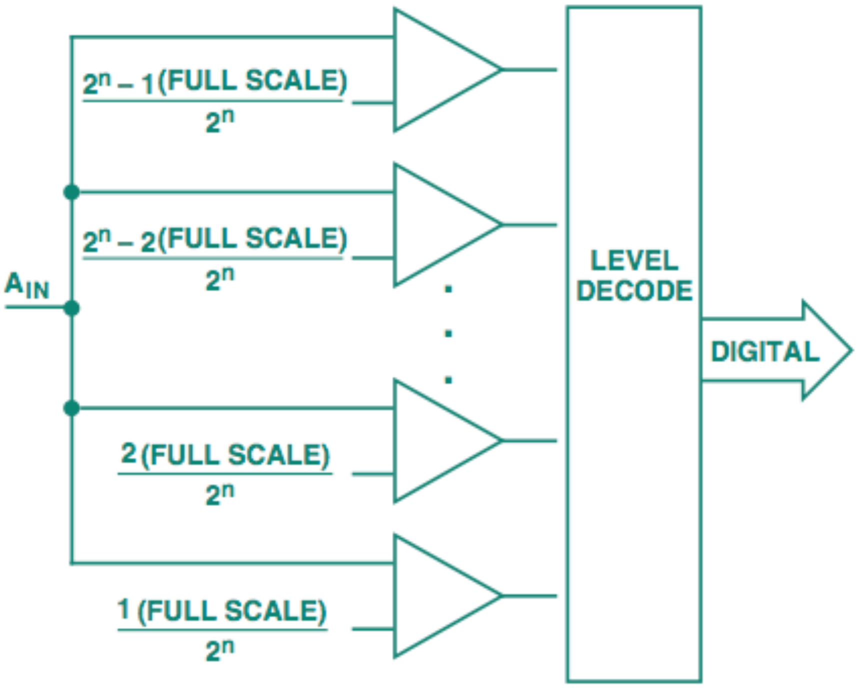
- Complexity/Cost
- Accuracy
- Throughput
- Latency

Converters

Flash Converter

Has 2^N-1 individual comparators connected directly to decode logic.

- Can convert in 1 clock cycle > fast
- a 16-bit converter would require 65535 individual comparators > expensive

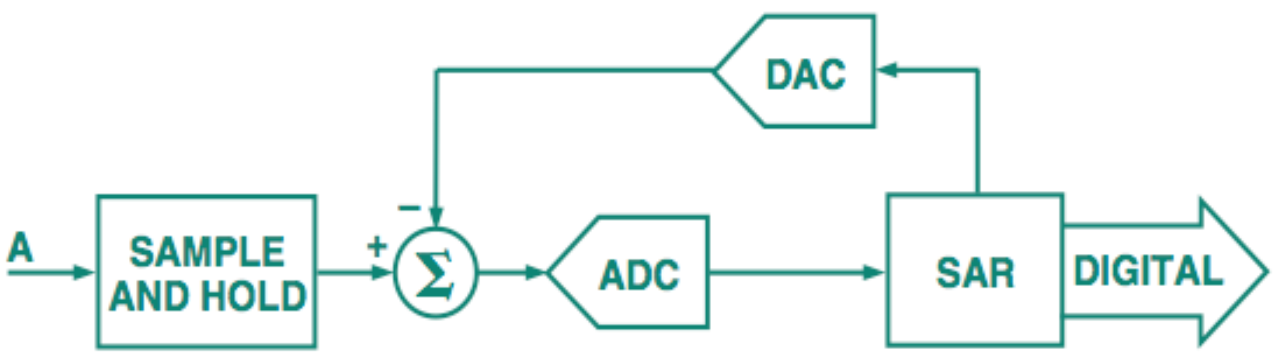


Converters

Successive Approximation

Has one comparator connected to an approximation of the input signal, derived from the previous stages

- Can convert in 1 clock cycle per bit > slow
- Same circuitry regardless of the number of bits > cheap

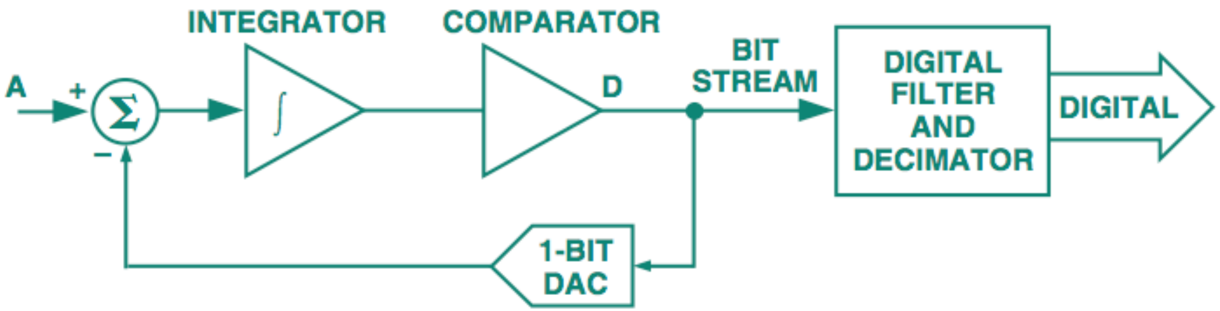


Converters

Sigma-Delta $\Sigma\Delta$

Has a 1-bit DAC whose output is based on whether the output of an integrator is greater or less than the input signal.

The density of the 1's in the bit stream represents the analogue signal



Physical Interfacing

Sigma-Delta ADCs have the apparent advantage that they don’t require a formal communications bus to interface to a microcontroller, the micro can dedicate a single input pin to the signal, sample it regularly and deduce the applied voltage by the ratio of high samples to low samples:

```
loop_for_1_sample {
    if (input_pin)
        hcount++;
    else
        lcount++;
}

v = hcount / (hcount + lcount)
```

Physical Interfacing

This method doesn’t work because of branch delays.

The following two snippets of code should produce the same result, but they don’t: The ‘if’ and ‘else’ branches don’t take the same amount of time to run!

```
loop_for_1_sample {
    if (input_pin)
        hcount++;
    else
        lcount++;
}

v = hcount / (hcount + lcount)
```

```
loop_for_1_sample {
    if (!input_pin)
        lcount++;
    else
        hcount++;
}

v = hcount / (hcount + lcount)
```

Temperature Sensors

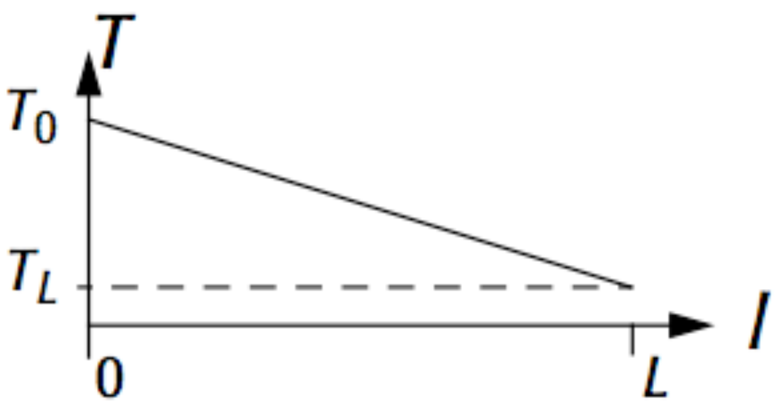
Every sensor is a temperature sensor, some are just better than others.

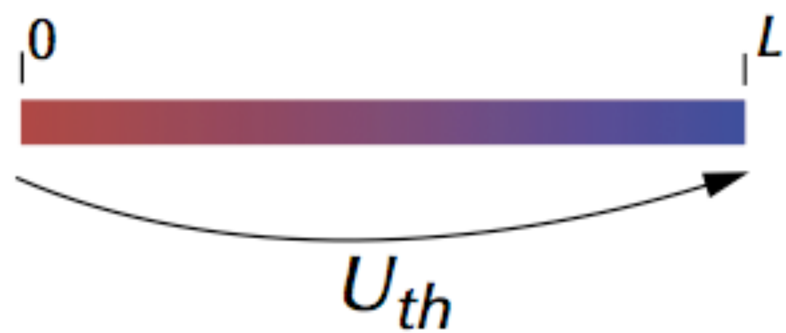
Almost all electronic elements are affected by temperature, sometimes this is a good thing, sometimes an effect that needs to be overcome.

Temperature Sensors

Thermocouples

Thermocouples are a type of temperature sensor based on the Seebeck Effect. Metals in a temperature gradient build up an electric field.



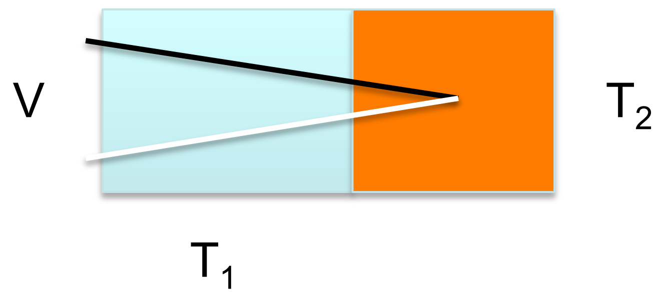


Temperature Sensors

Thermocouples

By using two metals with different Seebeck coefficients, a voltage difference can be measured. This voltage is proportional to the difference in temperatures and difference in Seebeck coefficients

$$V = (K_a - K_b)(T_1 - T_2)$$

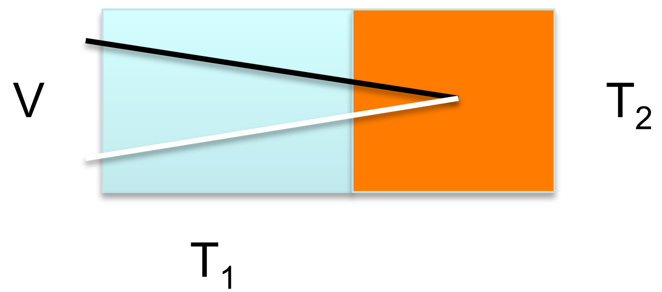


Temperature Sensors

Thermocouples

If T_1 and the Seebeck coefficients are all known, you can solve for the remote temperature.

$$V = (K_a - K_b)(T_1 - T_2)$$



Temperature Sensors

Thermocouples

There are many different types of thermocouples with different properties.

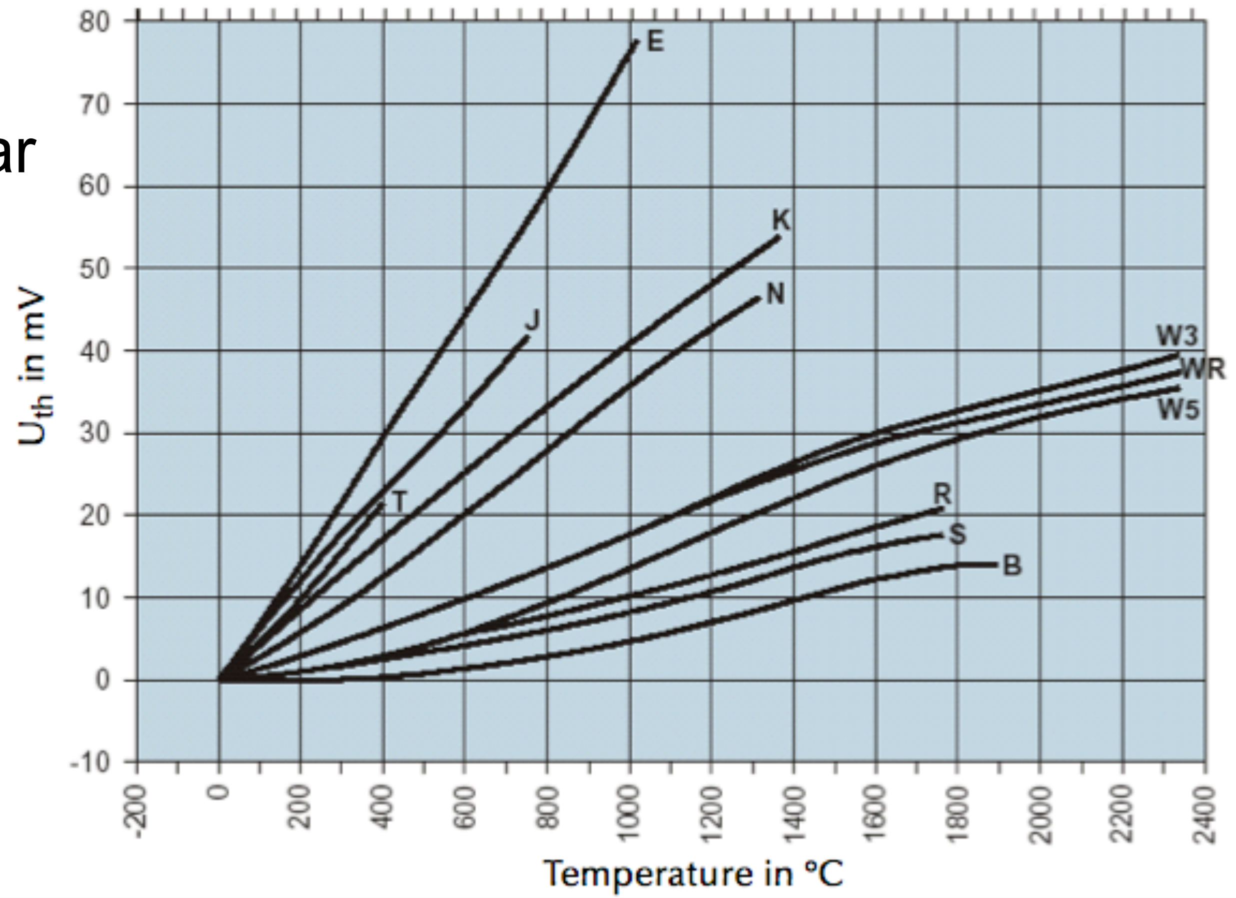
Good for high temperatures, but very small generated voltages

Short Name	Metals	T_{max}	Max V
T	Cu-Constantan	400°C	21.000 mV
J	Fe-Constantan	700°C	39.729 mV
K	NiCr-Ni	1000°C	41.310 mV
S	PtRh-Pt	1300°C	13.138 mV

Temperature Sensors

Thermocouples

Also, not too linear



Temperature Sensors

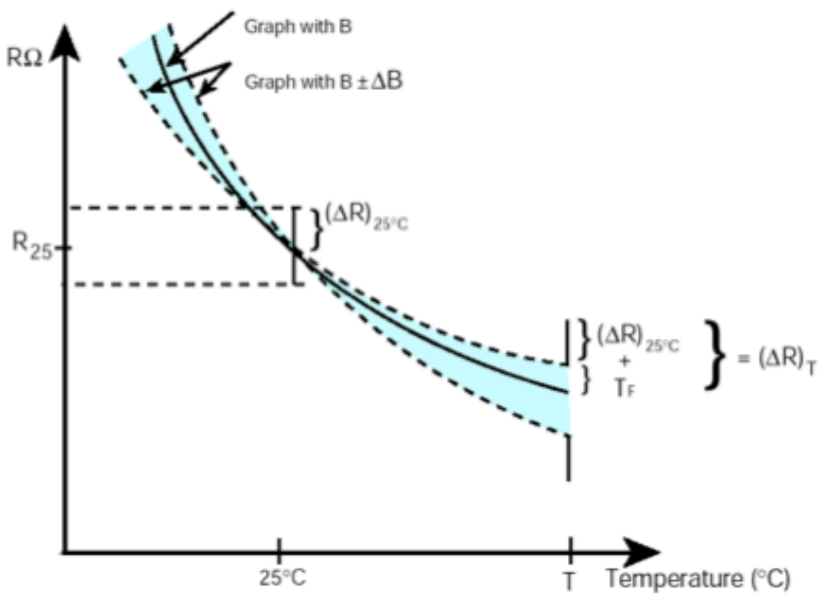
Thermocouples

- Cheap
- High Temperature
- Require precision sensing circuitry
- Need to know the temperature of one end already

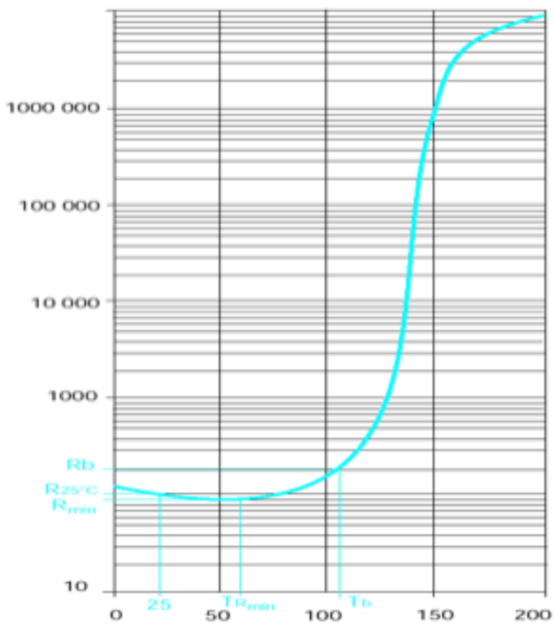
Temperature Sensors

Thermistors

NTC



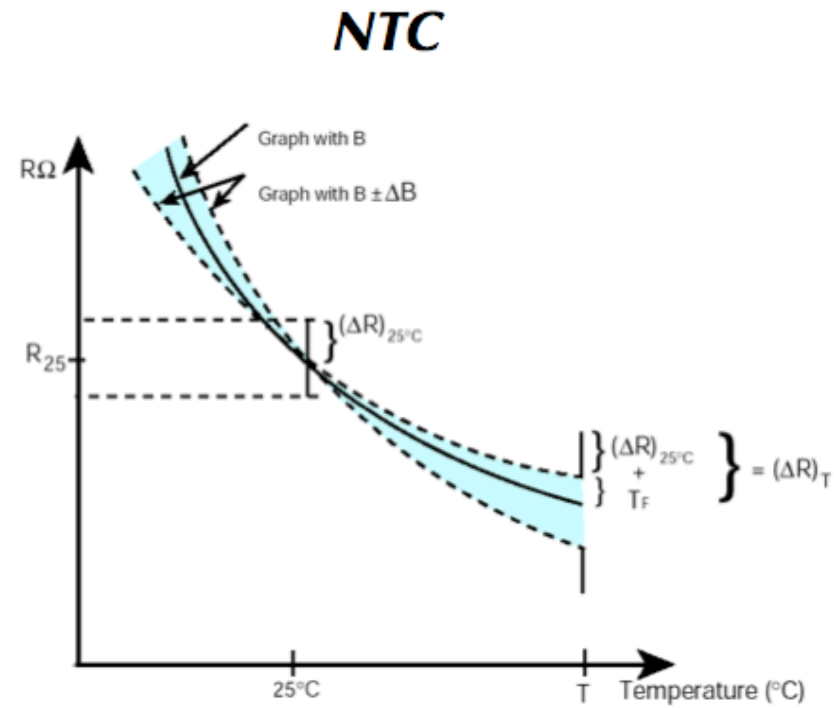
PTC



Source: AVX

Temperature Sensors

Thermistors



Negative Temperature Coefficient parts reduce resistance with temperature.

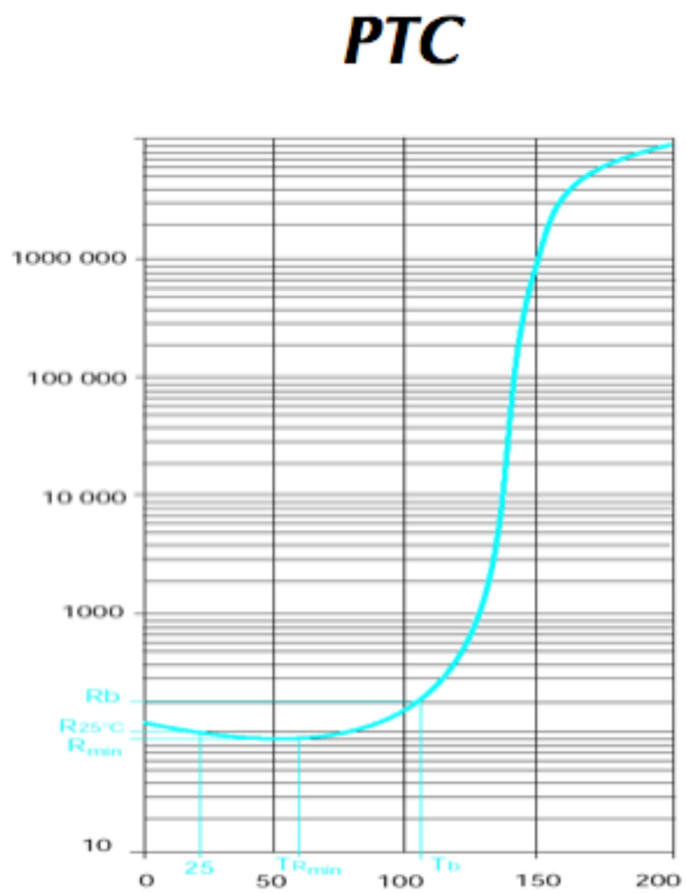
Work fairly well around room temperature but have increasing error outside that band

Temperature Sensors

Thermistors

Positive Temperature Coefficient parts increase resistance with temperature.

Very non-linear and don't work well around common temperatures



Temperature Sensors

Thermistors

Thermistors fundamentally provide some resistance as a function of temperature.

In order to measure the resistance, current has to be passed through. Current passing through a resistor dissipates power, which generates heat, which changes your reading.

This phenomenon is called ‘self heating’ and isn’t restricted to thermistors

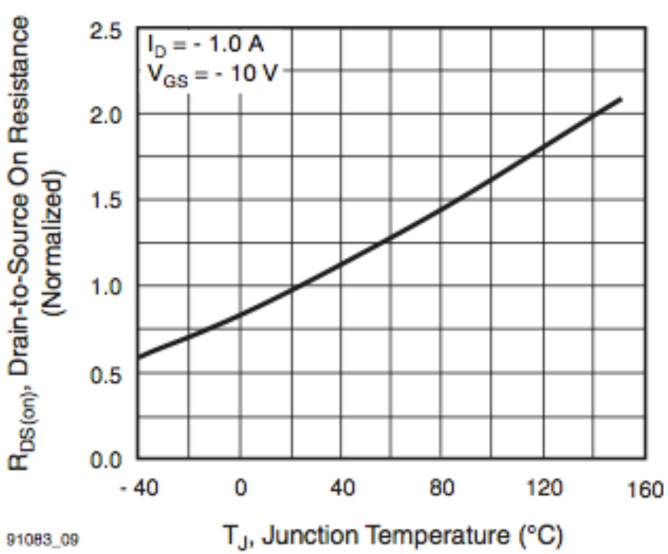
Temperature Sensors

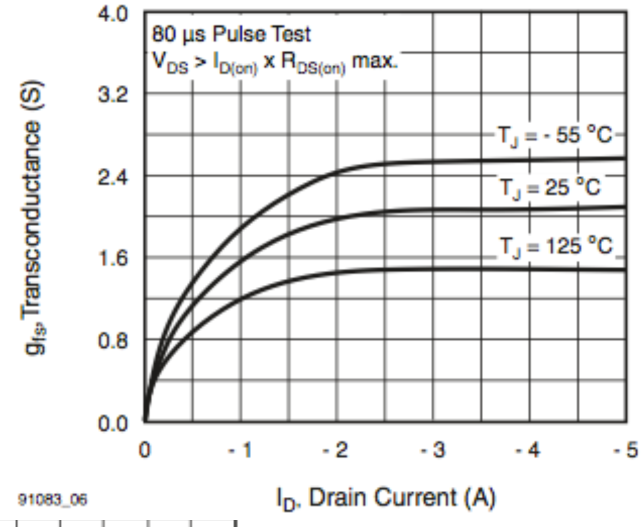
Semiconductor Sensors

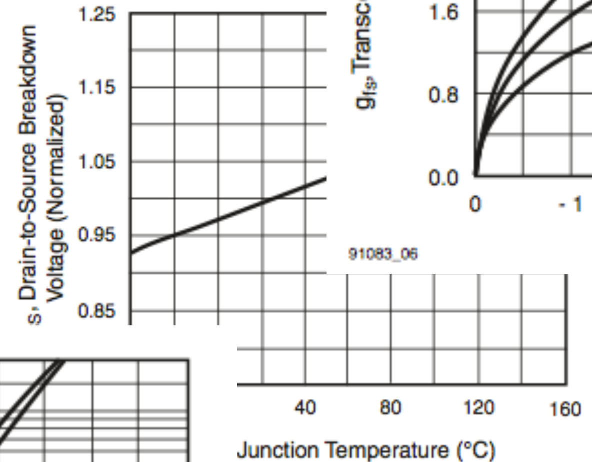
Different semiconductor structures exhibit either negative or positive temperature coefficient behaviour over a number of parameters. For example, the IRF9620S MOSFET from Vishay has at least five different parameters that change with temperature!

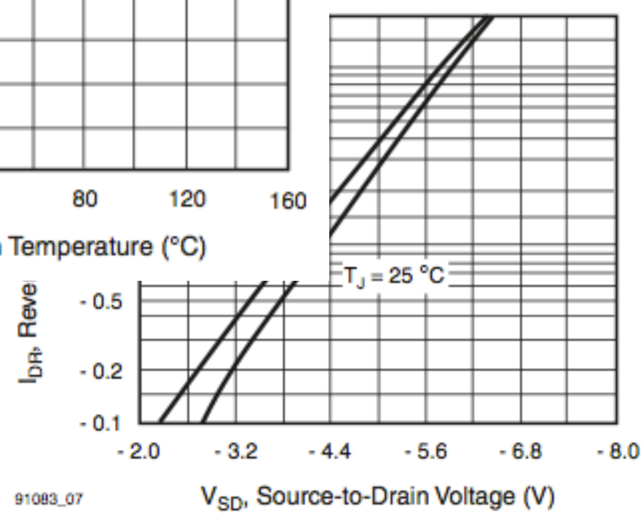
Temperature Sensors

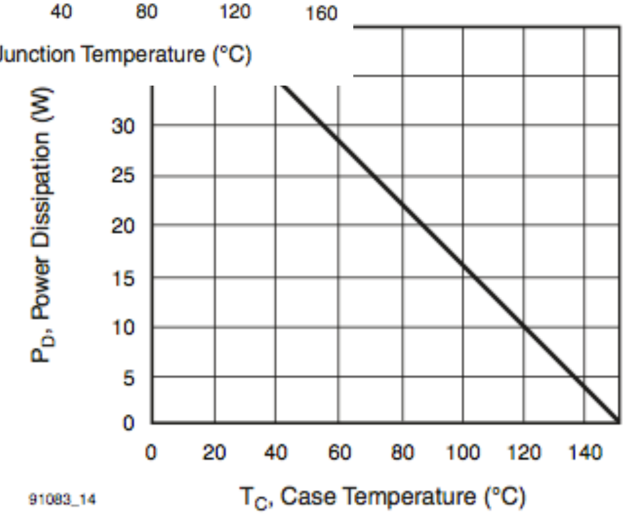
Semiconductor Sensors











Temperature Sensors

Semiconductor Sensors

Semiconductor sensors use these effects to provide a temperature reading. Other semiconductors actively try to minimise the effect of temperature on their operation.

Remember: Everything's a temperature sensor, some are just better than others!

Position Sensors

GPS is a very important sensing technology in modern technology.

GPS is the specific term for the American Global Navigation Satellite System (GNSS). Other countries are building their own GNSS so they don't have to rely on America.

Position Sensors

Global GNSS Systems:

- GPS (America)
- Galileo (Europe)
- GLONASS (Russia)
- COMPAS (China)

Position Sensors

GNSS systems are composed of three parts:

1. Satellites (Space Vehicles)
2. Receiver
3. Ground Station

Position Sensors

The satellites and base station contain high accuracy atomic clocks. Each time the satellites come in to range of the base station, their clocks are synchronized, keeping all satellites' idea of time within 1ns of each other. The satellites send a message whose content is related to the time.

For example, the GPS system might define that at exactly midnight the satellite sends an 'A'; then whenever a receiver gets that signal is knows that now, minus the travel time of the signal, is midnight. The actual signal is more complex than this, we'll come to that shortly.

Position Sensors

The satellites are in a ‘bi-sidereal’ orbit. This means that they orbit the earth exactly twice in a sidereal day (a day with respect to distant stars, not with respect to the sun).

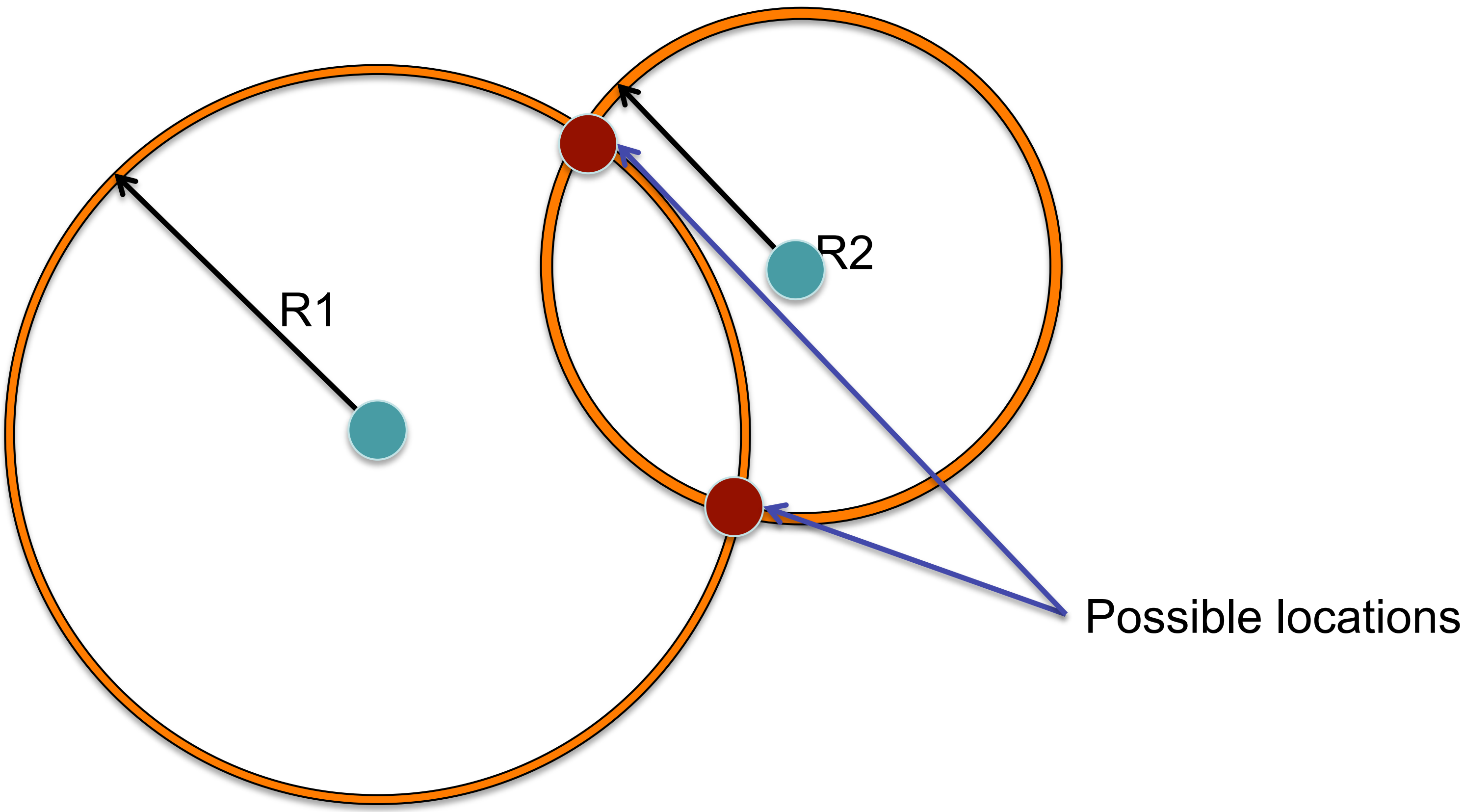
This in turn means that at exactly the same time each day, from the same place on earth, you will see the same positions of the satellites. This fact will be used later.

Position Sensors

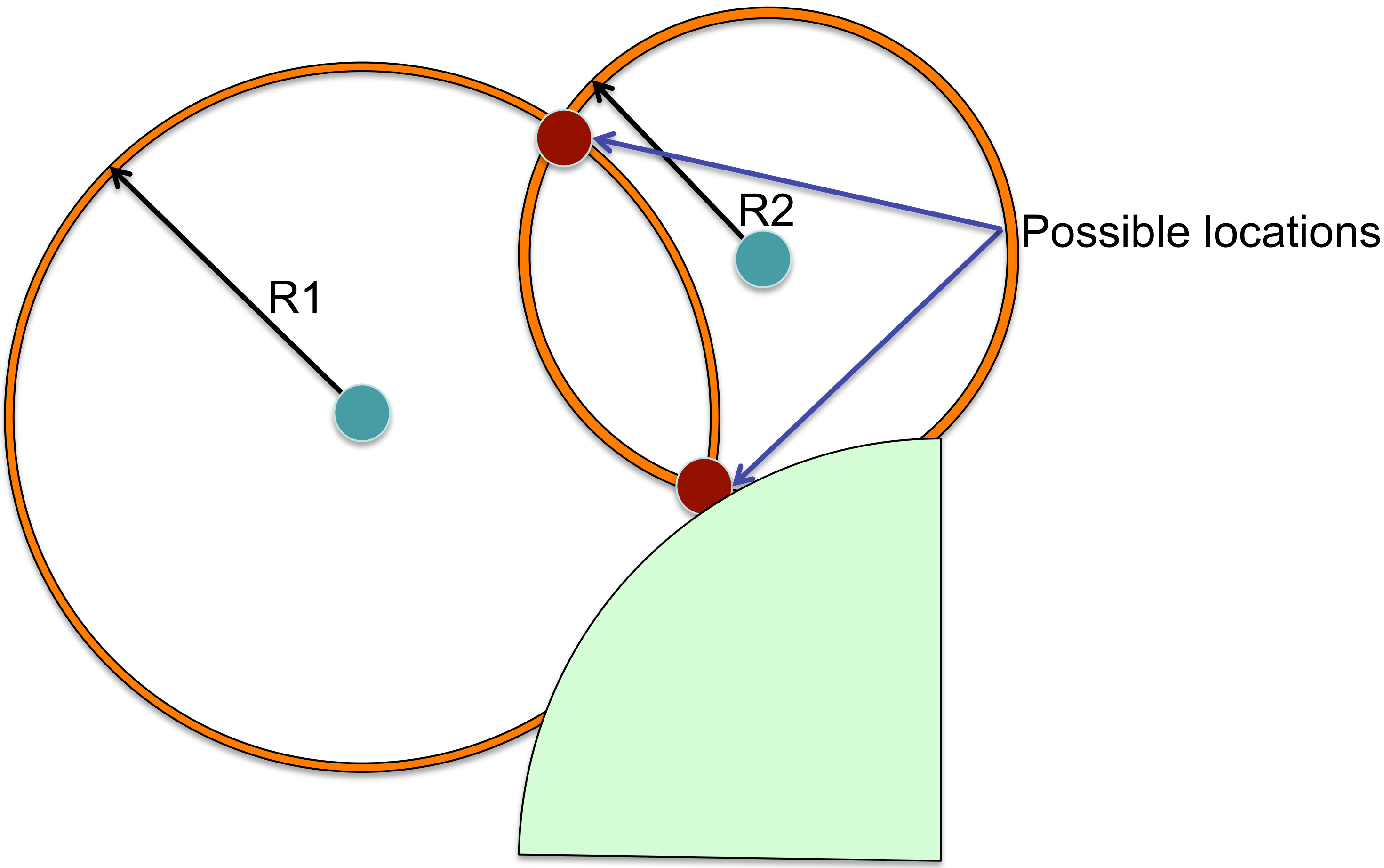
Receivers pick up the signals from each satellite. By comparing the known time of transmission of the signal, the receiver can calculate how long each signal took to arrive and therefore how far away each satellite is.

This is called trilateration

Position Sensors



Position Sensors



Position Sensors

GNSS systems require readings from four satellites to locate in four dimensions: X Y Z T (or Lat/Long/Alt/Time)

This number of readings will not solve the system to a single point (in general), but it will reduce to a small number of solutions. The correct one can be chosen based upon one extra piece of information: The receiver is generally pretty close to the Earth.

Position Sensors

The satellites transmit three different things in the same signal: Code, Carrier, Data.

The carrier wave is the underlying RF signal. There are actually three carrier frequencies in use, L1, L2 and L5. Most commercial receivers use only L1, which is approximately 1.575GHz.

Position Sensors

The satellites transmit three different things in the same signal: Code, Carrier, Data.

The code is a pseudo-random sequence modulated on to the carrier wave. The code signal won't repeat for millions of kilometers so determining the code phase gives the primary measure of distance to the satellite.

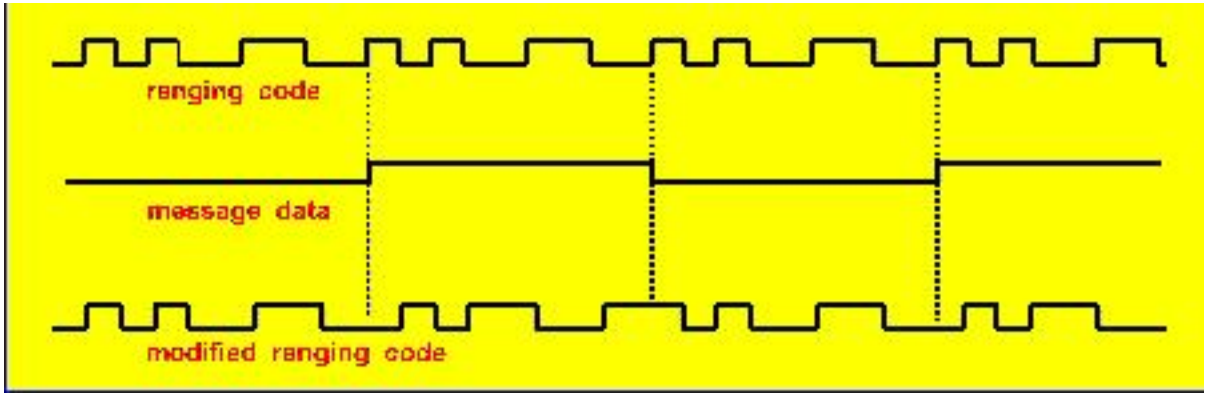
Position Sensors

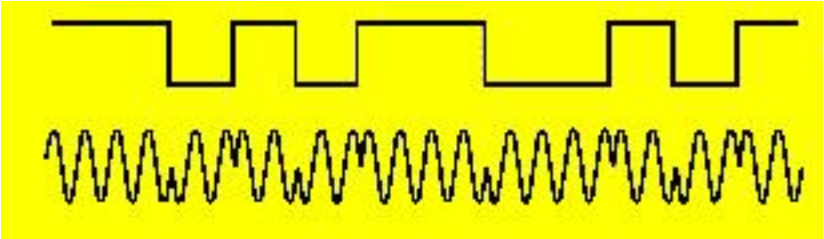
The satellites transmit three different things in the same signal: Code, Carrier, Data.

The data is modulated on to the code signal by flipping the code phase by 180 degrees. This is a low-rate data channel that allows the satellites to broadcast information to the receivers (apart from the normal ranging signal). This data typically contains information such as the satellite ephemeris data - where the satellites are expected to be in the future.

Position Sensors

The satellites transmit three different things in the same signal: Code, Carrier, Data.





Position Sensors

Sources of Error in GNSS

Clock Errors.

The 1 ns synchronization of the satellites equates to around 30cm of error in each reading.

General Relativity also makes a big difference. The satellites move in lower gravity than the receivers and they also move very fast. The combination of these two effects adds up to around 38 microseconds – or 11km - per day

Position Sensors

Sources of Error in GNSS

Clock Errors.

We have asserted that the time at the receiver is solved for along with position – why wouldn't the receiver just track time itself? Typical quartz clocks drift around 100ns – 3 metres - every second. The alternative is to fit atomic clocks to everyone's receiver but that's impractical.

Position Sensors

Sources of Error in GNSS

Satellite Orbit Errors

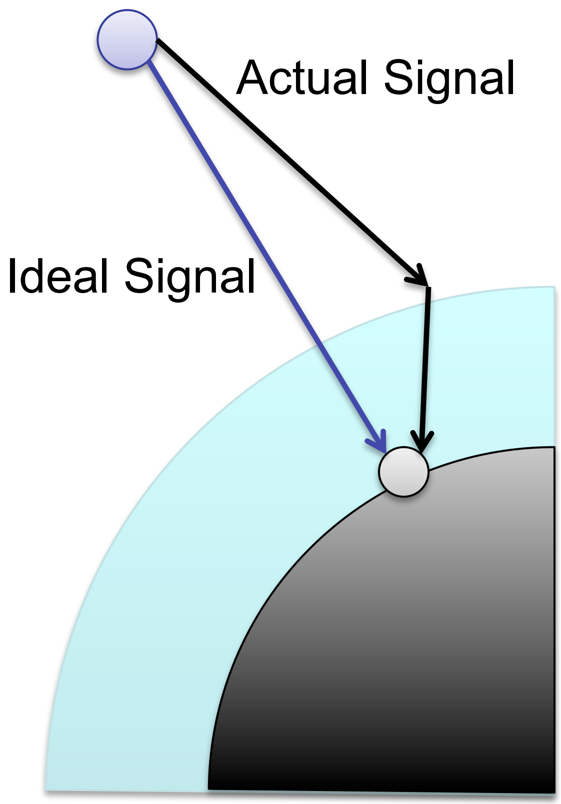
The solution of the satellite equations gives the distance to each satellite. In order to solve for actual position, the satellite locations must be known to great accuracy. In practice, the satellites are moving at around 14,000km/h but their position is still known within a few meters.

Position Sensors

Sources of Error in GNSS

Ionospheric Errors

These are the major sources of error in receivers. The signals from the satellites are distorted by their passage through the atmosphere. Ionospheric effects typically cause errors of 10 or more meters.

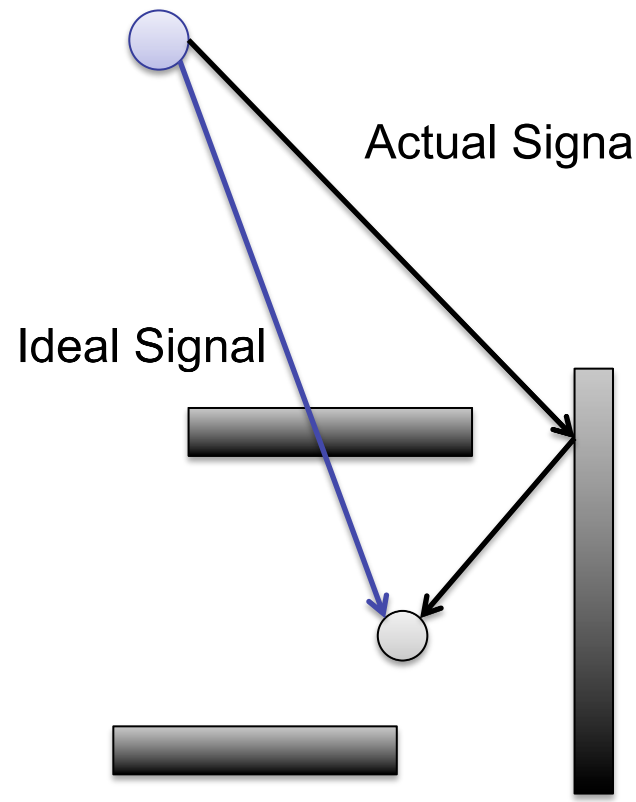


Position Sensors

Sources of Error in GNSS

Multipath Effects

The signal from the satellite may not reach the receiver by the shortest path. The extra length of the reflected signal will cause an error in the result.



Position Sensors

Solutions to GNSS Problems

The major sources of error in the GNSS system are common to all receivers in a small area. Multiple co-located receivers will all have the same errors due to ionospheric effects, satellite orbital errors, satellite clock errors and the like.

Position Sensors

Solutions to GNSS Problems

If one of those receivers has a precisely surveyed position, it can calculate the error due to these effects and communicate it to those receivers around it. The mobile receivers can subtract the known error and improve their position. This is known as Differential GPS (DGPS) and generally gives sub-meter accuracy depending on the distance between the surveyed and mobile receivers.

Position Sensors

Solutions to GNSS Problems

A simpler incarnation of DGPS is one where the correction signal is not relayed to the mobile receiver, but stored at the base. The paths generated from the mobile receiver can then be ‘fixed’ after the fact.

Position Sensors

Solutions to GNSS Problems

Going the other direction, these corrections can be transmitted back up to an augmentation satellite which in turns transmits them back to regular receivers through normal GPS signals.

Position Sensors

Solutions to GNSS Problems

The satellite-based augmentation system (SBAS) is not available all over the world - Australia isn't covered. Even in the areas that are covered, the SBAS may degrade signal quality as there aren't any guarantees as to how far away from the receiver the error signals were calculated.

Position Sensors

Solutions to GNSS Problems

Once a DGPS system has isolated the location to within a meter, the system location may be well enough known that the carrier phase information can be used to refine position further.

This is known as Real-Time Kinematic correction (RTK).

Position Sensors

Solutions to GNSS Problems

Recall, GPS actually works on several frequencies: L1, L2 and L5. Receivers that have access to more than one frequency can get much more accurate readings without DGPS or RTK.

The single largest cause of errors are the ionospheric distortions. These distortions influence different frequencies in different ways, so by receiving the same information on two bands, the dual frequency receivers can largely remove the distortions themselves.

Position Sensors

Other uses of GNSS

The main reason that a GPS unit takes a while to get a 'fix' is that the receiver must download all the ephemeris data through the satellite data channel at 50 bytes per second. This may take several minutes.

AGPS, assisted GPS, uses an external data channel (though, for example, 3G) to download this ephemeris data and skip straight to the calculation of distances.

Position Sensors

Other uses of GNSS

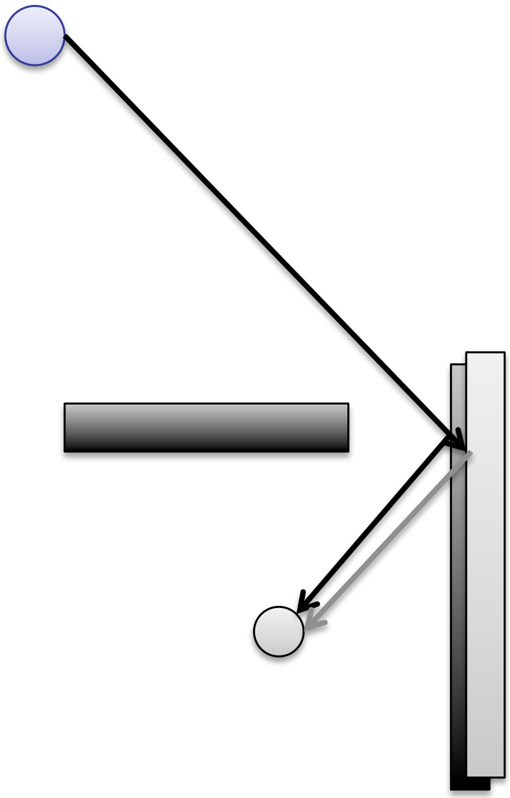
The last several slides worth of discussion have centered around using GNSS to calculate position.

In fact, GNSS is much better at giving velocities than positions. Positions require distances, ephemeris and so on. Velocities can be determined directly by monitoring the Doppler shift of the carrier signal from the satellites.

Position Sensors

Other uses of GNSS

Multipath signals can actually be very useful. At the same time each day, a particular satellite will come across exactly the same patch of sky. By monitoring the time of arrival of the multipath reflection each day, the position of the wall causing the reflection can be monitored.



Position Sensors

Other uses of GNSS

This method can be used to monitor the deformation of dam walls to to millimeter accuracy