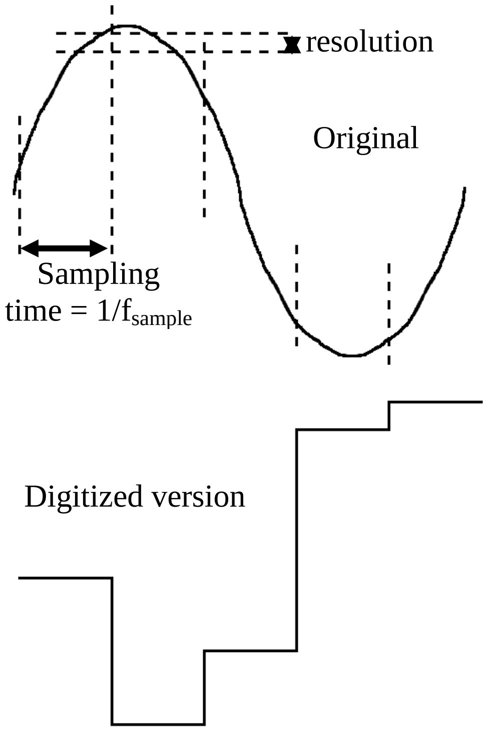
Analogue Digital Conversion and Measurement: Conversion & noise

The lecture course began with digital electronics and logic gates. We then went on to analogue signals. Here we are going to go close the circle, by discussing *how to convert analogue signals into digital data* for manipulation on the computer. We then will see how noise will influence these measurements and how you can try to reduce it.

(1) Sampling & Frequency:

The process of **digitizing** an analogue signal involves taking a signal that is a *continuous variation* in voltage with time and then deciding how to *approximate it by discrete voltages at discrete time steps*. **Information is lost** converting from analogue to digital. The number of subdivisions in the signal size (of a voltage signal on the vertical axis in the plot below) is the **resolution**. The number of subdivisions in the time (on the horizontal axis in the plot below) is the **sampling rate**. Note that this nomenclature has a *single signal parameter* in mind *which varies with time*, like the one shown in the upper part of the plot.



There is another type of resolution: in the CCD chip of a camera the sampling of the signal is not done in time but in the plane of the image which is *subdivided into pixels*. This is called the **spatial resolution**. The signal size is given by the charge collected during the exposure in each pixel. The precision of the brightness measurement in each pixel is defined is defined by the *resolution* of the charge-to-digital converter of the CCD chip.

In high demanding applications you may have all three ways of digitisation present. Take the LHCb Ring Imaging Cherenkov (RICH) detectors as an example. As most detector components of experiments at the Large Hadron Collider (LHC) it takes data samples at the rate of bunch crossings at the LHC, with a rate of 40MHz, i.e. it takes one sample every 25ns. The RICH detectors cover an area of about $3m^2$ of active area with a pixels size (spatial resolution) of $2.5x2.5mm^2$, that is about 500,000 pixels which are read out in parallel (currently at a reduced rate of 1MHz, but the upgrade foresees a readout at 40MHz). Each pixel has to be sensitive to single photons in the near UV and visible spectrum. A novel photon detection technology allows creating charge pulses with a charge equivalent of up to 5000 electrons (=0.8 fQ) from single photons in an electronic environment which features in its front-end an absolute peak noise level of typically less than 1100 electrons with a Gaussian variance of the noise of typically 150 electrons. By setting a threshold (in steps of approximately 25 electrons) as low as possible but still high enough to prevent any noise to pass it, any charge pulse larger than the threshold will be interpreted as a photon hit. Here the signal resolution is defined by the step size with which the threshold can be set, while the readout is binary (a pixel has seen a hit or no-hit) which minimises the data volume to be shipped for each event (occurring at each LHC bunch crossing).

Amplitude resolution:

Let's return to the simple time-dependent signal of the plot above. The shown amplitude (voltage) resolution of the digitised readout may or may not be sufficient for the application at hand. You would need to answer the following questions to find this out:

- What is the *largest and smallest signal size* you would like to be able to measure?

This defines the **dynamic range** your system has to cover.

- What is the *noise level* in your system? I.e. what are the smallest changes of your signal you still can resolve?

This defines the smallest sensible **step size** of your digitisation. If you make your step size *larger* than the noise fluctuations you *lose information* in the digitisation process. If you make your step size *smaller* than the noise fluctuations *you do not gain any further information* (you only increase uselessly the data volume...). So if you find that your system cannot resolve your signals or its substructures because they are smaller than the noise level you first have to work (hard) to *reduce the noise level in your system* before it would make sense to increase the resolution of the digitisation step.

Whether you can afford the optimal resolution (step size) in your digitisation process is another question. **Analog-to-Digital Converters** (**ADC**s), whether they convert voltage levels or (much less common) integrate over charge pulses, generally provide a numbers of digitisation levels, called **bins**, which come in powers of 2. In general n-bits code 2^n ADC bins. This commonly is referred to as an **n-bit** ADC.

This nomenclature is simply *caused by the data structure* the converted value is stored in: an 8-bit ADC, for example, stores the converted measurement in an *integer number represented by 8 bits* and therefore can represent 256 different ranges of analogue values (step sizes) by the numbers 0 to 255. The measurement results typically are displayed in a **histogram** displaying the number of *entries per histogram bin* across the *range of ADC bins*. If the dynamic range of an 8-bit ADC is 0...+5V than each bin represents a voltage range of 5V/256bins = 19.53mV/bin. So the size of your **Least Significant Bit** (**LSB**) is about 20mV and your typical **measurement error** due to the digitisation is **½ LSB**, i.e. ~10mV in this example. On top of that you will have the error from the noise in your system. So in applications you may find:

	n-bit ADC	number of ADC bins	ADC bin range	comment
2^1	1-bit ADC	2	0:1	**binary readout**, e.g. LHCb RICH
2^4	4-bit ADC	16	0:15	e.g. fast, data volume limited applications
2^8	8-bit ADC	256	0:255	e.g. for cost-effective, low spec applications
2^{10}	10-bit ADC	1024	0:1023	common research ADC
2^{11}	11-bit ADC	2048	0:2047	high grade research ADC
2^{12}	12-bit ADC	4096	0:4095	spectroscopy grade ADC
2^{13}	13-bit ADC	8192	0:8191	high spectroscopy grade ADC

The cost of ADCs goes fairly linear with the number of bins provided. For high precision research grade ADCs from low-volume production you can almost equal the number of ADC bins with GBP, for a single readout channel. The *high precision and linearity* also comes at the cost of a *slow conversion time* of the order of 10μs. Low and medium grade ADCs produced in high volumes for the consumer market are orders of magnitudes cheaper.

Now look back at the plot above with the example sampling of a time-variant single signal. You already may have noticed that the digitised wave form below the input signal is *inverted* with respect to the input. So an inverting ADC was employed. That is no problem as *for digital signals transformation and formatting operations are performed without further loss of information*.

Sampling rate / time resolution:

And once more we look at the example plot above but now we look at the **sampling in time**. You will notice that the sampling steps in time *look rather coarse*. Very significant approximations have been made going from the analogue to digital versions. We will have to examine how coarse the sampling in time may be in order to still get a *sensible representation* of the original signal, i.e. one from which the *main characteristics* still can be determined *without ambiguities* and with *limited errors*.

As for the vertical resolution you have to answer some questions to specify the parameters of your system. You need to know:

- What is the *fastest change* in your signal that you still want to be able to resolve?

The required **time resolution** will determine the **necessary sampling rate**.

- *How many successive samples* shall the system be able to store at the sampling rate before it (in many cases) is read out at a slower speed?

That will define how much *fast analogues and/or digital memory* (also referred to as **buffer** memory) you will need to record your *longest* and most *complex signal*. And it will be an important input for determining the **dead time** of your system, i.e. the fraction of time your system is *unable to take data* because it is busy.

To shed more light on the question of the *necessary sampling rate* for a measurement, let's have a look at an admittedly unrealistic, yet more familiar example: suppose you know someone with regular mood swings, in the morning this night-owl is grumpy but in the evening fit and in a good mood. You have to ask here the same questions as for any electronic signal:

1) What is the period of the signal (the mood swings)?
2) How often do I have to sample (meet the person) to get a reasonable idea about the character of the signal (the behaviour of the person)?

As for the period, in the example it is a day. So if you only meet this person in the mornings you get an incomplete impression of that person's behaviour. You need at least two regular visits to get the complete picture. Even then the timing of your visits, by chance, may fall in the times of the changes of mood and you don't get the right idea how severe the mood swings are, i.e. about the amplitude of the signal.

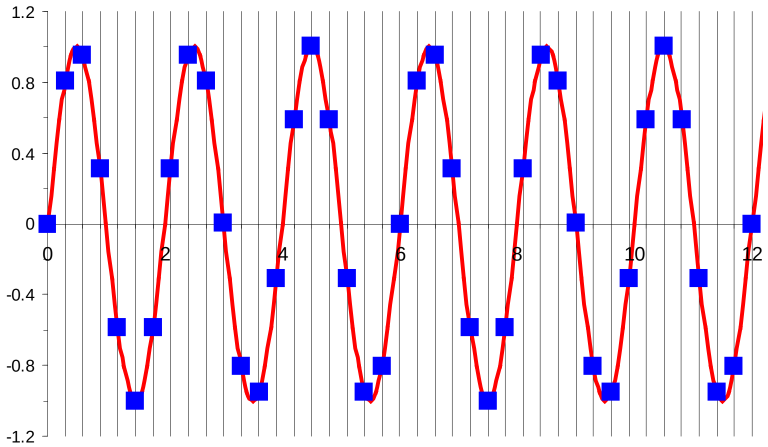
In summary: **two samples per period** are the *very minimum* you need to get correct characteristics of a *periodic signal*. To avoid the pitfall of getting the amplitude measurement wrong because you sample at a *fixed phase shift* with respect to the periodic signal you want a sampling rate which is getting *more than two samples per period*. This is known as the **Nyquist**

criterion, which states that the *sampling frequency* must be *at least twice the highest frequency in the signal*:

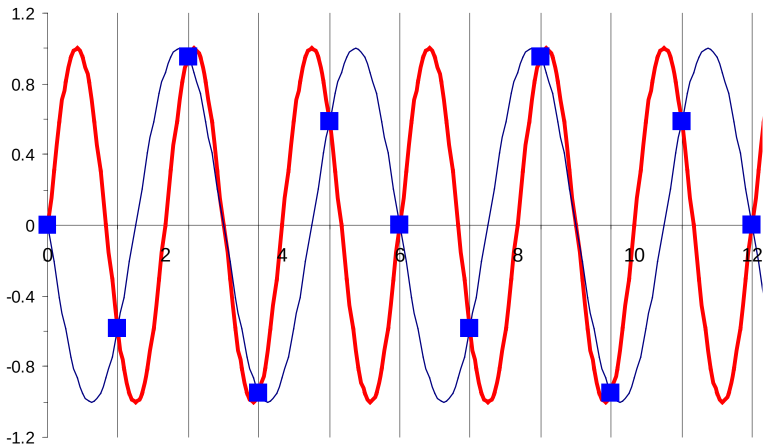
$$f_{sample} \geq 2f_{max}$$

If this condition is not fulfilled then a problem occurs which is known as **aliasing**. Here the sampled components of a signal with a frequency higher than the Nyquist limit ($f > ½ f_{sample}$) fake a signal of lower frequency. This is illustrated below.

More than 2 samples / period: gives good signal reconstruction



Under-sampling gives aliasing: poor signal reconstruction



A fake signal emerges at a lower frequency

The red sine wave is the original high frequency signal. When samples at high enough rate, like in the upper plot with approximately 7.5 samples per period, one can *reconstruct* from the samples well the *frequency* and *amplitude* of the signal. In the lower plot the same signal is only sampled with about 1.5 samples per period. As you sample at different phase positions of the signal you still get a *fair result on the amplitude* of the signal. But the *frequency* you get *significantly wrong*. If you fit the poorly sampled signal with the *hypothesis* of *having measured a sine wave* you get the *best fit* with a sine wave of a *significantly lower frequency* (blue sine wave). The *fake lower frequency signal* is the **alias** of the true higher frequency signal which was under-sampled.

In general the nature of a signal is *unknown*. In order to *control aliasing* the signal first has to be filtered by a low-pass which fulfils the condition: $f_{3dB} = f_{max}$, i.e. has to match the Nyquist criterion defined by the sampling capability (f_{sample}) of your system. Employing filters in this way is common. They are known as **anti-aliasing filters**.

The above discussion was done using sine-wave signals. As any signals can be build by superposition of sine-waves in principle the problem is solved for any signal. But there are two further points worth to be spelled out:

- If you have a square wave as input, for example, the frequency spectrum of a perfect square goes to infinite frequencies. *Any physical device producing or receiving this wave will act as a low-pass with some cut-off frequency*. This limit is called the **analogue bandwidth** of the device. The effect is that the square nature of the wave is reduced as the *edges of the square are rounded off*, the stronger the lower the cut-off frequency is. When a signal is composed from many sine-waves it may be that *part of the frequency spectrum of that signal is above the Nyquist limit* of your digitisation equipment. *Using an anti-aliasing filter will distort the signal* in the same way the analogue bandwidth limit. Typically it reduced the cut-off further with respect to the other components in the data acquisition chain.
- The discussion was done so far for *periodic signals*. If you sample such signals with a frequency near the Nyquist limit, i.e. at slightly above two samples per period, you may start out to sample at an unfortunate phase point of the signal and may first estimate the amplitude significantly too small. But as the phase shift of the samples with respect to the signal builds up over a number of periods you eventually get samples at the full amplitude of the signals as well, i.e. *the closer you sample to the Nyquist limit the more periods you need to fully reconstruct your periodic signal*.
  Your demands for the sampling will rise when you work with signal **pulses**. Then you do not have several periods available to reconstruct your signal, but you have to *get it right the first time*. The rule of thumb is that you would like to get *three samples on the fastest edge of your signal*, one at the base, one in the middle and one at the top, or any phase shift with respect to that. This is nothing else than demanding a sampling frequency which is four times higher than the Nyquist limit. Looking more closely it turns out that a sampling rate *three times higher than the Nyquist limit* already is sufficient.
  The case of the finer sampling in the above example fulfils this requirement. Say that your signal just consists of *one half-wave* of the red sine wave. At whichever phase shift of your samples with respect to the half-waves you get a good idea about the shape and magnitude of each individual half-wave.

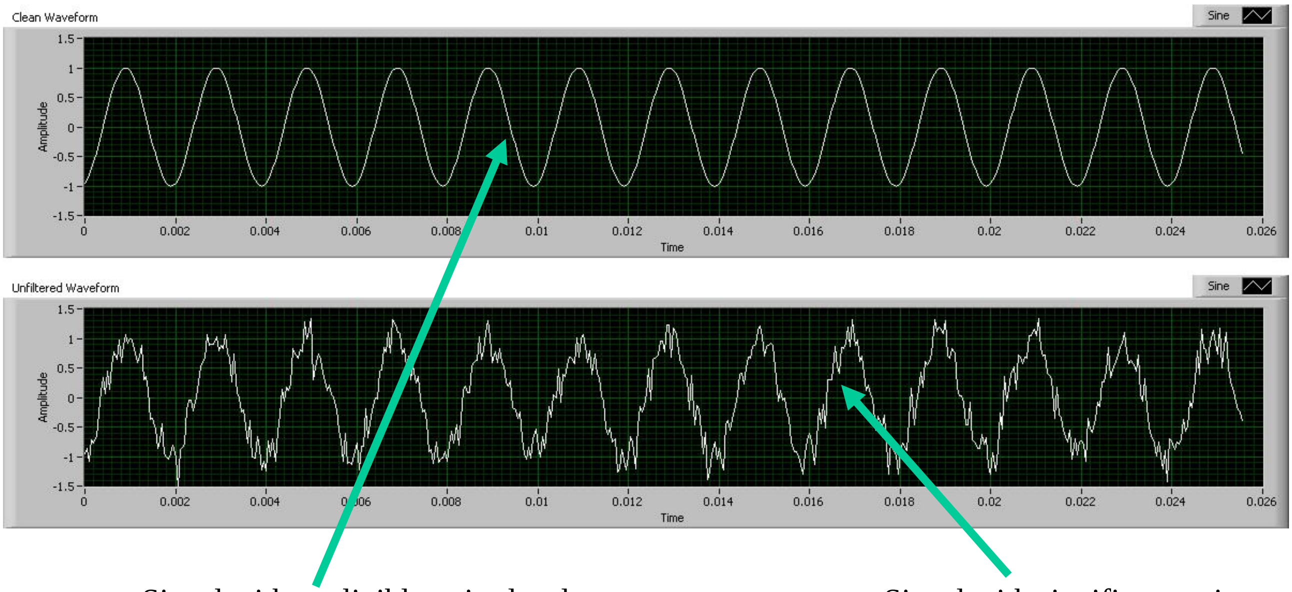
The cost of digitisation rises strongly with the sampling rate, I would say at least quadratic, at the leading edge of the technology even more. This is because you have to pay for the research and development going into the technologies which facilitate the extension of the available upper limits. The rise of cost with the sampling rate is only comparable with the rise of cost in the analogue bandwidth which increases as strongly.

The same Nyquist criterion and aliasing problems also arise with *spatial frequencies*. Better CCD chips in digital cameras feature larger number of pixels in the same area in order to be able to capture higher spatial frequencies. In low resolution CCD chips aliasing is noticeable in digital photographs by the appearance of regular grids or stripes. These days this has become rather rare as the available number of pixels per unit area in consumer devices has dramatically grown over the years. With the reduction of pixel size the limitation in CCD chips most often is due to noise. Cheaper CCD chips feature many pixels of very small area. Therefore only a small amount of light is shown on each pixel. In low illumination images the electronic noise in the pixels becomes visible. Expensive digital cameras feature much bigger

CCD chips and bigger objectives, but not necessarily more pixels anymore. But since the individual pixels (and the objectives) are larger they can collect more of the light and therefore have a much bigger margin before the electronic noise in the pixels becomes significant.

(2) Noise & Interference

Electronic noise is the *uncontrollable fluctuation* of your signal with time. It can have many sources, some may be reducible by a better layout of the measurement setup, others are irreducible. You *always* will have some level of noise in your measurement. The question is *how significantly the noise affects your measurement*. Look at the two traces of a signal below.



Signal with negligible noise level

Signal with significant noise

The upper trace has no visible noise on it, the lower trace a substantial noise level. Do you think you can reconstruct the original signal from the lower trace? Indeed you can. When you fit a sine-wave to it you still would get the frequency, amplitude and phase of the wave with a pretty good precision, i.e. a relatively low error. The error becomes low as you have measured a reasonably large number of periods to average out the fluctuations on the individual data samples. Still the errors on the fit of a sine-wave to the upper trace would be even lower.

The *significance* of how the noise affects your measurement can be quantified by the **signal-to-noise ratio** (**SNR**), which can be defined as (on a logarithmical scale, as engineers like it):

$$\mathrm{SNR} = 10\log_{10}\left(\frac{V_s{}^2}{V_n{}^2}\right)$$

where V_s and V_n are the root mean square (rms) signal and noise voltages, respectively. An rms noise voltage that is the *same* as that of the signal would lead in this definition to an SNR of zero. At this point the signal would be almost buried within the noise and would be *at the edge of being successfully reconstructed* from the measured data.

Two of the many possible sources of noise are *fundamental* and *unavoidable*.

Thermal noise (also known as Johnson noise):

Any resistor *always* generates a noise voltage. The origin of this noise is the random fluctuations in the distribution of electrons in the resistor caused by the *thermal movements*.

Above T=0K these movements are inevitable. They become stronger as the temperature increases and so the rms amplitude of this noise voltage *increases with the temperature*.

This noise has a very specific character. It has a *flat spectrum* (the rms amplitude is the same at all frequencies). As a result the *amount of noise varies with the bandwidth* of the measurement. Thus the Johnson noise voltage is:

$$V_{noise}(rms) = \sqrt{4k_B TRB}$$

Here $k_B = 1.38 \times 10^{-23} JK^{-1}$ is Boltzmann's constant, T is the temperature in K, R is the resistance in Ohms and B is the bandwidth in Hz. Because of this bandwidth dependent behaviour the irreducible noise of a device often is quoted in units of $\mu V / Hz^{1/2}$. The user can calculate the noise present in the application once its bandwidth is defined.

Current noise (also called shot noise):

An electric current is not a continuum. It is made up of the flow of many discrete electrons. Therefore there is a noise contribution due to the *statistical fluctuations* in the number of electrons *passing a barrier in a component*. There is a major difference to Johnson noise: if there is no current flowing then there is no shot noise! Quantitatively the shot noise is:

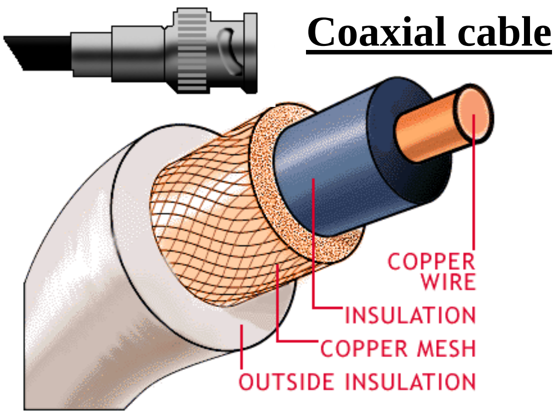
$$I_{noise}(rms) = \sqrt{2qI_{dc}B}$$

Here q is the charge on an electron, I_{dc} is the steady current and B, again, is the bandwidth in Hz. The formula is only actually true for *statistically independent electrons*, such as those diffusing through a barrier in a junction diode. Where there are *correlations* between the motions of the electrons, like in a resistor, the current noise is *significantly reduced*, often negligible.

Interference:

Changing electric and magnetic fields in the environment of your measurement setup will *influence the flow of charges in a circuit*, this is called **interference**. Aerial communication (mobile phones, wireless, Radio and TV aerials) works this way. But when you measure a (small) signal you want your cables *not to act as an antenna* and pick up RF (radio frequency) signals which would disturb your signal to measure as an *additional noise component*.
The question is, how do you stop your cables or traces on a printed circuit board acting like a radio aerial when you don't want it to? The fields couple to circuits in different ways.

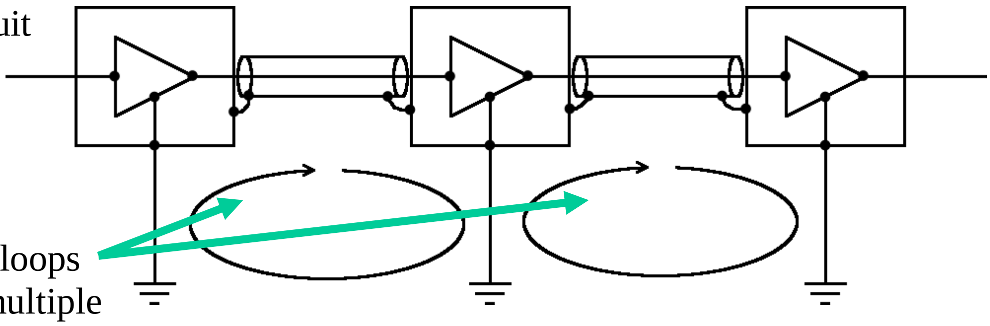
Capacitive coupling: Noise can be coupled into a circuit via stray *electric fields*. The charge carriers in a conductor will flow in response to an electric field, in order to reduce it in the centre of the conductor to zero. Capacitive coupling therefore can be removed by surrounding a cable by a conducting shield layer which is well grounded. This is what is done in a coaxial cable.



http://www.phy.davidson.edu/StuHome/phstewart/IL/spee

Inductive coupling: Noise can also be coupled into a circuit via stray magnetic fields. This coupling happens via *any loop created in a wire*. And loops can crop up where you don't expect them! Most common and annoying are so-called **ground loops**. They occur when you make a mistake in the layout of your measurement setup and provide *two or more paths* for your ground or return current (by ignorance or by accident...). The following picture shows such a situation:

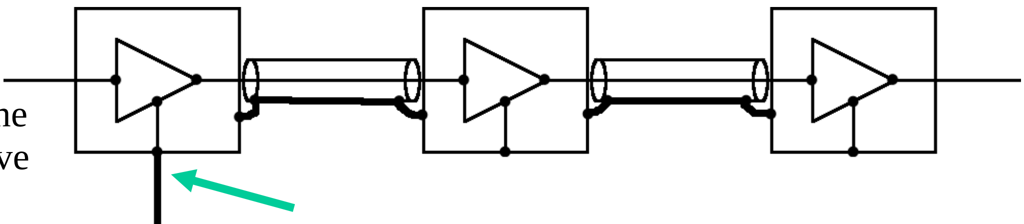
Problem: this circuit will pick-up noise, such as 50Hz from the mains and RF noise from the air



Ground loops due to multiple ground points.

There are several paths to the ground. Hence there are loops which pick up induced signals. These could be 50Hz pick-up from the mains as well as any RF signals. This has to be avoided! The layout of the setup has to be changed to a *single connection to the main ground point* and a *star-like distribution* of the ground to the locations where it is needed, to avoid internal loops as well. A fixed version of the setup above is shown here:

Solution: here the ground loops have been eliminated.



Single ground point made with thick Copper braid.

Coaxial wires are expensive and their connectors are space consuming. Often a much cheaper and space effective solution to transmit many parallel signal lines is sufficient: **twisted pair cables**. The two leads of a signal are twisted in regular short spirals with equal loop sizes. The current induced in one loop is *cancelled* by the next, leaving only small residuals along the line.

<u>Twisted pair</u>