# SAMPLING

Introduction

The focus of the course will be not the instrumentation sites, but the understanding of complex signals, we'll see algorithms and techniques to extract information from signals. First of all, we must study how the acquisition of data works.

Acquisition of data (or sampling): introduction

In general, we want to transform the information that we are measuring (like the temperature, for example) into a voltage signal because it's easier to carry the signal far apart from where the measurement takes place and so it's easier to collect all the data. Usually, we would like to obtain a linear relationship between the physical phenomenon and the output signal.

Let's imagine having a plot that represents all the data that we have collected as the one depicted in the picture. From this plot we can extract some information like, for example, the maximum value or the minimum value, but we are not able to obtain other parameters like the average due to the fact that we have a continues measurement that is not a function. In case of a function we would be able to compute the average since it is defined as the ratio between the integral of the function itself and the intervals:

$$avarage = \frac{\int f(t)}{T}$$
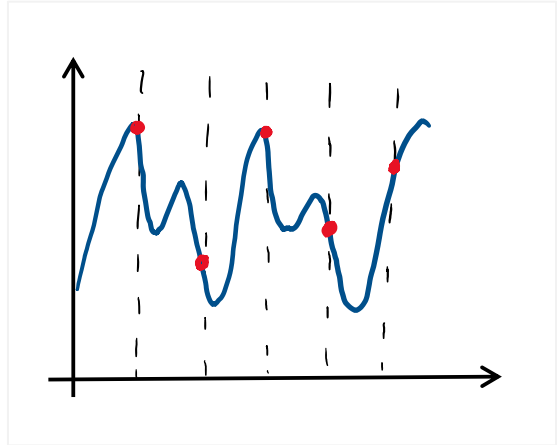
So, we have understood that one of the main problems is the fact that we are dealing with a continuous measurement and not a function. Another important element to consider is trying to understand how we should choose the time interval: this choice is up to us and it depends on what we think is meaningful. Depending on how many time intervals we choose we could increase or decrease the resolution[1] of our measurement that can be defined as the minimum variation in the input that causes a variation of the output, in other words it is the minimum variation that causes a variation of the output. Anyway, after choosing the time interval we can extract samples data[2]. With the term "sampling[3]" we mean understanding what we are seeing and reducing the information by extracting what we think is relevant. When we sample, we are losing some information. In general, we are losing whatever is in the between of the time interval that we have chosen. Since we are dealing with complex phenomena and with fast phenomena, we are going to see a device that allow the process in an automatic way; this device is the ADC converter. Let's now try to understand how it works.

Acquisition of data (or sampling): how it works

Let's, now, imagine having a transducer (T) that has two wires which are supposed to carry electricity in terms of voltage. How can we translate that voltage into a number? We can bring the voltage coming from the transducer into a component that compares that value with a standard one; this component is called SAR (Successive Approximation Register). The standard voltage used to make the comparison can be obtained


---
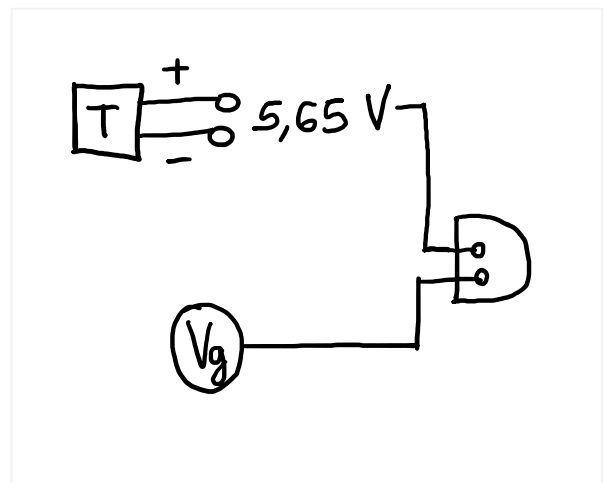
[1] Sensibilità
[2] Valori di campionamento
[3] Campionamento

thanks to another devise, called voltage generator (V_g). Since the voltage generator is able to generate a voltage between two number, we usually first generate a voltage in between these two values, then we compare this value that has been generated with the value that comes from the transducer: by doing a comparison between these two values we can understand if the voltage coming from the transducer is higher or lower than the one coming from the voltage generator. By repeating these steps, we end up finding our value; generally, the number of comparisons is equal to the number of bits. Of course, we can increase our resolution by improving the number of the intervals, or, in other words, by increasing the number of bits. Form these considerations we can easily understand that the resolution of our measurements is strictly connected to the number of bits of the converter. We'll see in just a minute the definition of resolution. Anyway, thanks to this technique of comparison we can also obtain a code in 0 (if the value that comes from the transducer is lower than the generated one) and 1 (if the value that comes from the transducer is higher than the generated one) that is easy to send to a computer.
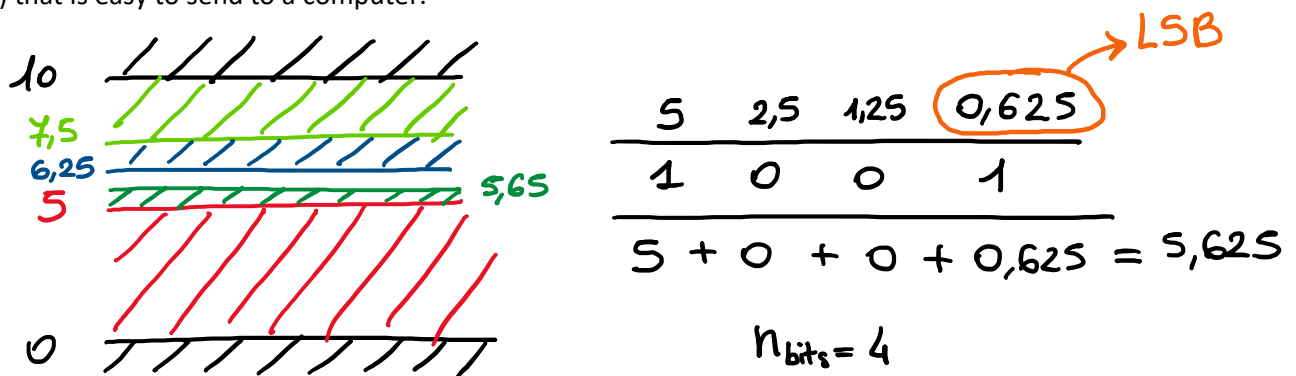
Of course, the number that we obtain by applying this technique is usually different form the real one, this error, that can be defined as the difference between the real voltage and the output result of our measurement is called quantization error and it is at maximum equal to $\pm LSB/2$ (we'll see that LSB stands for Least Significant Bit, for us it is the same of resolution):

$$quantization\ error = real\ voltage - output\ result$$

This method of obtaining a voltage output is called, as already mentioned, ADC which stands for Analogue Digital Converter. We can foresee how much the resolution is by taking the range of the converter and divided it by 2 to the power to the n choice of sampling (2^n). The number of choices is the bits and so we could also write 2^bits:

$$resolution = RES = \frac{range}{2^n} = \frac{range}{2^{bits}}$$

Examples:
- $0 \div 10$: range = 10
- $-10 \div +10$: range = 20

Usually a very good range is 24 bits, but also 16 bits can be considered a good range, too. The last value of the one obtained by using this method is called LSB which stands for Least Significant Bit, for us it is the same of resolution. So, we call resolution:
- the number of bits (informatic resolution)
- the least significant bit (informatic resolution)
- the definition we have previously provided (electrical resolution)

In general, when we'll talk about the electrical resolution, we'll refer to the definition in volt whereas when we'll say informatic resolution we'll mean the number of beats or the LSB.

Acquisition of data (or sampling): technological issues

So, until now we have understood what sampling is; as we already seen it can be defined as the process of extracting significant data from a population. Then we saw how to transform an electrical signal to a number, and we defined this process as ADC. Actually, there are a couple of technological issues that we must cover:

1. If we should design a converter where would we invest our money on? What is most crucial component for an ACD? In term of accuracy the voltage generator is the most critical part. Of course, also the number of bits is important, but the voltage generator is the most crucial one. We can define the accuracy of the ADC as the difference between two steps as depicted in the picture:
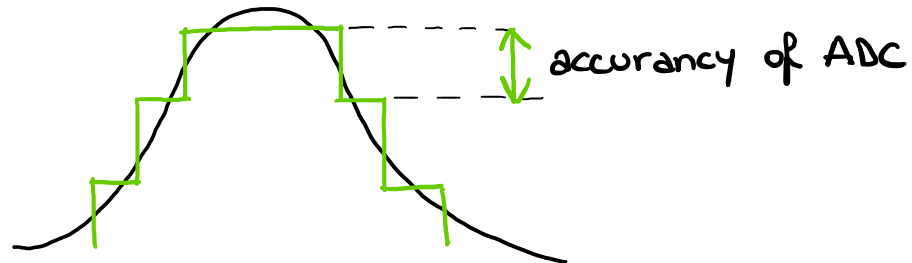
2. The time that is required to convert the signal into a number is the second important issue that we must cover. In fact, if the conversion process is too long, in that time the signal could have been changed. So, we need to freeze the voltage and then operate a conversion. Let's now try to understand how we can freeze the voltage. In this case we can freeze the voltage by installing a capacitor which is a device used to store electric charge that consists of one pair of conductors separated by an insulator. The new system will be the one depicted in the picture; as we can see we also have two switches and the negative wire is connected to the ground. Let's see how it works; when I decide to convert the signal, I close the switch on the positive wire and keep open the witch on the negative wire; thanks to this the capacitor accumulates the charges, then I switch off the switch on the positive wire so I can convert the signal. Due to the fact that there is nothing in nature that is able to keep the charges constant, the voltage that has been stored into the capacitor starts decreasing but in a slow way; of course the speed of decreasing should be slower than the time I need to do all my comparison and conversion and so it must be lower than the LSB. In other words, I have to freeze the signal and to do this we need to charge a capacitor, at this point the voltage is not coming anymore from the transducer but it comes from the capacitor, the capacitor cannot do anything else than losing its charges. By a practical point of view, we cannot use a huge or infinite capacitor since I would spend too much time to charge and to discharge it, and this is not convenient. When the conversion is over, I have to discharge the capacitor and to do this I close the switch on the negative wire so that the capacitor is connected to the ground.
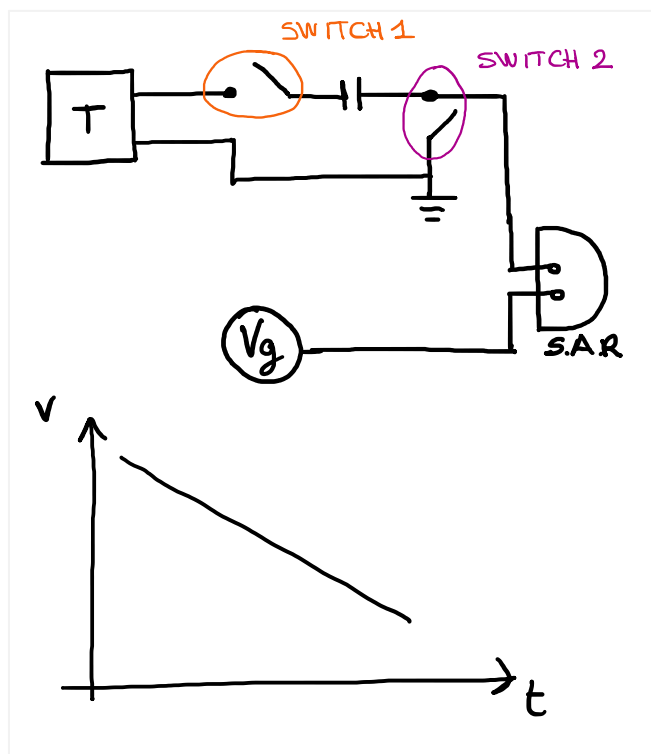
Another important point to keep in mind is that most of the time the users are not interested in seeing the measurements expressed in bits or voltage, because they only care of what they are measuring, like the temperature, for example.

Example: let's imagine having a thermometer with:
- Sensitivity: $S = 0,1 \, V/°C$
- Range: $\pm 10 \, V$
- Resolution: $12 \, bits$

Let's compute:
- Electrical resolution: $Res_v = \frac{10-(-10)}{2^{12}} = 4,88 \, mV$
- Resolution in number of levels: $Res_n = 2^{12} = 4096$
- Resolution in engineering units (it is a link between the sensitivity and the electrical resolution):

$$Res_{E.U} = \frac{Res_v}{S} = 4,88 \, mV \cdot 0,001 \frac{V}{°C} = \frac{4,88 \, mV}{100 \, \frac{mV}{°C}}$$

Finally let's highlight that with 16 bits the electrical resolution is in the order of the micro volt whereas with 24 bits the electrical resolution is in the order of nano-volts. In this case we should use amplifiers because otherwise the value measured is not "valid".

Acquisition of data (or sampling): problems

Let's now discuss a little bit about the different errors that we can end up with during the measurements if we choose the wrong range into the converter and so if we choose wrongly the values on the vertical axis (in the next paragraph we'll discuss the errors that can occur when we act wrongly on the horizontal axis, the time):
- Problem 1 = if we are measuring an amplitude lower than the LSB we'll end up having a 0 since the converter is not able to catch that signal due to the fact that the range on which we are working on is not big enough
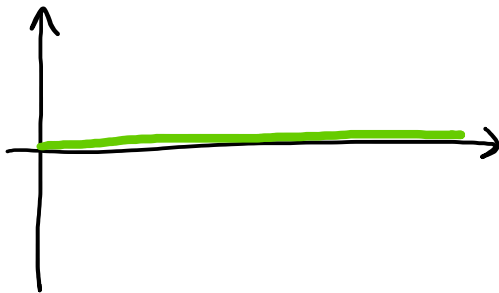
- Problem 2: clipping = it is a form of distortion that limits a signal once it exceeds a threshold. Clipping may occur when a signal is recorded by a sensor that has constraints on the range of data it can measure. For example; if we are dealing with a harmonic signal that goes form -10 to $+10$ but we are using a range of $(0, +10)$, we'll obtain something like the figure shown since the negative signal is not converted because it out of the range of the converter itself
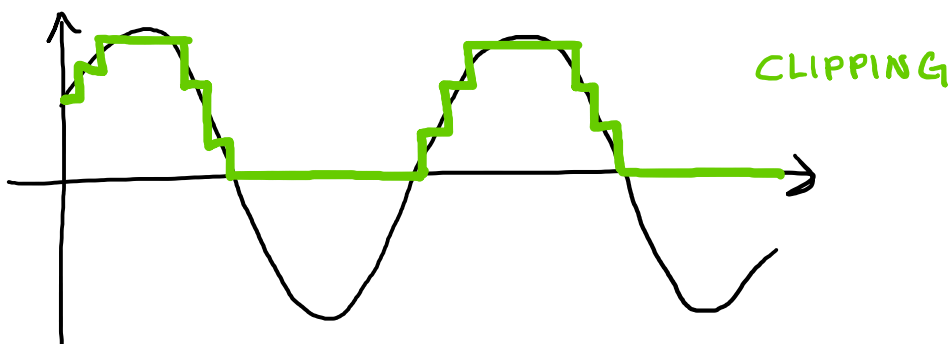
- Problem 3 = it can occur when we use the wrong range in terms of unit of measurements in fact if the range is expressed in V whereas the signal is in mV we could not be able to convert correctly the signal itself.

Time of sampling: introduction

Until now we have studied the conversion only by the amplitude point of view. Let's now discuss about time. When do we do the sampling? When do we start sampling and how often we need to do that? The time of sampling is crucial because we can introduce strong errors in our sampling. Usually errors are related to fast measurements, but they can occur also in slow measurements.

We can define the time of sampling[4] T as the product between the number of samplings N and the time interval dT or the ratio between the number of samplings N and the frequency of sampling $f_{sampling}$:

$$T = N \cdot dt = \frac{N}{f_{sampling}}$$

Time of sampling: aliasing

Let's now study the following plot and let's assume that the points indicated are at a distance of $1\ ms$; in this case choosing $1\ ms$ as the sampling time is not smart because we would extract the same value each sampling; a nice idea would be do measure twice per cycle. We can define the sampling frequency $f_{sampling}$ as the average number of samples obtained in one second (samples per second). The sampling frequency is important because if we choose it in a wrong way, we could reconstruct a signal that is different for the real one; this problem is called aliasing. An example of aliasing is what we can see in movies when we see a car that is moving forward whereas the wheels seems to move backwords. There is a

theory that we are going to follow according to which the sampling frequency must be higher than twice the signal frequency; this theory is called Shannon theory:
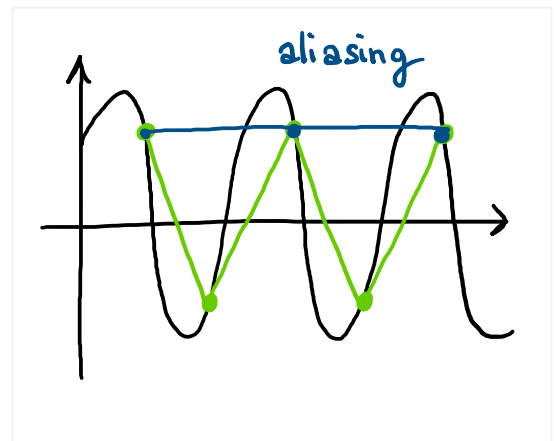
$$f_{samplig} > 2f_{signal}$$

The main problem of this theory is that we could not know the frequency of the signal and so it's like a paradox because if we know something, we do not measure it, usually we measure it when we do not know the signal. So, how can we solve this paradox or problem? The only way is unfortunately to limit the damage status to our measurements; since I cannot know the signal until I have measure it, but if I do not respect this condition I do an error, the only thing I can do is removing from the signal all the things that do not depend on this relationship. So, to avoid this problem, we usually add an antialiasing device which is an apparatus that removes all the component of the signal which have a frequency higher than twice my sampling frequency. Since we use it in this specific way, we call it filter. If we study the ratio between the output and input as the function of frequency, the filter has a function that can be plotted as the following one:


---
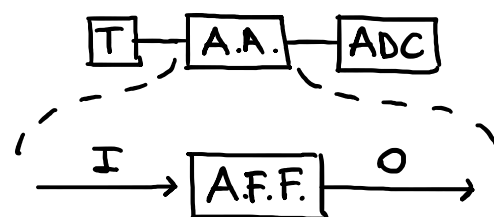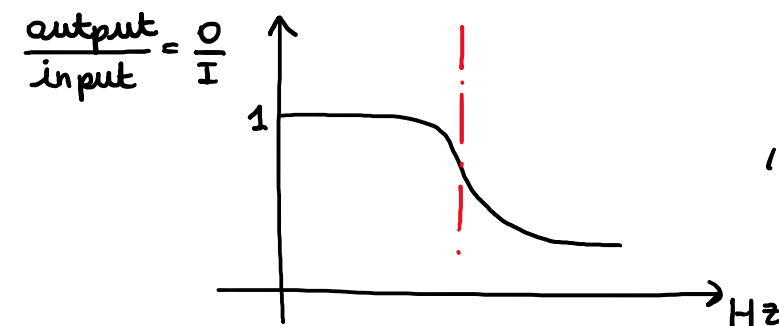
[4] Periodo di campionamento

Of course, the higher the frequency of the signal is, the lower is the ratio output/input (the idea is that I choose the sampling frequency, the red line; if the frequency of the signal are lower than this value, I keep them, otherwise if they are higher, they tend to be zero and so I neglect them). This type of filter is called low-pass filter; it can be defined as a filter that passes signals with a frequency lower than a selected cut off frequency and attenuates signals with frequencies higher than the cut off frequency. In other words, we are killing all the components that do not respect the Shannon theory. Usually I can write:

$$f_{aliasing} = f_{sampling} - f_{signal}$$

Example: let's assume having a sampling frequency at $1000\ Hz$. In this case the limit, according to Shannon theory, would be $500\ Hz$, this frequency is called Nyquist frequency or cut off frequency. I can plot it to clarify everything. In the yellow zone of the plot I have frequencies that pass that should not passed and frequencies that should have been passed but they didn't pass.
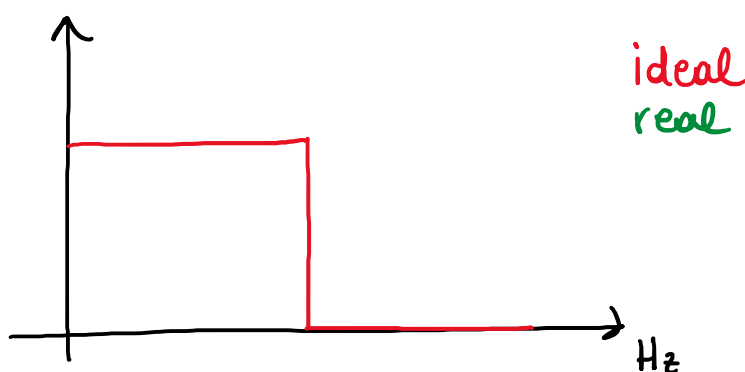
Usually the cut off happens at 60% of the ratio output/input which is more or less $\sqrt{2}/2$. But we must be sure to not have aliasing and so we take the cut off frequency and multiply it per 2,56. This value comes from the point after which the attenuation of the signal is 100dB and so very negligible. The concept is that the filter that we choose is characterised by a specific cut off frequency, we know that we are able to sampling correctly only signal characterised by the Nyquist frequency (which is equal to half of the sampling frequency), but if this frequency is higher than the cut off frequency we for sure not be able to sampling anything and so we'll choose a sampling frequency equal to the double of the cut off frequency. In addition to this to be extremely sure to sample correctly we'll actually choose a sampling frequency equal to $2,56 \cdot f_{cut\ off}$.
[La frequenza ci cut off è propria del filtro che si sceglie: sapendo che il filtro taglia tutto ciò che sta sopra ad essa e sapendo che si riesce a campionare correttamente dei segnali che hanno una frequenza minore della metà di quella di campionamento (cioè $f_{Nyquist} = f_{sampling}/2$), se tutto ciò che sta sopra a $f_{cut\ off}$ si perde, allora si sceglie una $f_{sampling} = 2 \cdot f_{cut\ off}$. Poi per stare sicuri, siccome il filtro non è netto, ma ha un gradino addolcito, si prende $f_{sampling} = 2,56 \cdot f_{cut\ off}$.]
The only question that we haven't answered yet is: at which frequency do I keep the signal? I must simply use the common sense.

Exercise: imagine having three signals: a, b and c with 8 bits and with three ranges that can be $\pm 10V;\ \pm 5V;\ \pm 25V$.
  a. $3\ V$ @ $1000\ Hz$
 b. $6\ V$ @ $2500\ Hz$
  c. $0,01\ V$ @ $100\ Hz$

Let's answer to these questions:

1) <u>Which range should we choose?</u> The range that we should choose is $\pm 10V$ since is the smallest range in which the three signals that are given are defined. Of course, we could have chosen also $\pm 25V$, but by choosing this range we would have decrease the resolution of the measurement. I cannot for sure choose the range $\pm 5V$ since one signal would not be included ($6\ V$)

2) <u>Which is the sampling frequency?</u> In order to answer this question, we must analyse the higher signals since it would be one easiest to be subjected to aliasing. So, we start from $f_{signal} = 2500\ Hz$. From this we could decide one value as our cut off frequency $f_{cut\ off} = 2750\ Hz$, for example (every value higher than our f_{signal} will be fine, of course we must use the common sense). After computing the cut off frequency we can determine the sampling frequency by multiplying it per 2,56: $f_{sampling} = f_{cut\ off} \cdot 2,56 = 7040\ Hz$.

FREQUENCY ANALYSIS

Introduction: experimental measurement of the sound

Let's see a chain of measurements: we have a measurement microphone which is made of a cartridge inside which we have a capacitor. The capacitor is made of two armaments: one is fixed whereas the other is a membrane that moves as soon as it gets in contact with the sound wave. Basically, this capacitor is able to measure the sound pression that arrives to the membrane as a wave. After this device we have the data acquisition board; inside this instrument we have both the anti-aliasing filter and the ADC converter. So, at the outlet of the anti-aliasing filter we still have analogue signal; only after the ADC we have the digital signal. In general, the setup configuration can be depicted as:
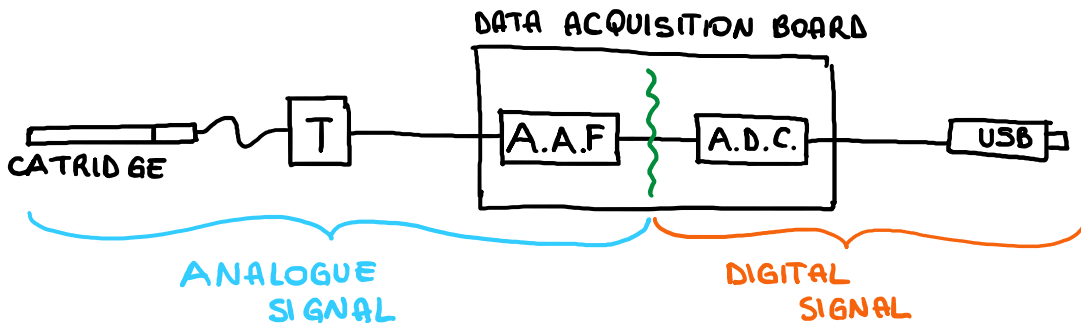
Of course, the transducer is connected to the data acquisition board thanks to a cable and the data acquisition board is usually connect to the computer through a USB link. It's important to highlight that usually we would spend more money on the analogue cable so that we can have a more precise measurement; spending a lot of money on the cable that connect the data acquisition board to the computer is useless and worthless.

As usual, knowing the range and the number of bits of the data acquisition board we can compute the resolution in volt. In this particular type of data acquisition board used in class for the experiment we have:

- $range$: $\pm 5V$
- $bits = 24$

$$Res_V = \frac{10}{2^{24}} = 0{,}6 \mu V$$

Now, knowing that the audible sound range is around $20000\ Hz$ we can choose the sampling frequency as $50000\ Hz$. After defining the set up and the sampling frequency we can now proceed with the experiment and start our measurement. We can leave out[1] the values obtained because they are not important; the point of everything is that after the measurement we usually would like to try to report what we have seen in the plots and to do this the easiest thing to do is reporting the main frequencies that we saw on the plot itself. By doing this we are splitting the main signal into different components. This is the base of the frequency analysis.

Frequency analysis

From now on we are going to study the frequency analysis; this approach helps us to identify the main different components of the signal itself. It is based on the idea of splitting the signal into different contributions and for each component of the signal we have to report both frequency and amplitude. Usually all this information (amplitude and frequency) can be reported in a plot called spectrum; so, the spectrum is a way to represent the frequency analysis in term of amplitude and frequency, or, in other words it's the representation of the signal in frequency time.

How can we do this frequency analysis? What are the techniques and the methodologies that we can use? The first methodology that we can think of is the Fourier technique (since we have already studied it in other courses) but before this method another technique was developed and used: the band method. Let's start study this first method and then we'll move onto the Fourier technique.

---

[1] trascurare

Frequency analysis: band method

Before Fourier technique was developed, a particular type of filter called band pass filter was used. A band pass filter can be defined as a device that passes frequencies within a certain range and rejects frequencies outside this range. In general, the band is characterised by two extremes: a low frequency that we can indicate with f_L and a higher frequency indicated with f_H. In the following picture we can wee both the typical shape of the plot of a band pass filter and the configuration of the measurement chain system:
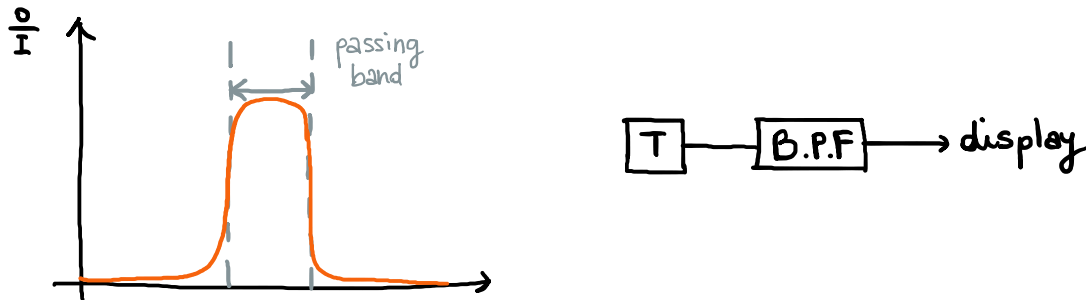
Example: referring to the example of the sound we can write the first line of the table for the first band than we change the band but keeping constant the bandwidth (A is the amplitude).

L (Hz)	H (Hz)	A
700	900	$1,7\ mV$
900	1100	$54\ mV$
1100	1300	$99\ \mu V$

Of course, the width of the band is a choice up to us. When the number of bands is increasing usually the bandwidth is decreasing and vice versa. A common approach of the band analysis is using the octave band. Octaves band is a band that had a frequency which is in terms of percentage constant with regards to the centre of the band. Usually when we have the centre frequency f_c, we can determine the two extremes by computing the following calculation:

$$f_L = \frac{\sqrt{2}}{2} f_c \qquad f_H = \sqrt{2} f_c = 2 f_L$$

Unfortunately, the band method has a major limitation: if we want to study in a precise way the signal, we need a lot of bands with extremely small bandwidth and so the computation cost will be extremely high. On the other hand, if we are ok with having a small number of bands, they'll be characterised by a very large bandwidth and so the solution obtained will not be precise what so ever.

Frequency analysis: Fourier method

To overcome the limitation of the band method we can use the Fourier approach. The Fourier theory states that each function can be reconstructed or seen as the sum of a *sine* function. We are not going to see the mathematical side of this theory but only the measurement side of it. In general, we can say that every function or, in our case, every signal, can be seen as:

$$S(t) = A_0 + \sum_{i=1}^{n} A_i[\sin(\omega_i t) + \varphi_i]$$

where $\omega_i t$ is related to the frequency of each component of signal, A_i it's the corresponding amplitude and φ_i is the phase. Usually we are going to neglect the phase. The coefficient A_0 indicates how much is the energy related to the frequency when $\omega = 0$, this component is called static component and it can also be identified with the term "DC" component. Of course, the Fourier series is defined only for periodic signals. Sometimes we can use the Fast Fourier Transform (FFT) despite the simple Fourier transform. This type transform is

extremely useful since it is already defined in Matlab. We'll see later on the difference between these two transforms.

Laboratory: it's extremely important to remember that in Matlab the "fft" function gives a complex function as the output. So, the solution is not normalised at all and so we must work on it. Usually the passages that we must follow in order to obtain the correct solution are:

- Considering only half of the vector otherwise I have double of the real pics. This is simply because in Matlab the output of the "fft" is a vector organized in the following way:
    - Even[2] numbers:

DC	f_1	f_2	f_3	f_{Nq}	$-f_1$	$-f_2$	$-f_3$

    - Odd[3] numbers:

DC	f_1	f_2	f_3	$-f_1$	$-f_2$	$-f_3$

- To normalise all the value, we must divide the result of the "fft" per the number of samples N and then multiplying everything per 2 except the DC term and the Nyquist frequency:

$$normalised\ value = \frac{result\ of\ FFT}{N} \cdot 2$$

The Fourier transform is of course a very powerful instrument, but it has two strong limitations and therefore some common sources of errors. The limitations are:

1. Issue regard frequency resolution = since it is the sum of the $i = 1,2,3,\dots,n$ component and it doesn't take into account also the contribution of other i like for example $i = 1.5, 2.5, 3.5, \dots$ I have a limitation in the resolution; we have, in fact, a jump between ω_0, ω_1 and so on (in general we have $\omega_i = i\omega_0$). The conclusion is that I can measure only frequencies that are entirely inside the period. I have frequency representation equal to $1\backslash T$.
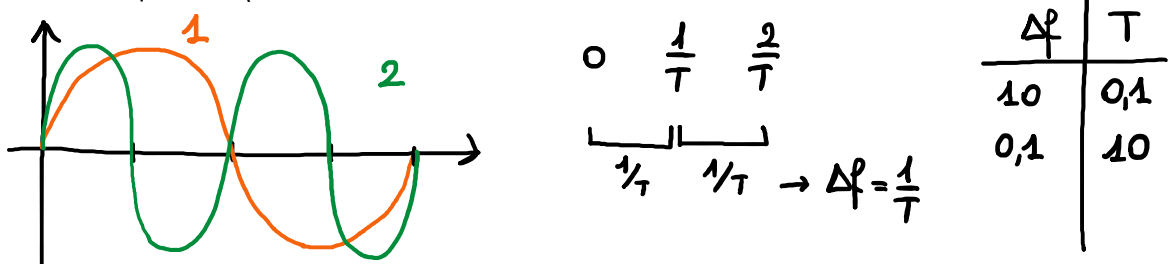
If I need a better frequency resolution Δf, I need to increase not the sampling frequency but the time acquisition. The frequency of sampling f_s is equal to the number of sampling n_s divided per the period T and so the link between these two quantities is:

$$\Delta f = \frac{1}{T} = \frac{f_s}{n_s}$$

$$f_s = \frac{n_s}{T} = \Delta f \cdot n_s$$

2. Issue regarding amplitude = the amplitude that Fourier uses it's a number, it's not negotiable value, but it's fixed. So, if the signal is not stationary Fourier ca be applied but it's not a good representation of what physically it's happening.

---

[2] pari
[3] dispari

Leakage: introduction
Let's now try to go in depth on the limitation of the Fourier transform. In order to do this let's imagine having a very plane signal (with plane we mean a clear signal, without noises) characterised by a period T and an amplitude A. Let's compute both the frequency of the signal, as the number of cycles over T, and the resolution frequency Δf and let's depicted in the same plot the output chart that we would obtain by applying the Fourier transform versus the real spectrum that we were supposed to have if we could work with all the frequencies and not only the discrete ones:

$$f_{signal} = \frac{1,5}{T}$$

$$\Delta f = \frac{1}{T}$$
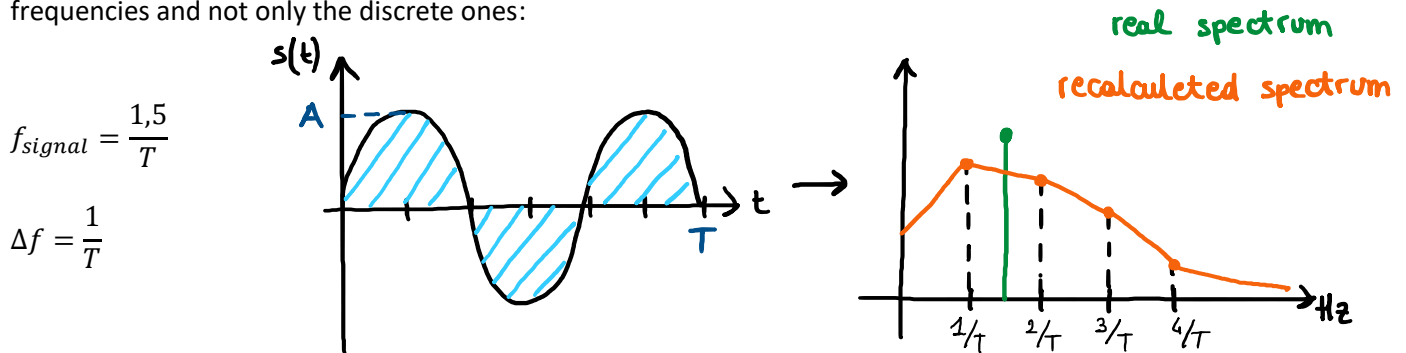
As we can see from the chart, the output of the Fourier transform is not a continuous graph; the spectrum, in fact, is characterised by the presence of some points. If we connect these points, we obtain a broken line and not a continuous one. This type of output, which can be considered as a discrete output since it's not continuous, is a consequence of the issue n°1 that we have previously mentioned. So, we can say that Fourier transform is not able to reconstruct perfectly the signal. Those points in the spectrum represent how the Fourier transform has distributed the energy of the signal in the discrete frequencies that were available.

Let's now highlight another aspect: from the plot of the original signal we can evince the fact that the mean value is equal to 0 and so, in the spectrum, we were supposed to have at frequency equal to 0 an amplitude, and so a value of the energy, equal to 0. But this doesn't happen in the spectrum that we have plotted in fact we have a certain amplitude that corresponds to frequency 0. This occurs since the Fourier transform usually compute the integral of the signal over the period and, in this particular example, the integral of the signal is not equal to 0 and so, by a mathematical point of view, we'll have a value of amplitude different from 0. So, the Fourier transform lead us having an error in the amplitude value that correspond to the frequency equal to 0. Because of this we always must pay attention to the solution provided by the Fourier transform since it couldn't have any meaning by a physical point of view (by a mathematical point of view it is correct, whereas by a physical point of view no). Actually, the same thing that we have just notice for the amplitude associated to the frequency equal to 0 occurs also for the other amplitudes but in those cases it's not so easy to compute it. This is a consequence of the fact that since we are "shifting" from a continuous signal to a discrete one (and so as a consequence of issue n°1), we have that the overall energy of the real signal is split into the spectrum in a wrong way. This means that during the reconstruction of the signal, by using the Fourier transform, the energy is spread around to preserve the total energy of the signal, since the distribution is discrete. So, what we see in the spectrum is an artefact of the calculation, made only to preserve the energy of the signal. This phenomenon is called leakage. Let's now study the cause of this phenomenon. Leakage can occur when not every component has a inter period submultiple of the acquired period which can be written as:

$$\frac{T}{T_{signal}} = \frac{f_{signal}}{\Delta f} = c = integer\ number$$

So, leakage can be defined a smearing[4] of power across a frequency spectrum that occurs when the signal being measured is not periodic in the sample interval. It occurs because discrete sampling results in the effective computation of a Fourier series of a waveform having discontinuities, which result in additional frequency components.

Before studying how we could eliminate or, at least reduce the leakage, let's talk a little bit about one of the main consequences of the leakage. As soon as I obtain a spectrum in which we have leakage we for sure have

---

[4] Spalmamento, divisione

masking. Masking is a phenomenon according to which in a leakaged spectrum we are not able anymore to recognize the presence of another sinewave signal or not.

Leakage: solution
To eliminate this problem, we can think of neglecting the last part of the signal that is the cause of having an integral different from zero. By doing this we have some good and same bad things:
- Good: the mean goes back to 0
- Bad: we would have the problem of the right-angled corner[5]
- Bad: we would have to look at the signal and decide which part of it should be cut and what should be kept

So, we conclude that this method of reducing or eliminate the leakage is not smart, let's now study something better. We can actually attenuate the leakages in two different ways:

1. Improve the frequency resolution = by improving the resolution we reduce the integral of the extra energy over T. This method, seen in a graphical representation, works perfectly since we are reducing the spacing of the frequency and so we are also reducing the spread of energy. Since Δf is reducing, we can say that when the period tends to infinite the leakage tends to 0. This method of reducing or trying to eliminate the leakage is actually not a very good idea since considering a long period the signal can change.
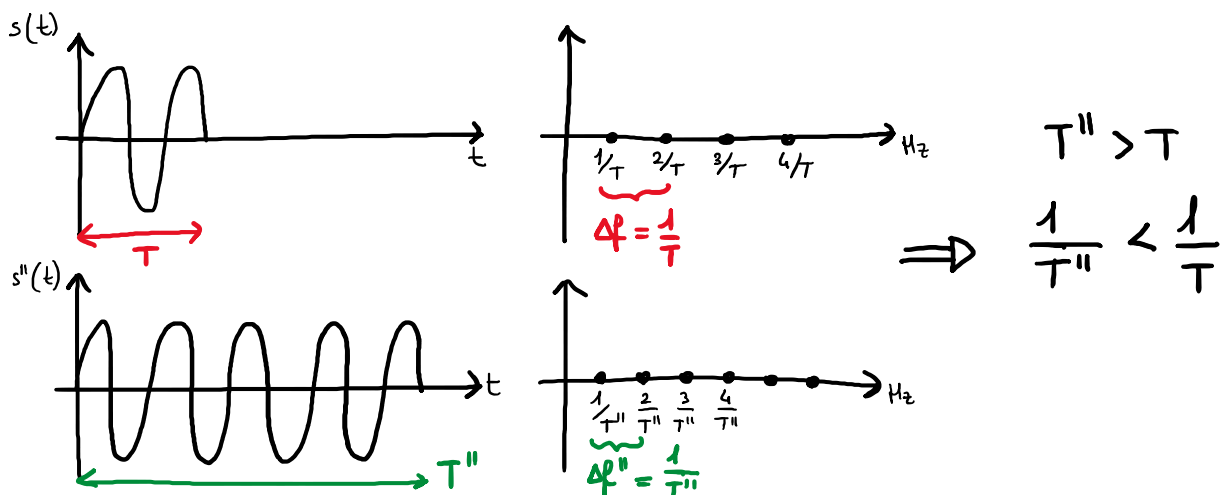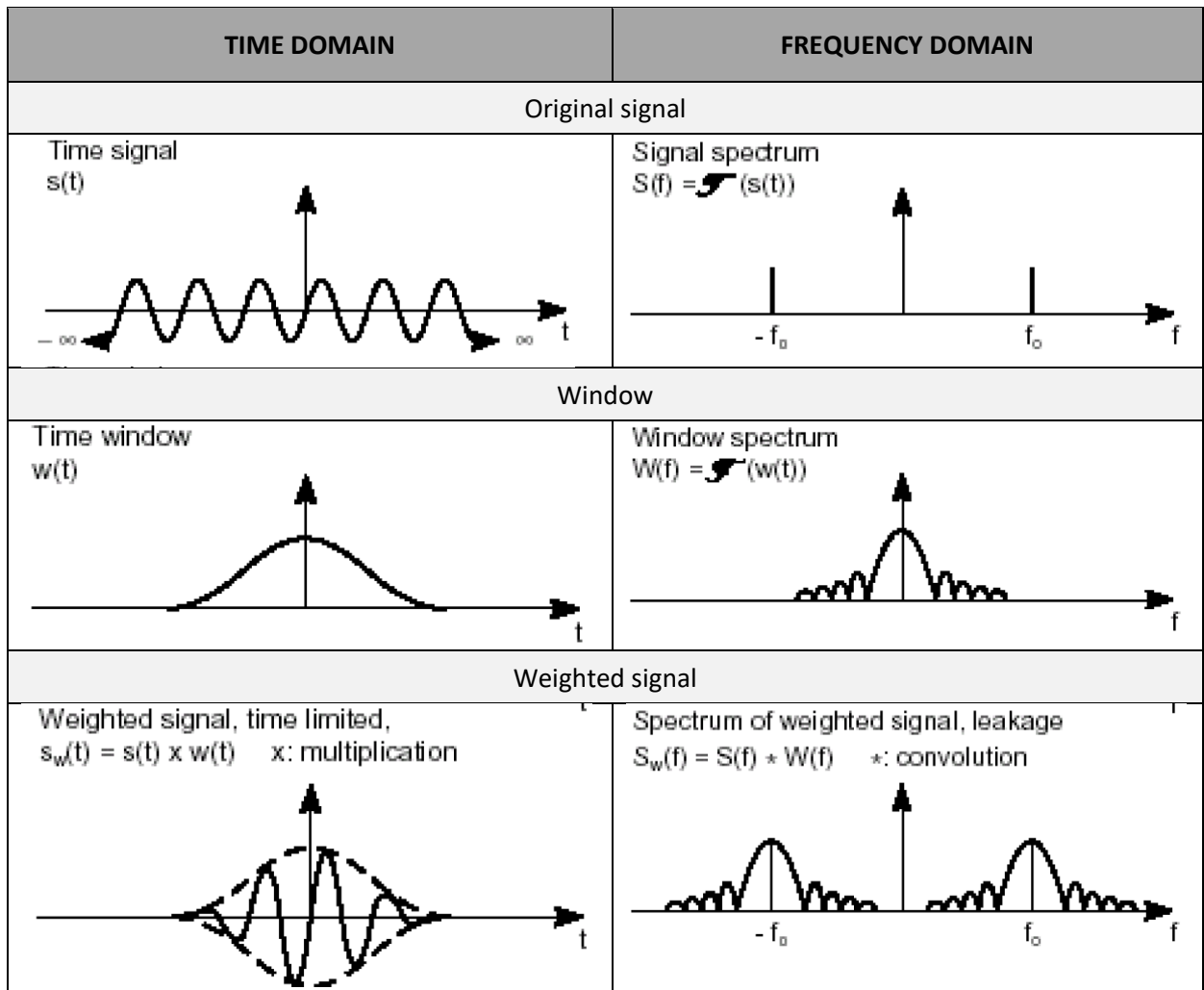

2. Windowing = we can think of multiplying the signal per a numerical function which is something like the one depicted in the following plot so that we end up with a new signal (by a "mathematical" point of view multiplying the signal per a function is like multiplying the array of the signal with an array of numbers that I have generated and chosen, depending on the function I've chosen). By doing this we actually have reduced the energy of the signal (the energy of the signal is underestimated) but we have also reduced the problem of leakage.
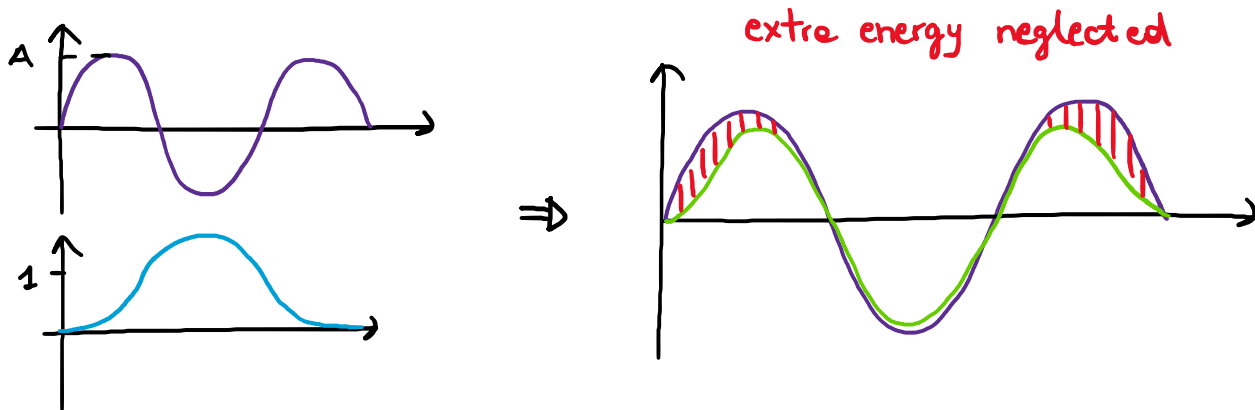
The function that I've chosen, and I have multiplied is called window. So, we can also define the window as a function that is able to removes the extra cycle in an original signal that is the main responsible for

---

[5] Angolo retto

leakage. Since they are function, they are defined in time domain, but they can be shifted in the frequency domain, too. Let's try to understand better this point and how is the typical shape of a window's spectrum by looking at the following image:

TIME DOMAIN	FREQUENCY DOMAIN
Original signal	
Time signal $s(t)$	Signal spectrum $S(f) = \mathcal{F}(s(t))$
Window	
Time window $w(t)$	Window spectrum $W(f) = \mathcal{F}(w(t))$
Weighted signal	
Weighted signal, time limited, $s_w(t) = s(t) \times w(t)$ x: multiplication	Spectrum of weighted signal, leakage $S_w(f) = S(f) * W(f)$ *: convolution

So, the shape of a window's spectrum is made of a lot of lobes. There is a main lobe which aim is to increase the effect of the interested frequency and then there are a lot of side lobes which main effect is to decrease the effect of the unexpected frequency.

There are many kinds of windows. The most common window is called "Hanning window" and it has a shape like the one we have just mentioned:
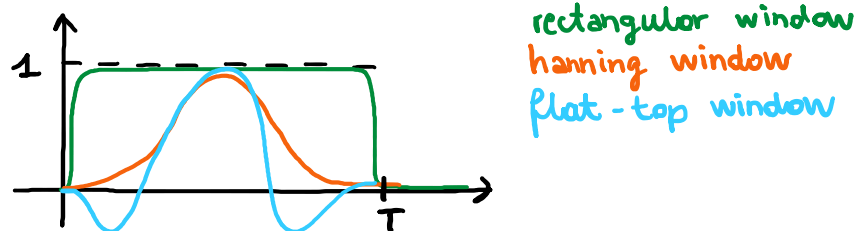
We could also choose a more selective window or a less selective window depending on what we are interested in, so:

- More selective window = I'm more interested in the frequency selectivity (having the right frequency) and less interested in the energy of the signal
- Less selective window = I'm less interested in the frequency selectivity but more interested in the energy of the signal

When we don't want to apply any window, but we are using a software we must select and apply a rectangular window that actually doesn't change the signal whatsoever.

It's extremely important to highlight that the windowing technique doesn't help struggling against aliasing; in fact, when we use the windowing technique, we are working with an already sampled signal whereas when we talk about aliasing, we are working on the acquisition part and aspect. Another important point to keep in mind is the fact that if I apply a window to a signal in which leakage doesn't occur, I'll end up with a wrong spectrum, so applying a window can be done only after checking if there is leakage.

Exercise: Let's imagine having a signal with a lot of components, two of them are known. We must set the sampling parameters in order to have a good situation for the analysis. The two components are:

- $1\,V\ @\ 1000\,Hz$
- $2\,V\ @\ 999\,Hz$

First of all we must say what are the main devices that we need to use: we need the transducer, then we have to put the anti-aliasing since we don't know the other components of the signal (if the text would say that the signal is made only of this two components we could also avoid putting an anti-aliasing filter), finally we have to put the ADC. Now we have to decide the sampling frequency. We can choose the $1200\,Hz$ as the cut off frequency. We can then compute the sampling frequency:

$$f_{sampling} = 2{,}56 \cdot f_{cut-off} = 2{,}56 \cdot 1200 = 3072\,Hz$$

Then we can compute the Nyquist frequency:

$$f_{Nyquist} = \frac{f_{sampling}}{2} = 1536$$

At this point we can determine the minimum frequency resolution: $\Delta f_{min} = 1000 - 999 = 1Hz$ but we want to be sure to avoid having leakage and the spreading of energy. I want to see the spectrum goes down between the two frequency so that I can be sure that the energy of each frequency is related only to that frequency: so, we want a situation similar to the one depicted in the picture and so the frequency resolution must be at least equal to

$$\Delta f \leq \frac{1000 - 999}{2} = 0{,}5Hz$$

Thanks to this we can have a good spectrum without misunderstanding. Now we can finally compute the period at which I should sampling to avoid leakage (inside the period T I must have an inter number of cycle):

$$T = \frac{1}{\Delta f} \geq 2\,S$$
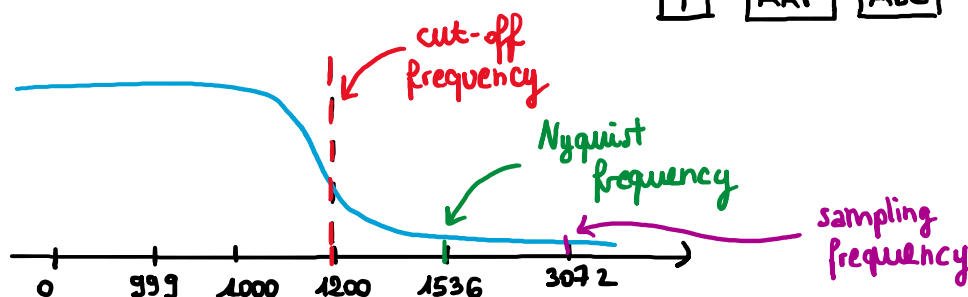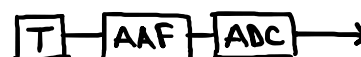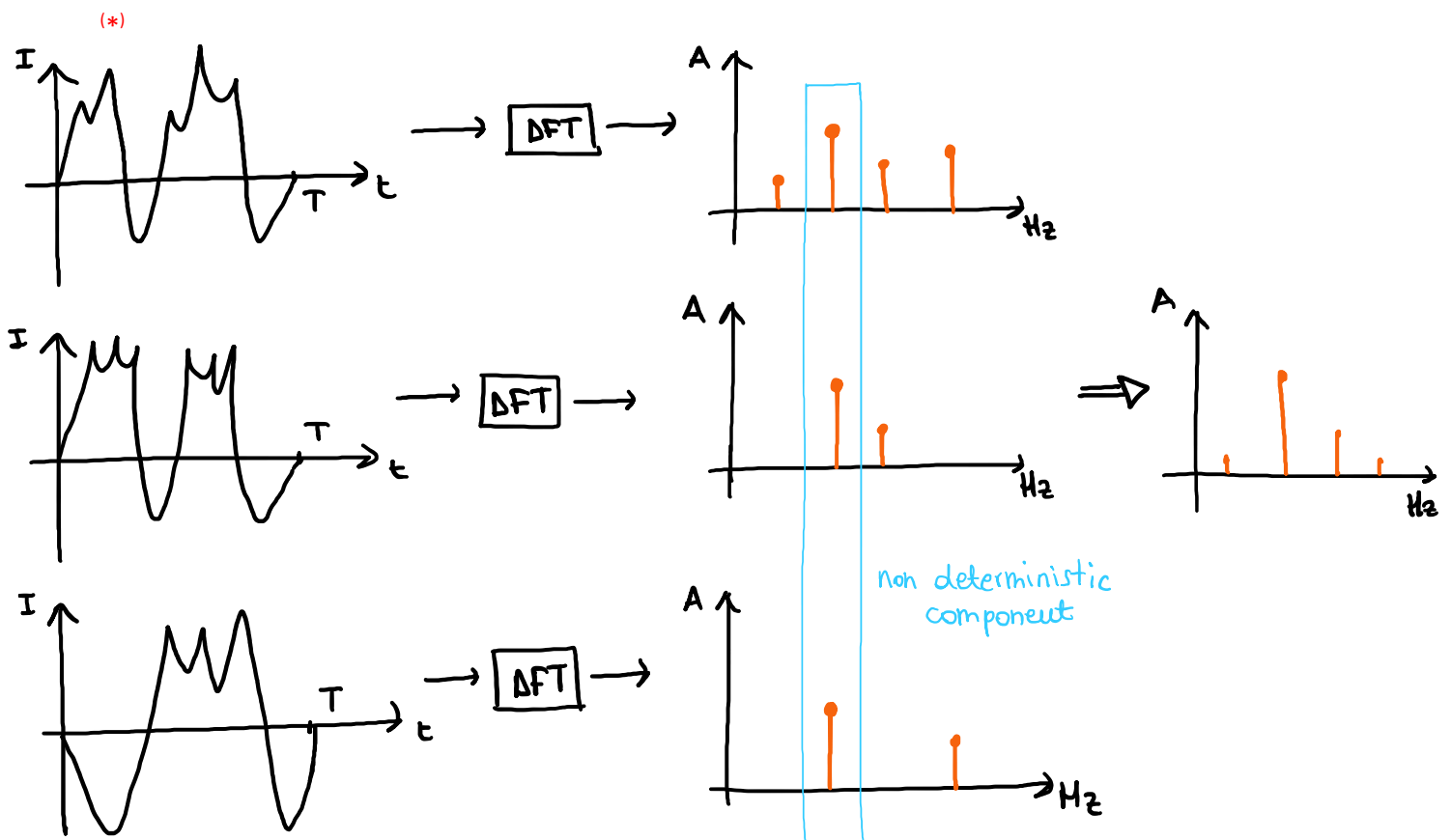
Deterministic and non-deterministic components

Let's imagine that we have to measure the acoustic noise in the classroom, due to the air conditioning system. Of course, we need to proceed with a frequency analysis, and we decide to use the FFT to perform it. In the room, besides our signal of interest we also have someone which is continuously whistling and someone that sometimes claps. We can notice that one of the two extra signals is always present (the whistle) whereas the other one is random because it appears and the disappears immediately (the claps). In term of signal process, we can divide the components of the signal into 2 big families:

1. Deterministic components = components which are always present in the signal. In this family we have:
   a. Components that we want to measure
   b. Other components which we don't want to measure, and which are responsible for making our measure complex. These deterministic components that we don't want to measure are called disturbances.
2. Random component (or noises) = these are non-deterministic components that can come directly from the physical signal or from the instrumentation that we using to perform our measurement

So, in our example we have 2 deterministic components: one is our signal of interest and the other one is the whistle and then we have one random component which is the clap. Let's now, first of all, focus our attention to the random components.

Averaging methods: introduction

If we have a complex signal can we reduce the noise in order to improve the accuracy, which is the overall quality of my signal's spectrum? In order to reduce the noises and increase the accuracy we can think of take different recordings of the same phenomenon and then "took them into account together" thanks to the operation of averaging. We could carry out two different common types of averaging. Both of these two techniques are based on the idea of doing a several recording with which we feed the digital DFT (digital Fourier transform) with each record. Then, we compute a spectrum for each record that we have previously done. At the end we'll have a lot of different spectrum which are in the same number of the recording that we have done. Actually, before applying the DFT it's important to check if the phenomenon is stable or not, in fact if it's not stable, we'll see that doing the average is basically useless. So, once I verify that the phenomenon is stable, I can apply the DFT and transform the time domain signal into a spectrum. We can expect that in each spectrum we will have some components that are repeated (common components that are the deterministic ones) and other ones that appears and disappears (the non-deterministic components).

Of course, the amplitude of the non-deterministic components will be for sure influenced also by the noises and so in the spectrum we are able to read not the real value but the real value plus the noise. So, we'll never be sure about the value of the amplitude of the non-deterministic components since they are influenced by the noises and we'll always have noises in our measurements. After obtaining these spectrums we have to do the average, let's now study he two methods that we can use to do the average.

RMS averaging method

Let's imagine having only 5 components, to make things easier, with only 3 measurements. Of course, the average cannot be done in columns since it wouldn't have any meaning; we need to make the average in rows and by doing the average in rows we are actually computing the average of the amplitude of each measurement at the same frequency. If we write everything in a table, we would end up with something like that:

|	I measurement	II measurement	III measurement	Average
Component C_0	A_0^I	A_0^{II}	A_0^{III}	$\overline{A_0}$
Component C_1	A_1^I	A_1^{II}	A_1^{III}	$\overline{A_1}$
Component C_2	A_2^I	A_2^{II}	A_2^{III}	$\overline{A_2}$
Component C_3	A_3^I	A_3^{II}	A_3^{III}	$\overline{A_3}$
Component C_4	A_4^I	A_4^{II}	A_4^{III}	$\overline{A_4}$

After computing the average, we can now draw the spectrum. So, this spectrum would be made by the average of the amplitude recorded previously in each measurement. By doing this we are introducing a more reliable evaluation of the measurement for each component and, in addition to this, we are reducing the noises components because, for instance, one random component presents in one measurement will not be present, or at least will have a smaller amplitude, in another measurement. Because of this we obtain a spectrum which is more reliable in term of amplitude and also clearer with regard to the noise because we have reduced them. A second good effect of using this technique is that we are reducing a little bit the leakage and the masking. Only if the deterministic components have an average value which is lower than the average of the noise, we cannot be able to find that component (and so would end up with a new sort of masking).
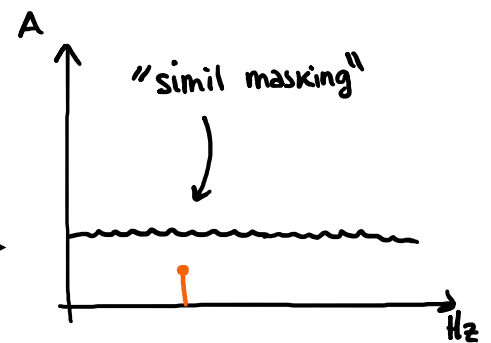
Now that we have understood how this works, we would like to understand how many records we should take. Of course, the higher number, the better will be. In general, the main drawback is that we must pay attention to the stability of the signal. The problem is the that sooner or later the signal will change and so we must be sure to do the sampling when the signal is stable and it's not changing (we usually cannot measure for a lot of time because we are not sure about the stability of the signal).

Imagine now having a super stable signal and not having problem with storage data, in this particular and impossible case if I take an infinite number of measurements, I still cannot eliminate the noises. I would be able to not have noises only if all the values would be 0, even with only one value different from zero we'll not end up with an average equal to zero.

How can we improve this technique? How can we draw to zero the noise components? Thinking of using a threshold and so deciding to neglect all the components smaller than this threshold is not a solution because in order to choose a threshold we should know the signal and so, in this case, I wouldn't do any measurement. Another possible solution would be summing to our original signal a second one with the opposite sign. This means having the same amplitude but with opposite phase, in this case in fact the average would be equal to 0.
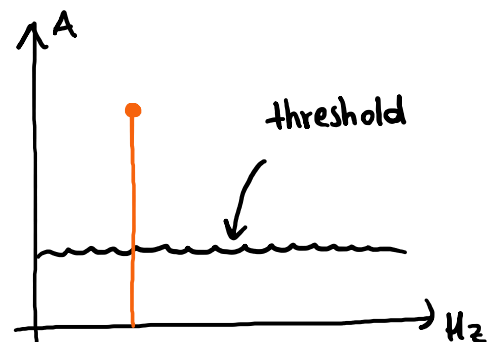
Until now we have done something which is not precise and coherent: we have talked about the average only in term of amplitude where we know that the output of the FFT is a complex number. So, until now we have completely neglected the complex part. Actually, in order to reduce the noise, is extremely useful to consider the all complex number which represent a vector.

Laboratory: let's now try to understand by a more mathematical and informatic point of view what does doing the RMS averaging means and how does it work. First of all, we must point out the fact that we start with a matrix y_tot which is made of:
- Rows = each row is an acquisition of the signal, a function of the time
- Column = each column represents an instant in which the signal is sampled

This matrix is, of course, a function of the time. Then we must apply the fast Fourier transform and normalise it in order to obtain a matrix, that we can call y_f_tot, which is a function of the frequency and not of the time. So, in this matrix we have that:
- Rows = each row of this matrix, represents a spectrum
- Columns = are referring to a specific frequency, after the normalisation the last column will be referring to the Nyquist frequency

The value inside this matrix are complex values and so in order to proceed with the RMS averaging we must consider only the amplitude. We then do the average of each column by using the following formula:

$$A_{f_1} = \sqrt{\frac{A_{f_{1,1}}^2 + A_{f_{1,2}}^2 + \cdots + A_{f_{1,n_{avg}}}^2}{n_{avg}}}$$

At the end we'll have a vector of real values, and each element of the vector is the average of the amplitude of a specific frequency:
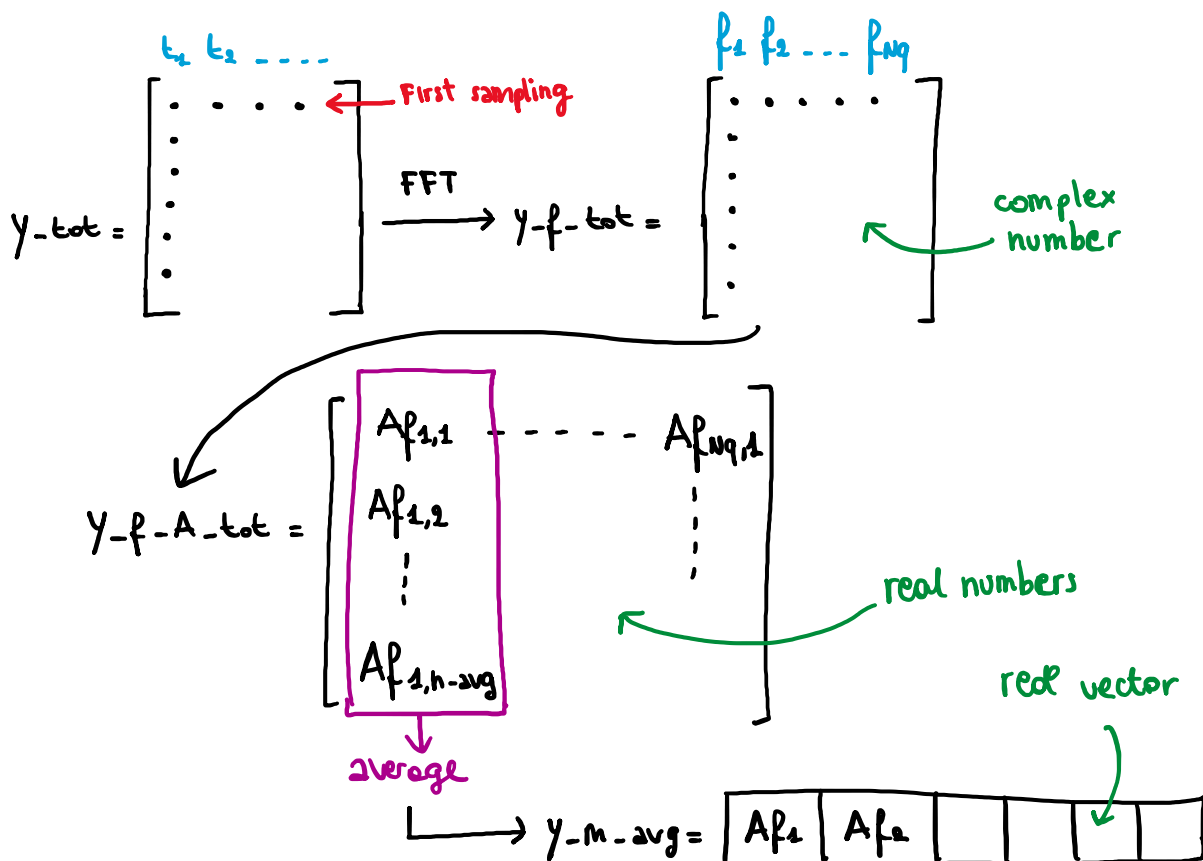
Vectorial averaging method

By considering also the complex part, when we do infinite measurements, we can easily obtain the same random components with the same amplitude but different phase.

The key point is that all the reasoning works if and only if the deterministic components have the same phase in all the measurement. This method of taking the complex average, also called the vector average, works only if the deterministic components are characterised by the same phase, otherwise also them would be equal to zero. So, we must be sure that the starting point of the measurement is always in the same point of the physical phenomenon.
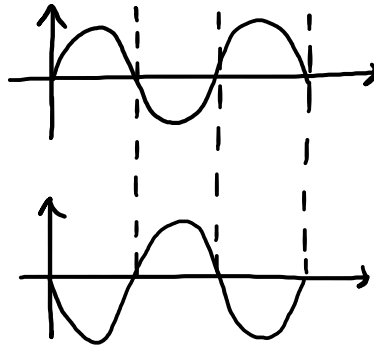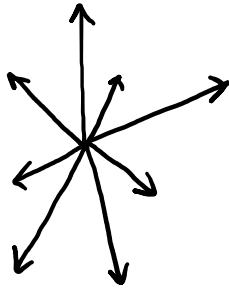


not good situation because we have different phase

Now we can conclude that the example that we have done previously [(*)] is not the perfect since the third measurement starts going down and not going up as the other two measurements. So, the idea is that if we don't have a safe phase reference this method is useless or at least doesn't provide any right conclusion.

This vectorial average cannot be use for the example of the air conditioning since we don't know the phase reference; in this example we should have used the first approach of average.

Example: in a reciprocating engine the trigger signal is necessary to have a safe phase reference. The trigger is saying to the ADC when it has to start the acquisition. Thanks to this device we are sure that the deterministic components are characterised by the same phase. As we have already shown with the air conditioning example, we are not always able to have the safe reference phase.
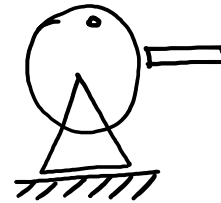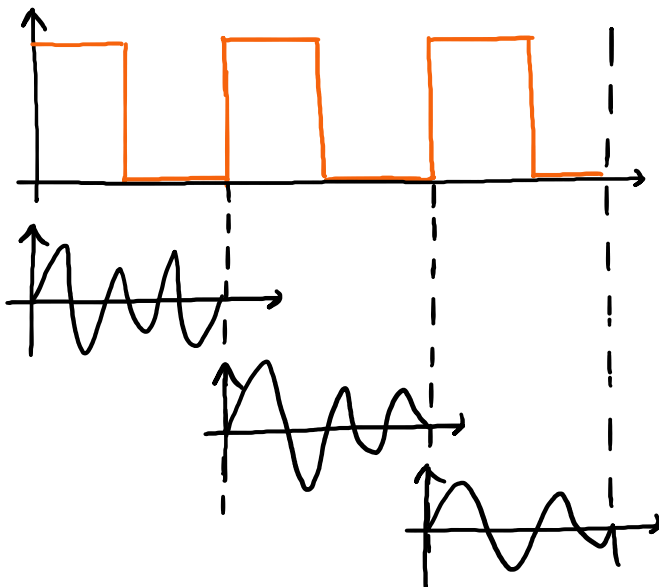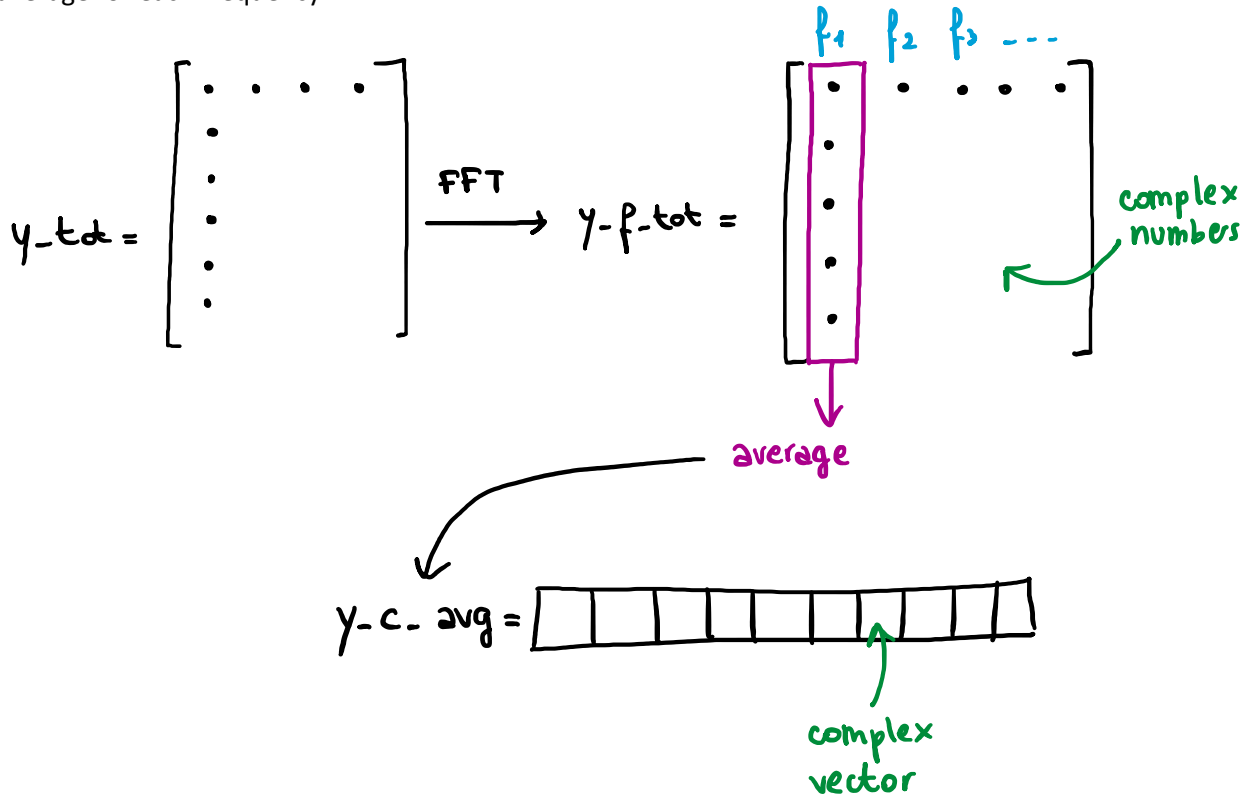
Comparison between RMS and complex averaging method

First of all, let's highlight once again that when we have a noise, the energy of the signal is randomly distributed in all the signal components and so we wouldn't have a correct value of amplitude. Anyway, we can easily notice by the following picture that the complex method reduces the noise much more than the average method this because the RMS does the average of all positive value and so we end up with a positive value that will always be different from 0; on the other hand since the vectorial method does the average of both positive and negative values we could end up with a very small value which can also be equal to 0. The positive value that we obtain with the RMS method is called carpet of noise and it's typical only of this method. So, this carpet of noise can never be equal to 0. In addition to this we can conclude that:

- Increasing n_{avg} (which means increasing the number od the sampling, the number of the rows) we'll end up with a more precise spectrum
- Increasing the noise, we'll effect negatively on the measurement since we'll be less precise

Let's also repeat once again another key aspect: it's extremely important to have a reference phase because if we don't have it, we must use the RMS method, otherwise with the complex method we'll have a wrong solution.
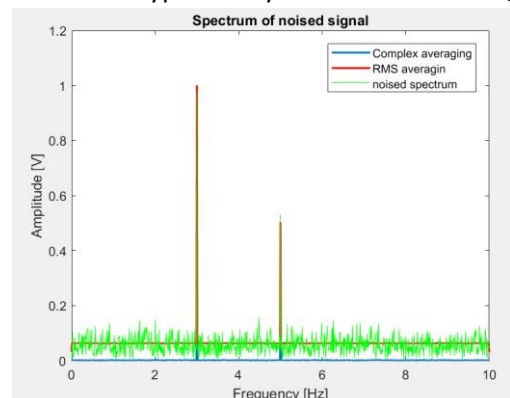
Rotating machines: introduction of synchronous acquisition

In the chapter on vectorial averaging, we have talked about the importance of having a safe phase reference and we have talked about the example of the rotor as an example in which we can usually use this type of average. So, the complex average is usually used with rotating machine since it's easier determine a phase reference. Although talking about rotating machines is very specific, it covers a lot of different fields because in reality most of the phenomenon are characterised at the back by the presence of a rotated machine.

Let's recap a little bit what we have said about the rotating machines: the main point is that the rotating phenomenon will be characterised by a period which is a submultiple of the acquisition time T. Let's now try to understand how we can improve, in a rotating machine, our analysis, with some engineering tricks and observation, knowing that the bad aspect of the FFT is that we could have leakage (fake energy that cover the spectrum; the second bad aspect is that we need a stable signal).
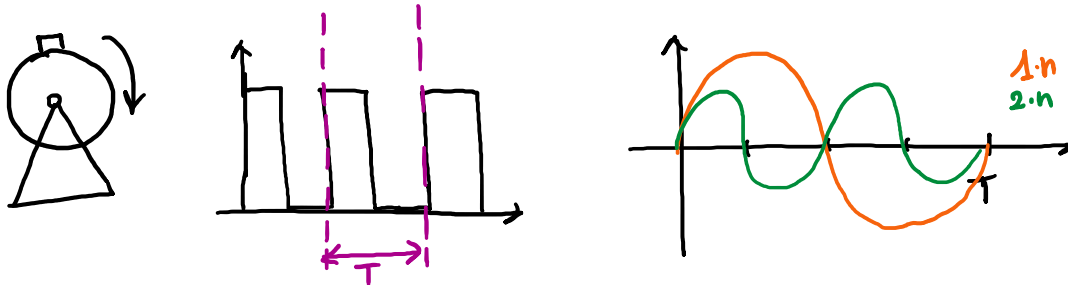

By exploiting this property of the rotating machines, we could in some way synchronise the acquisition time with the period of the phenomenon; by doing this we could increase the quality of the measurement; let's try to understand why. If it is true that at every cycle the phenomenon begins again, we can change the time acquisition and choose, as the acquisition time, the time of the complete revolution of the rotor. If we are able to do that, we will know for sure that in our acquisition time T, we'll have only an integer number of cycles. By doing this we are stating to have a signal which physically satisfy the same condition of the discrete Fourier transform and so we don't have leakage at principle. So, by choosing as the acquisition time the period of the phenomenon we are avoiding the problem of leakage.

Synchronous acquisition

The idea described previously has some problems since we don't know how long T will be and so we don't know how many samples we need to take in T. To solve this problem, instead of taking just one point as reference, we could apply a smaller wheel, with a number of defined teeth, to the rotor; this wheel rotates with the rotor itself and, thanks to a second transducer we will know for sure where and when we start a new cycle (as soon as we have passed the total number of the teeth of the wheel we know that a new cycle starts).
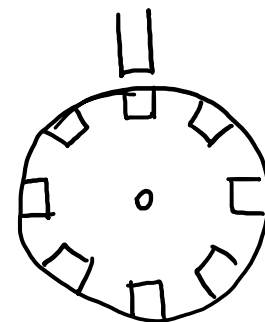
So, up to now our ADC was driven by an inside clock whereas in this case the start of the converting is no more determined by the internal clock, but it comes from the trigger itself. So, what we do here is passing from a time domain acquisition to an angular space domain acquisition. We don't acquire anymore according to the time, but we are acquiring according to the angular position of the trigger. So, we must change the axial dimension of the acquisition from the time t, to the angles α. The acquisition time will be swapped from T to bigger α which is written as A. Of course, also the frequency resolution will change, as the following plots state:
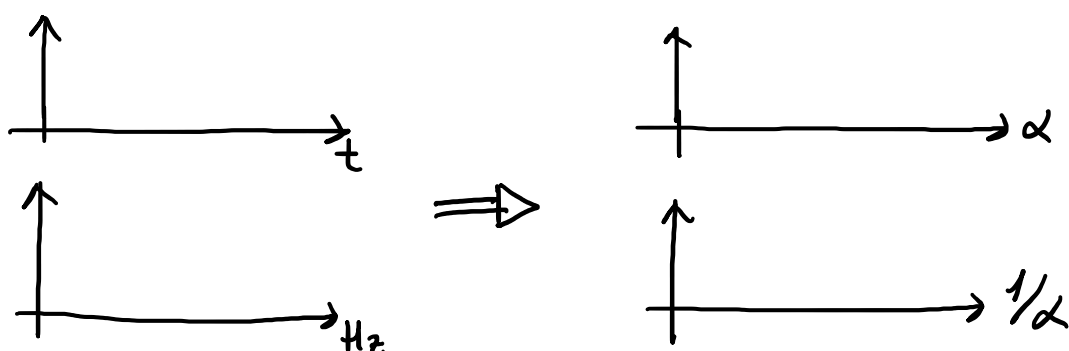
The configuration described until now is simply called synchronous acquisition since it goes together with cycle. Let's now try to understand better what the main advantage of this configuration is. With only one sensor (phase reference, only) we could still have leakage whereas with two sensors (phase reference and clock reference) we have the best acquisition possible because we don't have leakage anymore. So, the benefit introduce by this is that I reduce leakage and I know that physically all the average that we are making are referring to the same place, and this for sure is another improvement. Of course, when we do this synchronous acquisition, we cannot apply the window because it's absolutely useless since we don't have any leakage at all.

Once we have the signal, we can refer to a specific point of the rotor. If we know where the signal start, we also know what the angle between the reference point and the mass that is unbalancing the rotor is. If we know where it is, we can put a counter balance mass on the opposite side. So, the position in which we put this mass depends of where it occurs whereas the value of the mass depends on the amplitude of the signal.

In case of synchronous acquisition if there is a lower frequency with respect to one revolution of the rotor, we observe leakage; so, in case of a rotated machine we have leakage it could be a symptom of this fact. What is the highest frequency I can get from this analysis? It will depend on how many samples I have inside my cycle. The higher the number of teeth, the better the resolution. But the higher number of teeth, the faster the conversion must go. We have some technological issues that we must take into account.

With this technique of acquisition, we remove most of the limitation of the FFT. Actually, in real world we don't use teeth wheel, but we use encoder that are based on optical properties and they are cheaper.

Power spectral density

If the frequency resolution Δf is small, we have a high acquisition time T and the energy is spread out to a bigger number of bands. So, many times we could have very different spectrum depending on the choice of the acquisition time T. In order to avoid this, the ratio between the autospectrum S_{AA} of the signal and the frequency resolution is computed. The autospectrum S_{AA} is an indicator of the energy of the signal and it's usually used to compute other quantities. The autospectrum can be computed as the product between a complex vector and the conjugate of the complex vector (we have already said that each acquisition of the signal is a complex vector). So, the ratio between the autospectrum S_{AA} of the signal and the frequency resolution defines the density of energy:

$$PSD = \frac{S_{AA}}{\Delta f} = \frac{A \cdot A^*}{\Delta f}$$

This process of normalising the spectrum with respect to the frequency resolution is called power spectral density (PSD) and it's usually done in order to obtain results independent from the frequency resolution Δf. By doing this, all the acquisition with different resolutions will be with the same modulus. Usually this technique is used when the spectrum is continuous.

Laboratory: computing the autospectrum as the product between the vector and its conjugate it's like computing the absolute value of the signal, let's try to understand why:

$$A = a + ib$$

$$A^* = a - ib$$

$$S_{AA} = A \cdot A^* = (a + ib) \cdot (a - ib) = a^2 + b^2 = \left(\sqrt{a^2 + b^2}\right)^2 = |z|^2$$