Laboratory of Physics III

Experience 6:
Sampling Theory

University of Trento, Physics Department

Authors: Matteo Zortea, Elena Acinapura
Date: December 2020
Group: A05
Experience number: 6

Contents

1 Introduction

While quantities in the natural world vary with continuity in time and space, when we perform a measure we always extract discrete values of the quantity of interest - the so called *samples*. Therefore, we need a method to establish a relationship between discrete samples and the underlying continuous function, with the goal of reconstructing a function that is as close as possible to the original one.

In this experience we will firstly show how to measure discrete samples of an electric signal varying in time. One of the circuits that can perform such task is the *sample and hold*. Sample and hold circuits find many useful applications: for example nowadays computers work with a digital logic and, in order to translate an analogical signal coming from the real world to a digital one, one needs to know signal's value at discrete times. A sample and hold circuit provides then a way to keep the signal fixed for the time necessary to complete the analog-digital conversion, assuring that the entire process is performed on the same signal value. Another examples of application is a measure of a signal that varies too rapidly to be observed in a precise way, and, again, a sample and hold can be used to keep the signal fixed for the desired amount of time, once triggered.

Secondly, we will show how to apply the sampling theory of Nyquist and Shannon to reconstruct a signal from the samples, and we will compare the reconstructed signal to the original one. We will also show that approaching the problem of signal reconstruction with an intuitive and naive method fails in obtaining the original signal, and therefore that the theory of Nyquist and Shannon is necessary, brilliant and powerful.

2 Frequency Setting

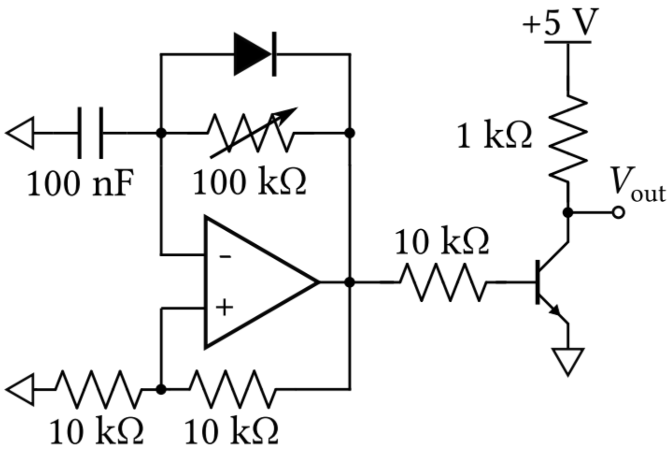

Figure 1: The above circuit was used to generate a square wave used to trigger the sample and hold and consists of a combination of a relaxation oscillator and a 0 V − 5 V switch.

The first circuit we implemented was the one reported in figure 1. The purpose of this circuit is to provide a stable square waveform which can be used to set the sampling frequency of the sample and hold device.

2.1 Circuit analysis

The left part of the circuit is a relaxation oscillator, while the right part, the one with the BJT transistor, can be regarded as a $0\,\mathrm{V} - 5\,\mathrm{V}$ switch.
Here's a brief recap of how a relaxation circuit work. We assume the initial conditions to be

$$V_x \simeq V_{sat} \qquad V_+ = \frac{V_x}{2} \simeq \frac{V_{sat}}{2} \qquad V_- \simeq 0$$

This is justified because a small voltage on the non-inverting input of the op-amp produces a huge V_x (limited by saturation), and by applying the voltage divider formula on the non-inverting loop one can obtain the result for V_+. Since the time taken to propagate the signal in the above process is small compared to typical capacitor charging time ($\tau \simeq RC \simeq 640\,\mathrm{ms}$) the assumption for V_- is justified too.
As V_- increases the voltage difference between the non-inverting and inverting inputs of the op-amp is reduced, and when $V_- > V_+$, the output voltage's sign changes according to $V_x = A\,(V_+ - V_-)$. In the same manner one can describe the negative output phase, with the capacitor discharge taking part this time. The role of the diode here is to lower the global inverting loop resistance during the negative output phase (the dynamic resistance of the diode decreases rapidly as the voltage across it increases), hence reducing the discharge time and the time that the oscillator spends in this state. The frequency of oscillation is determined by the value of τ and of the resistors in the non inverting-loop. As is visible in figure 1, we used a variable resistor to adjust the frequency at our intended value.

When the output of the oscillator is high, that is $V_x \simeq V_{sat} \simeq V_{cc}$, the base-emitter junction is in direct bias, the BJT is in active state and current flows through the transistor. The greater the base-emitter potential difference, the greater the current flowing through the transistor, the greater the potential drop between the $5\,\mathrm{V}$ supply and the collector. One can calculate that if $V_x = V_{cc}$, the current will be high enough to bring the transistor in saturation state, which means $V_{out} = V_E = 0\,\mathrm{V}$. On the opposite, when the output of the oscillator is low, the base-emitter junction is in reverse bias, hence the transistor is in interdiction state, no current flows through it and so there's no voltage drop between the $5\,\mathrm{V}$ and the collector: $V_{out} = 5\,\mathrm{V}$. Since the oscillator spends less time in the negative output phase, as previously stated, we can expect V_{out} to be high for just a small fraction of the total period. The ratio between this time and the whole period is known as *duty cycle*.

We set the frequency of oscillation to be $\simeq 1\,\mathrm{kHz}$. We say "$\simeq$" because the measured frequency resulted to be very unstable and dependent from the environmental conditions: touching the op-amp was enough to cause a change in frequency of up to 0.1%. So we adjusted the frequency to be1 kHz within 3 significant digits. We then measured the period of the square wave and the duration of the "high" part of it to estimate the duty cycle; we obtained

$$\text{Duty cycle} \simeq 3.5\% \tag{1}$$

3 Sample and Hold

We then used an integrated circuit, the LF398, which is called *sample and hold* circuit and whose internal electric diagram can be synthesized as in figure 2. We used this circuit to store the signal's value at discrete times and we then tried to reconstruct the original signal via software starting from these discrete values.

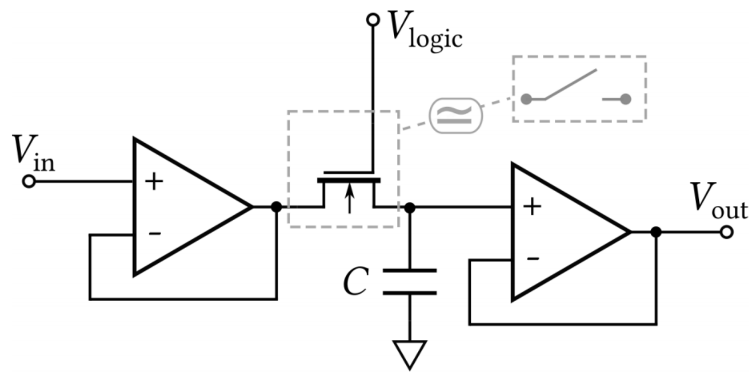

Figure 2: Sample and hold circuit diagram

The input signal is connected to the first operational amplifier, on the left, which works as a buffer to mask the output impedance of any previous stages of the circuit. Then a switch is used to decide whether the signal should be "read" or not. If the switch is open, the signal doesn't reach the capacitor, while, if the switch closes, the signal propagates to the capacitor. The capacitor, after being charged with the signal when the switch is closed, only sees the rightmost buffer, whose input impedance is ideally infinite, and hence remains charged until the switch closes again.
The switch is controlled by a clock generated by the relaxation oscillator; we need the duty cycle of the clock to be as short as possible, because the signal that remains "trapped" on the capacitor is the last value registered before reopening the switch, and we want it to be as close as possible to the value registered at the opening of the switch. In other words, we want the switch to open and close very quickly, so that it takes "a snapshot" of the input signal.
The result is represented graphically in figure 2.

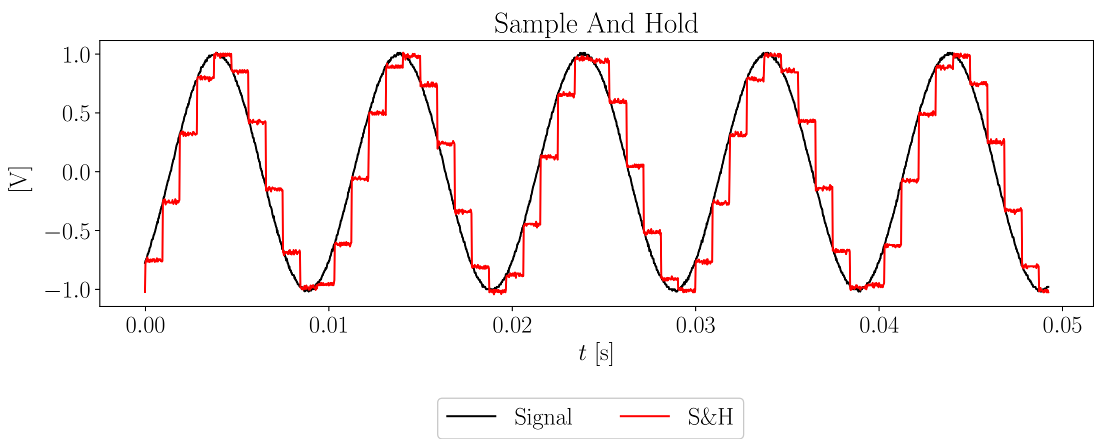

Figure 3: The output of the Sample and hold (S&H) is a square wave that "traps" and keeps the value of the input signal at every rising edge. By reading the value when the S&H is horizontal, one can obtain samples of the input signal, with a frequency equal to the clock's one.

We then tried to match the sampling frequency generated by the relaxation oscillator and the frequency of the input sine wave. By looking at the samples, one could see a constant value in time (in first approximation) whose amplitude was equal to the one of

the sine wave. This can be explained in the following way: by setting the two frequencies at the same value, we have that the sampling period is exactly equal to the sine period, hence we are always sampling the input wave at the same point and it seems that we measure a constant value.
Even though the two frequencies were matched carefully that was an unstable condition: we could observe the above described phenomenon only for a few seconds because of external fluctuations (especially due to thermal instability). We repeated the frequency matching at intervals of one minute for 6 times, and saw that the fluctuations had a magnitude of about 0.1% of the frequency.

4 Signal reconstruction

The major part of this experience was to try to reconstruct the original signal by means of the samples collected with the previous circuits and to face the problem of *aliasing*. We used as input function firstly sinusoidal waves at 50 Hz, 100 Hz, 200 Hz and 900 Hz and then triangular waves at the same frequencies; we used the sample and hold circuit to sample those waves, using the set frequency of about 1 kHz.

4.1 Finding the samples

We collected time series for the input and output of the sample and hold circuit and from that we obtained the samples of the input signal by reading the value when the output is "flat". This happens at a frequency given by the clock's one, so one might think that we could just take a point in the first flat, and then move to the other by "jumping" one period forward. However, as explained before, the frequency of the clock appeared to be very unstable, so that we had an uncertainty of about 0.1% in the frequency. That uncertainty was big enough to compromise the search for the flat intervals, because a change of 0.1% in the frequency often resulted in missing out one of the flat intervals, especially at higher frequencies. Finding the right frequency by trial and error wasn't feasible because of the big number of time series, so we opted for another method.
We noticed that the oscilloscope recorded a fixed number of points in every interval, call it K, and hardly any point was taken at the rising and falling edges. So we just found the first point of the first flat interval, read it value, jumped K points forward and there we were on the next flat. We included some control to be sure that we didn't find a point on a rising or falling edge, and that we didn't miss a flat, but that's just boring code.
Lastly, because we need to know the sampling frequency in order to reconstruct the signal via the sampling theory, we calculated it in the following way: let ΔT_i be the time interval between the recognized samples i and $i+1$; we estimated the sampling period T_s as the average of the ΔT_i over the $N-1$ little intervals (where N indicates the number of samples):

$$T_s = \frac{\sum_{i=1}^{N-1} \Delta T_i}{N-1} \tag{2}$$

and then simply $f_s = 1/T_s$.
An example of the results of the sample recognition is represented in figure 4

4.2 Reconstruction and aliasing

After having extracted the samples of the original signal, we moved on to per perform the reconstruction of the function, obtaining a function $r(t)$. Let us call the samples $V(nT_s)$,

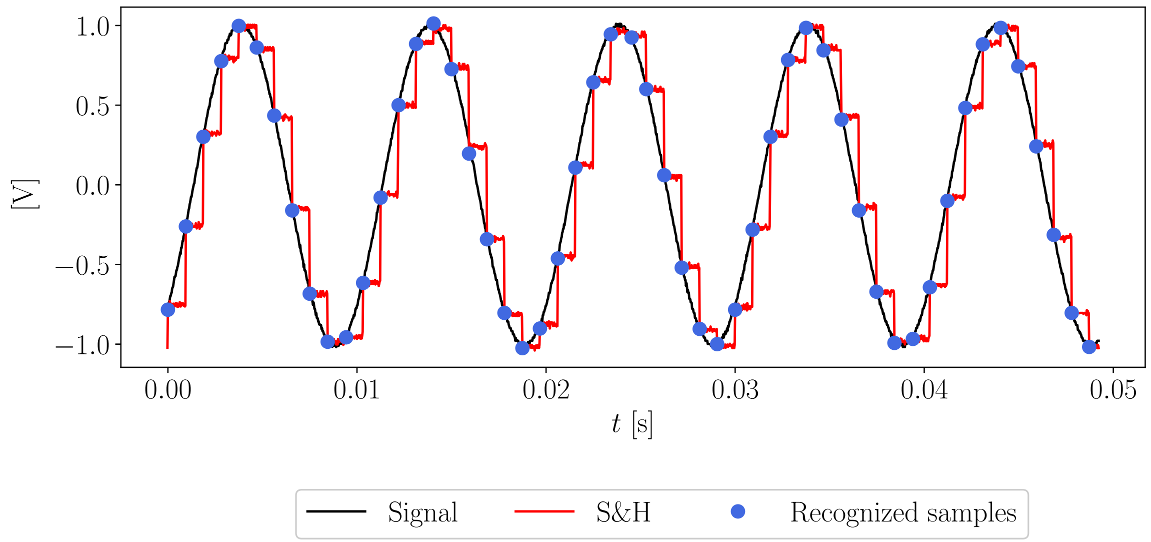

Figure 4: Recognition of the samples for one of the signals. As is visible, our code correctly recognizes every "flat" interval of the S&H output.

with $n \in \mathbf{Z}$. We then calculated $r(t)$ according to the Nyquist and Shannon theorem, that is:

$$r(t) = \sum_{n=0}^{N} V(nT_s) \text{ sinc} \left[\frac{\pi}{T_s} (t - nT_s) \right] \tag{3}$$

where N is the number of samples that we had. Theoretically, the summation should be extended to every $n \in \mathbf{Z}$, but obviously had a limited number of samples. Figures 5 and 6 compare the aliasing signal and the original signal for the sinusoidal and triangular waves at different frequencies.

4.3 Analysis via the Nyquist and Shannon theorem

The Nyquist and Shannon theorem states the following: consider a function $V(t)$ that has a Fourier spectrum that is $\frac{\pi}{T}$ band-limited for a certain T, i.e.

$$\widetilde{V}(\omega) = 0 \quad \text{if} \quad |\omega| > \frac{\pi}{T} \tag{4}$$

Then, if we sample the signal $V(t)$ with a sampling period $T_s = T$, the following equality holds:

$$V(t) = \sum_{n \in \mathbf{Z}} V(nT_s) \text{ sinc} \left[\frac{\pi}{T_s} (t - nT_s) \right] \tag{5}$$

Otherwise stated, if the spectrum of the function $V(t)$ is nonzero only for frequencies less that a certain $f_0 = \omega_0/2\pi$, then we are able to reconstruct the actual original function $V(t)$ with formula 5 only if we choose a sampling frequency f_s such that

$$f_s = \frac{1}{T_s} > \frac{\omega_0}{\pi} = 2f_0 \tag{6}$$

On the contrary, if we sample that function with $f_s < 2f_0$ and we try to reconstruct the function with formula 5, we will reconstruct a signal different from the sampled one: this phenomenon is known as *aliasing*.

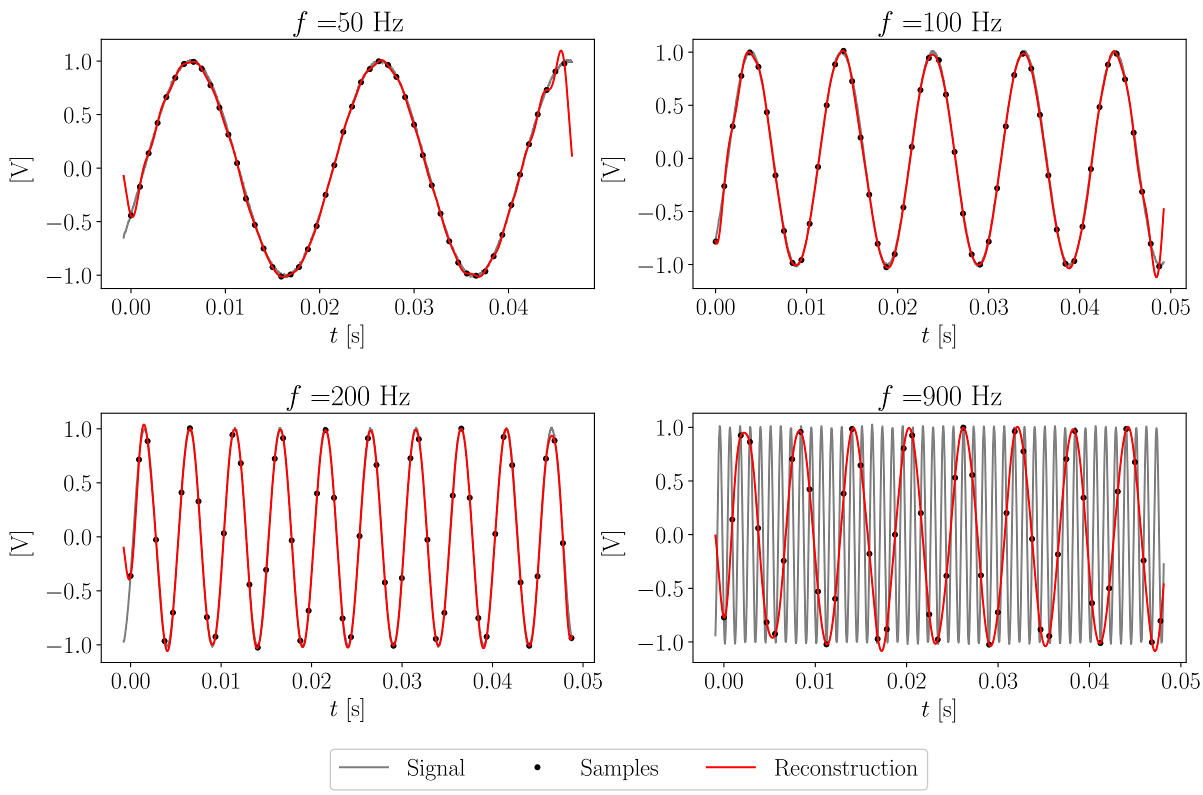

Figure 5: Reconstruction of sinusoidal waves at 50 Hz, 100 Hz, 200 Hz and 900 Hz.

In our experience we used both sinusoidal and triangular signals. A sinusoidal wave with frequency f_0 has a spectrum that is a Dirac-δ at $\omega = \pm 2\pi f_0$. Therefore, we will be able to reconstruct it properly only if we choose a sampling frequency $f_s > 2f_0$. We used $f_s \simeq 1\,\text{kHz}$, so we expect to be able to obtain the original signal of the last has a frequency up to 500 Hz. Indeed, we can see in figure 5 that the reconstruction process returns the "right" function only for sinusoidal waves with frequencies 50 Hz, 100 Hz and 200 Hz, while the reconstruction of the 900 Hz sine function returns a sine with a different frequency. Say f_a the frequency of the reconstructed signal, by means of the sampling theory one can predict that, if $f_0 < f_s < 2f_0$, as it is the case for $f_0 = 900\,\text{Hz}$ and $f_s = 1\,\text{kHz}$, the following relation holds

$$f_0 < f_s < 2f_0 \implies f_a = f_s - f_0 \tag{7}$$

We checked if the reconstructed signal for the 900 Hz sine function followed this relationship by finding the frequency f_a with a fit. We obtained that $f_a \simeq 98\,\text{Hz}$, which is in good agreement with the expected value of $1\,\text{kHz} - 900\,\text{Hz} = 100\,\text{Hz}$.

What about triangular waves? Well, a triangular wave with frequency f_0 can be written as a sum of sines with frequencies that are multiples of f_0:

$$V(t) = \frac{8}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{(2k-1)^2} \sin\left[(2k-1)2\pi f_0 t\right] \tag{8}$$

that means that its spectrum is not even bounded, since the series continues to infinity! However, the harmonics that really "make up the shape" of the triangular wave are the

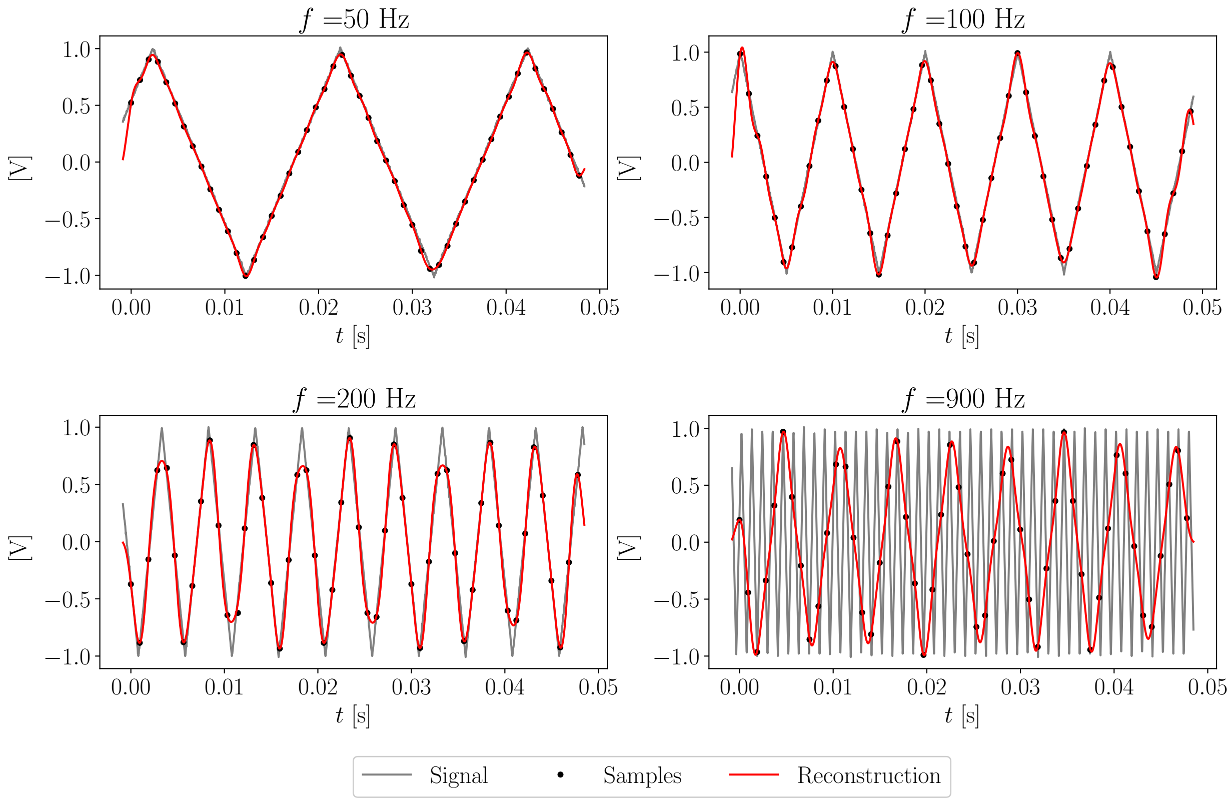

Figure 6: Reconstruction of triangular waves at 50 Hz, 100 Hz, 200 Hz and 900 Hz.

first ones, since the coefficient $1/(2k-1)^2$ rapidly decreases with k. For the sake of simplicity, we can consider the main frequencies of the triangular wave to be f_0, $3f_0$ and $5f_0$.
When we sample a triangular wave at 50 Hz, we can consider that we are sampling mainly the sum of three sines at frequencies 50 Hz, 150 Hz and 250 Hz. If we use a sampling frequency $f_s = 1\,\text{kHz}$, all the main harmonics are less than $f_s/2$, so we expect to be able to reconstruct the triangular wave quite correctly. This, in fact, is verified in figure 6. If, instead, the triangular wave has a frequency of 100 Hz, it mainly consists of three sines with frequencies 100 Hz, 300 Hz and 500 Hz. Hence, only the first two harmonic frequencies fall below $f_s/2$, while the third one is exactly $f_s/2$ (which, we remind, is not a "good frequency" for the signal reconstruction). And indeed, we see in figure 6 that the reconstructed function is similar to the original one, but starts to show some imperfections. Those imperfections are amplified when we sample the triangular wave at 200 Hz, for which only the fundamental harmonic falls below $f_s/2$. Lastly, the reconstructed function is completely different (*alias*) from the original one when we try to sample the triangular wave at 900 Hz, for which not even the fundamental harmonic is above $f_s/2$.

4.4 Reconstruction with linear extrapolation (bad idea!)

In the previous paragraph, the function sinc $\left[\frac{\pi}{T_s}x\right]$ we used to perform the reconstruction with $x = t - nT_s$, is generally called *kernel*. As an experiment, we tried to reconstruct the original function by using a different kernel based on linear extrapolation, which means that we simply connect two consecutive sampling points with a straight line. This means using the following kernel

$$k(t) = \begin{cases} \dfrac{T_s - |x|}{T_s} & \text{if } |x| \leq T_s \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The result is represented in figures 7 and 8.

As we can see (and expect), this method works good enough only when the function

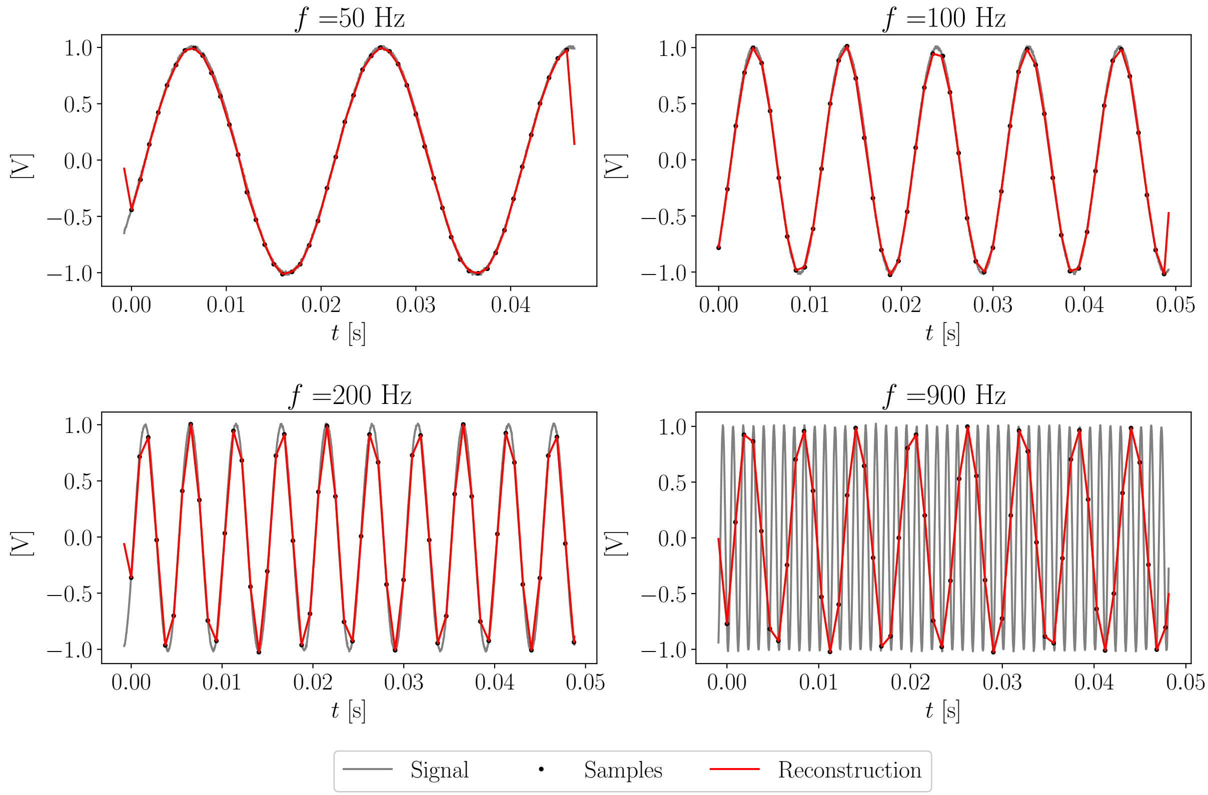

Figure 7: Signal reconstruction using a *linear*-kernel

is approximated good by a straight line between two sampling points. On the contrary, it fails in the neighborhood of the extrema, when the slope of the signal changes more rapidly for the sine, and suddenly for the triangular wave. The main problem with linear extrapolation is that it isn't able to reconstruct a sine even when its frequency is about 1/10 of the sampling frequency! (Look for instance at the plot for $f = 100\,\text{Hz}$ in figure 7). Because a lot of functions[1] can be written as a sum or integral of sinusoids, the fact that linear extrapolation easily fails to reconstruct a sine is a very serious problem.

[1]At least, the most "perbene"

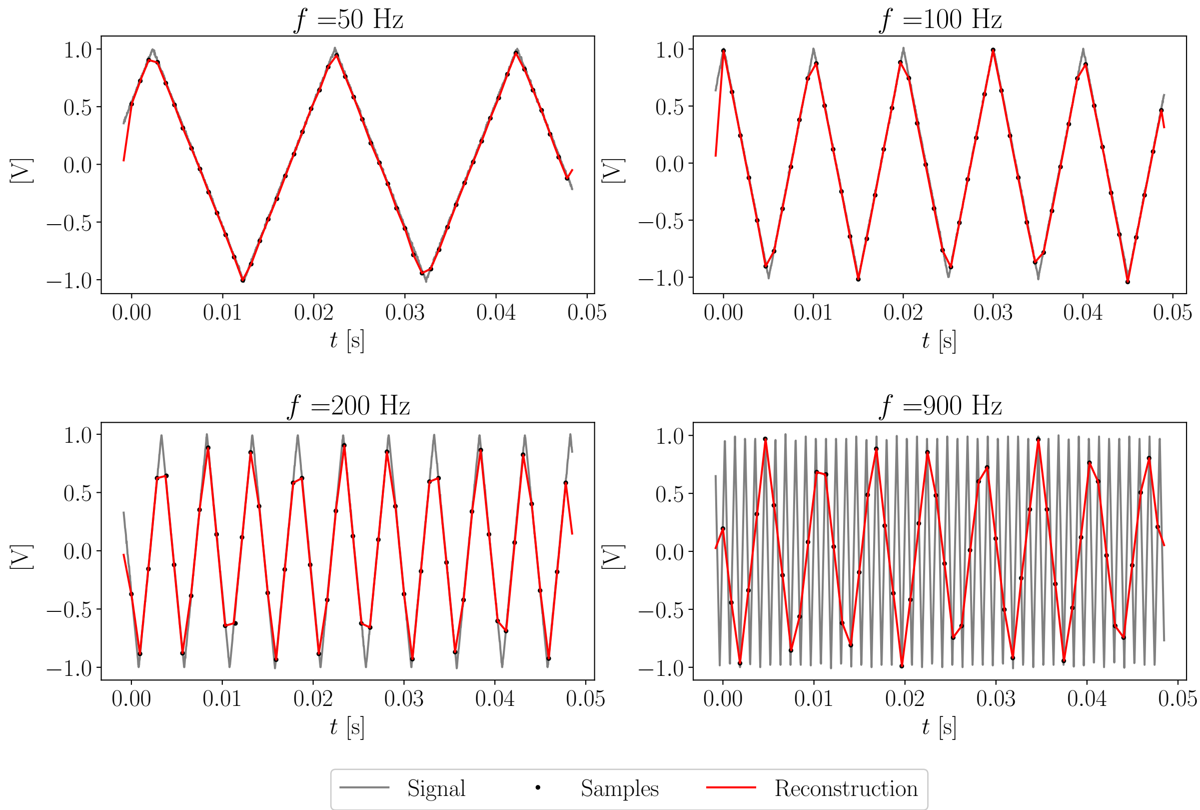

Figure 8: Signal reconstruction using a *linear*-kernel

5 Conclusions

In this experience we first implemented a clock generator that was an application of the relaxation oscillator, a circuit studied in the previous experience. The oscillator's output was used to trigger a sample and hold, a device that can sample signals given as inputs.By using this setup we could study the oscillator frequency's stability, a parameter strictly dependent on external conditions such temperature.
In the second part of the experience we used the $S\&H$ to sample an external waveform by varying its frequency and studied the relationship between the signal frequency and sampling's frequency. We encountered the aliasing phenomenon, which was studied by using both a *sinc* kernel and a *linear* kernel, obtaining that the former is generally a better choice in therms of accuracy at a given sampling frequency.

Appendix

We report a picture of the implementation of the used circuit.

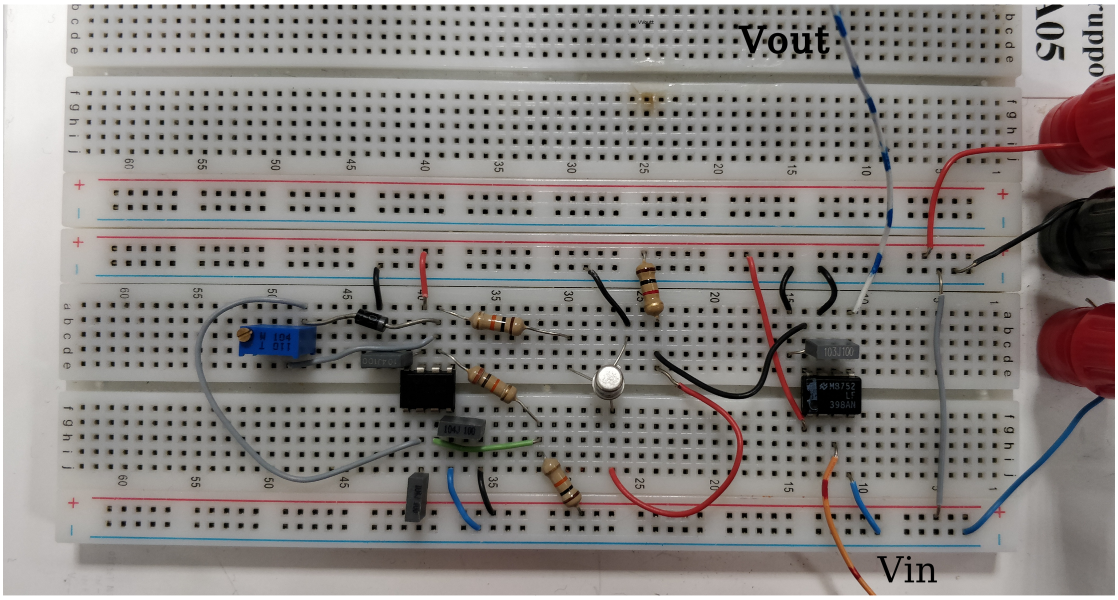

Figure 9: Implementation of the sampling circuit. Starting form the left, there are the relaxation oscillator, then the BJT switch and lastly the sample and hold. The white-and-blue striped wire going upward is the output signal, while the red-and-orange wire going downward is the input signal.