Version N° 1.0

LaurTec

Understanding Undersampling Technique

Author : *Mauro Laurenti*

email : info.laurtec@gmail.com

ID: AN4002-EN

License

The usage of any project, tutorial, software given by Mauro Laurenti (the Author) imply the acceptance of the following license.

The projects, tutorials, software (material) supplied herewith by Mauro Laurenti is intended for use solely and exclusively for personal use.

All the material is owned by the Author, and is protected under applicable copyright laws.

All rights are reserved.

Any use in violation of the foregoing restrictions may subject the user to criminal sanctions under applicable laws (Italian or International ones), as well as to civil liability for the breach of the terms and conditions of this license.

Commercial use is forbidden without a written acknowledgment with the Author.

Personal or educational use is allowed if the application containing any part of the given material doesn't aim to commercial use or monetary gain of any kind.

ALL THE MATERIAL IS PROVIDED IN AN "AS IS" CONDITION. NO WARRANTIES, WHETHER EXPRESS, IMPLIED OR STATUTORY, INCLUDING, BUT NOT LIMITED TO, IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE APPLY TO THIS MATERIAL. THE AUTHOR SHALL NOT, IN ANY CIRCUMSTANCES, BE LIABLE FOR SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES, FOR ANY REASON WHATSOEVER.

Introduction

All engineers know the sampling theory and the fact that you need to sample the signal at least at twice the maximum frequency contained within the signal of interest. By the way merely applying the Nyquist criteria or Shannon theory, can lead to design more expensive system than actually required. The article shows how to properly interpret Nyquist criteria and Shannon theory, thus making you aware of the undersampling technique as possible solution to decrease the system costs.

Sampling Theory

Converting an analog signal to a digital one requires an analog to digital conversion. This process is frequently described in the literature among the sampling theory chapters. Sampling a signal means converting it from continues time to discrete time domain. While this is where we are going to focus in this paragraph, converting a signal to a digital one, also requires converting the amplitude within a finite number of values. Out of the discrete values is made the numerical sequence that represents the analog signal in the digital domain. The amplitude discretization is performed by the ADC (Analog to Digital Converter), which normally also embeds the circuitry required to get the discrete time sequence. Let's take a closer look to the first process, sampling the signal; Figure 1 shows that process.

The signal of interest s(t), is a continues time signal, with a certain amplitude A(t). The amplitude is a real number, thus between two values A(x) and A(y) there are infinite number of values. The function s[n] is a continuous time function, which exist just in specific times, which are the signal samples. The time interval among the samples is a fixed time, determined by the sampling function δ(t). This function represents the driving signal that close the switch for a very short time, ideally an instant, and let the signal s(t) pass through. Closing a switch is a simplified equivalent model to represent the multiplication among s(t) and δ(t). So, summarizing, the sampling process is the process that let us get the s[n] function just closing the switch at fixed interval and measuring the value of s(t) at that time.

Observing the functions s(t) and s[n], it's possible to understand that s[n] has less information than the original function s(t) . So just applying this understanding it's easy to see that the more samples we take, more information we keep and vice versa. Reducing the sampling rate let us reduce the difficulties to design high frequency circuits and also have the advantage of requiring less data memory. By the way reducing the sample rate below a certain point creates a lost of information that does not allow us to go back from the s[n] to s(t). The limit below which, we should not sample are

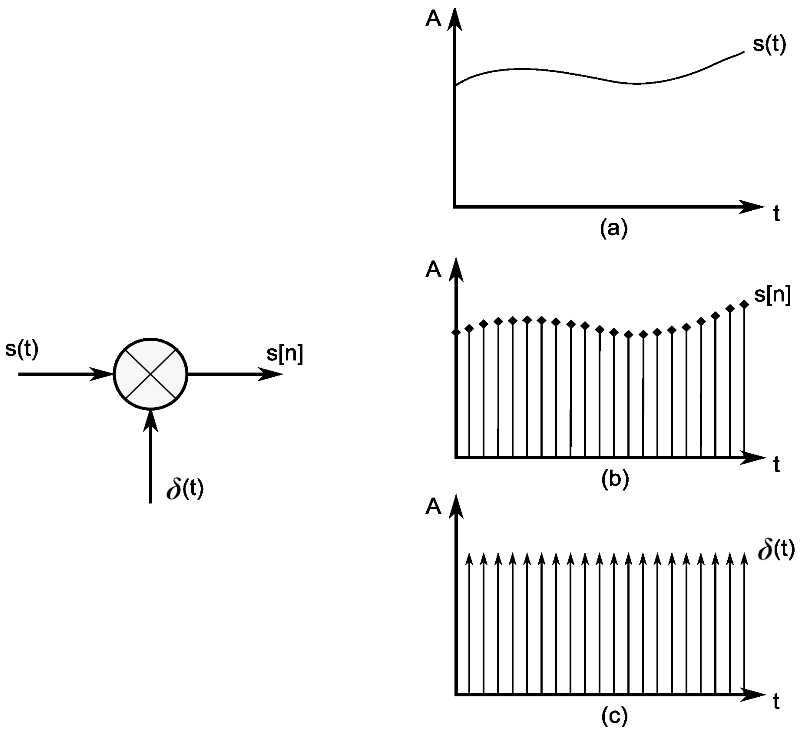

Figure 1: *The continuous time domain signal s(t) is sampled through a δ(t) function to get the discrete time domain sequence s[n]. The amplitude discretization of s[n] values represents the numerical sequence that identifies the digital signal.*

described within the Nyquist criteria and Shannon theory.
Nyquist criteria, states that, to convert a signal from s(t) to s[n], the minimum sampling frequency, to avoid lost of information, should be at least twice the maximum frequency fmax contained in the signal:

$$f_c \geqslant 2 f_{max}$$

This constrain considers that the signal is a baseband signal (located around zero) and it is also bandwidth limited. Thus, we could also state that the minimum sampling frequency should be twice the bandwidth of the signal. Nevertheless, it is the maximum frequency that is taken in to account to determine the sampling frequency.

Shannon theory is just a generalization of Nyquist criteria and it considers the case in which the signal leis anywhere in the spectrum. So, considering a signal of interest s(t) with a limited bandwidth B=fmax-fmin, the minimum sampling frequency to avoid lost of information should be twice the bandwidth, independently of the maximum frequency in the signal.

$$f_c \geqslant 2\mathrm{B} = f_{max} - f_{min}$$

This consideration as it is stated, could lead to a much smaller sampling frequency in case the signal does not lie around zero. Let's consider an example on the frequency domain, on which we sample a base band signal s(t) of bandwidth B, as shown in Figure 2.

Notice the symmetry of the base band signal, around zero, this symmetry indirectly states that we are treating real signals, as could be any signal we measure on the lab bench. As shown in Figure 2 (a), sampling a signal means creating copies of it around the sampling frequency and its harmonics. In case (a), fs has been chosen to be higher than twice the signal bandwidth, creating replications which do not interfere with the baseband signal. In case (b), the sampling frequency fs has been chosen smaller than twice the bandwidth. What is now happening is that the left replica is overlapping the baseband signal. This overlapping, known as aliasing problem, creates distortion since it is not possible to separate the original signal from the sampled one. This is the reason why the sequence s[n] does not represent s[t] any longer (lost of information), while case (a) does let us get s(t) out of the sequence s[n].

Well, at this point, some one could say, I can filter out the signal from the right side of case (b) and still get the original signal information. This is not possible because during the sampling process, what is actually happening is that the signal gets copied around the sampling frequency and its harmonics. So the actual spectrum would have been as shown in Figure 3.

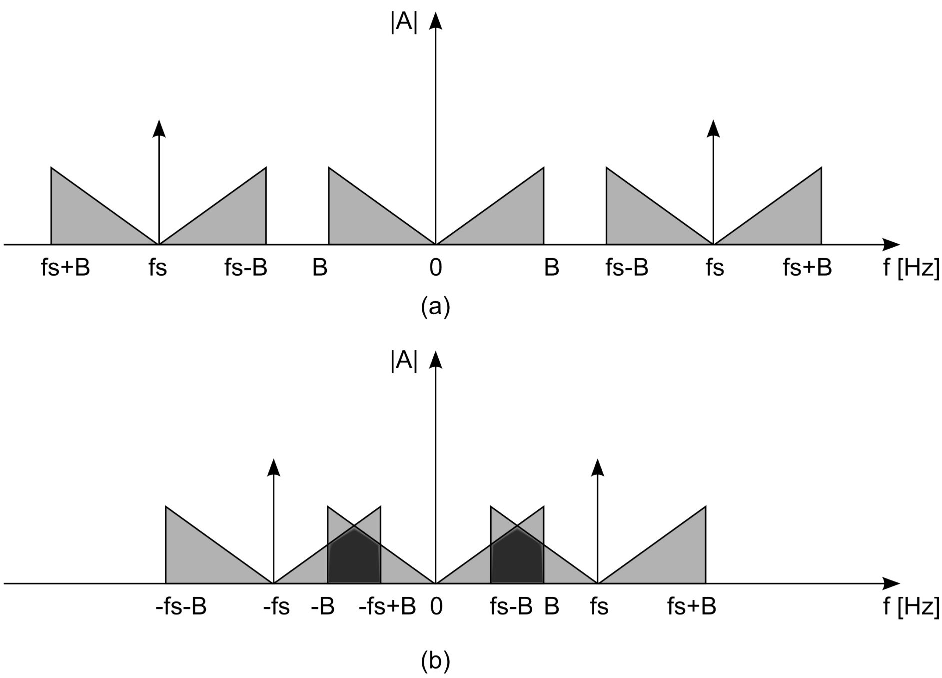

Figure 2: *Signal replicas around the sampling frequency. Figure a) shows the case with fs>B while Figure b) shows the case with fs<B and the aliasing problems associated with it.*

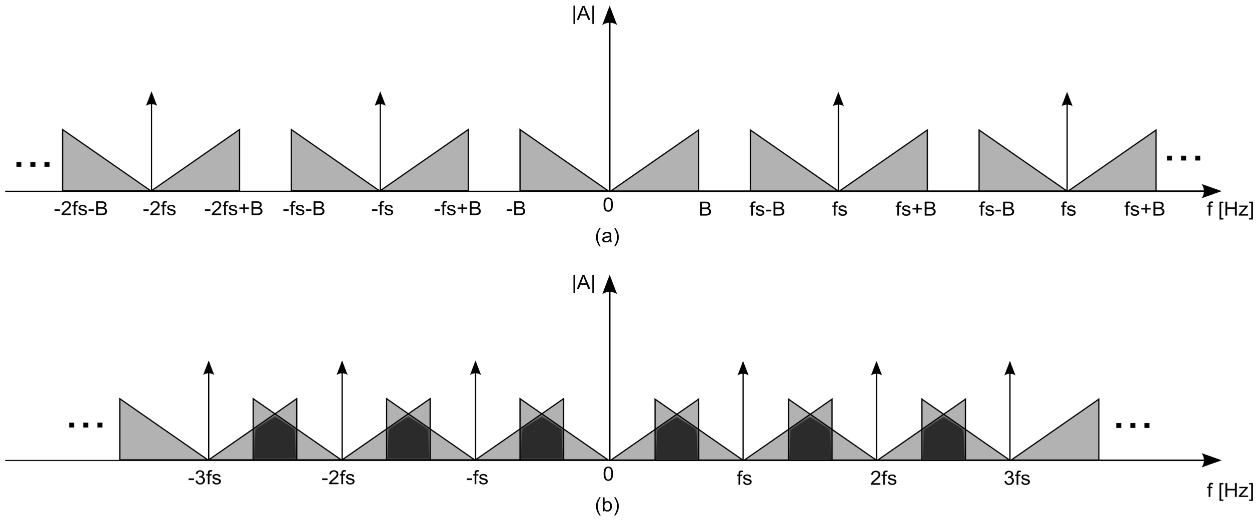

Figure 3: *Signal replicas around the sampling frequency harmonics. Figure a) shows the case with fs>B while Figure b) shows the case with fs<B and the aliasing problems associated with it.*

From this new perspective it is possible to see that all the replicas get actually distorted by the overlapping.

Analyzing Figure 3 (a) is possible to understand the importance of having a filter before the sampling process, that get rid all the signals outside the band of interest. In case of baseband signals the filter should be a low pass filter, while for others, it should be a band pass filter.

If we do not filter the noise out of the band of interest, the sampling frequency harmonics will fold back noise far away from the signal of the interest, thus interfering with the base band signal. The importance of having a filter is also a key point of making the signal of interest limited in bandwidth. The filter design is a complex process and depending on the constrains related to the signal, sampling rate and last but not least system cost, it could required quite a lot of effort.

So, harmonic sampling could be a problem in case we have not properly filtered out the unwanted signals, by the way properly exploiting it could actually benefit our design; harmonic sampling leads us directly to the undersampling realm.

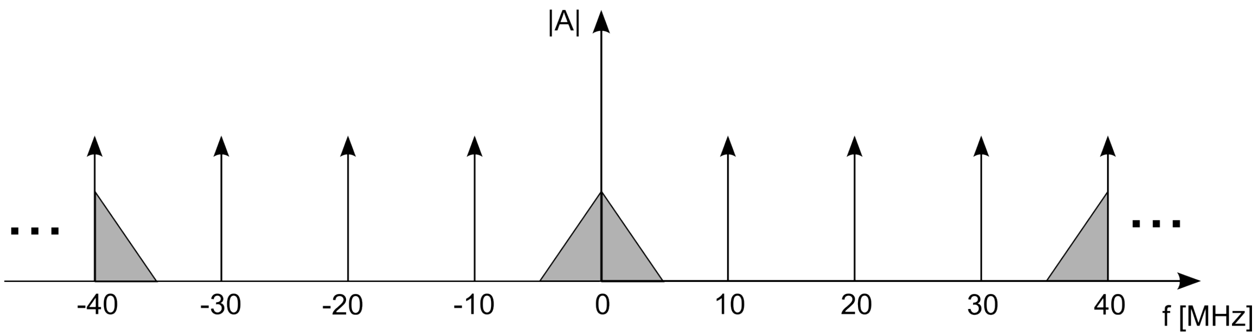

Figure 4: *Undersampling example with a video signal (5MHz Bandwidth) located between 35MHz and 45MHz. Just the replicas due to the 4th sampling harmonic at 40MHz are depicted.*

Undersampling technique

Let’s start with an example, which is in general worth more than thousand words. Consider a video signal with a bandwidth B of 5MHz, located between 35MHz and 40MHz, as shown in Figure 4.

As explained before, to avoid other signals to be folded back, it's required a band-pass filter surrounding our signal. As stated by Shannon theorem, the minimum sampling frequency should be 10MHz. If we pick this value (just as theoretical choice) we see that the fs sampling harmonics, in particular 4fs is creating a signal replica around zero, as a baseband signal. This means that the ADC will have at its input a baseband signal among the other replicas (not shown for clarity). The new baseband signal is an exact copy of the original one located between 35MHz and 40MHz, just the original frequency information get lost.

By the way this is not a problem since the signal location and sampling frequency are known a priori. Note that the signal that get translated around 0 has been inverted, this is, once again, not a problem because the DSP (Digital Signal Processor) or the FPGA (Field Programmable Gate Array), after computing the FFT (Fast Fourier transform) can invert the signal once again without too much mathematical effort. Nevertheless, depending on the sampling frequency and where we want the signal to be folded back, the frequency inversion could be avoided, but as mentioned is not a big deal. Choosing the sampling frequency is normally done considering many constrains, so the application drives what can be accepted and what should be avoided.

Observing with attention what has been just explained, should have shown you that undersampling has done what is normally made during a signal demodulation process. Demodulating a signal whatever nature it is, AM, FM, IQ, are often made mixing the input signal (channel of interest) with a local oscillator to get an intermediate frequency. This create copy of the signal at known frequency point making filtering and other signal processing to be easily done for all the channels of interest.

How to chose the right ADC

The example just shown, has demonstrated that the ADC could be used as a sort of mixer, doing at the same time the digital conversion and down-converting the signal. Even if this is a good opportunity to design a radio system without even using one mixer, it does not mean that any low sampling frequency ADC can actually do the job. Even if the sampling frequency that could be used is many times smaller than the highest frequency contained in the signal of interest, the ADC should have dynamic performances that must be excellent beyond the signal frequency. If this won't be met, the replicas we get will be highly distorted making the undersampling approach useless.

This is why most of ADC used for telecommunication, with a sampling rate in the range of 50-160 MHz, they have excellent dynamic performance much beyond the Nyquist bandwidth (fs/2).

At first glance, the major dynamic parameter to consider is the FPBW (Full Power Bandwidth), this parameter quantify the input bandwidth of the ADC and can give an understanding of the ADC working frequency. By the way the actual parameter used in telecommunication is the SFDR (Spurious Free Dynamic Range) which quantify how signals of different amplitudes presented at the same time in the input of the ADC, can be easily distinguished. This is a key parameter in the design of base stations, since in such design the front-end will deal with signals coming from cell phone located anywhere.

This situation could lead having signals that can be quite different in amplitude. For base station applications, depending on the technology used for transmitting the signal, the ADC should have SFDR in the range of 60-90dBFS (dB Full Scale). The SFDR is normally given on the datasheet on Tables format, for specific known application frequencies but it can also be found in a plotted fashion, versus input frequency and other key parameters.

Other parameters to keep in mind are the SNR (Signal Noise Ratio) and the SINAD (Signal to Noise and Distortion). SINAD is normally more important than the SNR, since, as source of noise, besides considering the internal noise, it also includes the Distortion due to non linearity behaviour. Both the SNR and the SINAD are shown on the datasheet the same way the SFDR is.

Last but not least the THD (Total Harmonic Distortion) can be quite important along with

H2 and H3 (Second and third harmonic) and the IMD (Intermodulation Distortion). These parameters are quite important in choosing the right ADC since can let the designer understand how much distortion is expected due to non linearity and due to adjacent radio channels.

While the ADC choice is very important, all the other components around the ADC must be also carefully chosen avoiding signal degradation. The front-end should have the right amplifier with low noise, low distortion and able to work at the frequency of interest. Clock distribution is also very important; the Jitter affecting the clock or phase noise, should not introduce too much noise, degrading the ADC performance; this is often an overlooked problem while working at high frequency.

Applications

After some theory let's take a look once again to the real world. Everyday we do use our cell phone and we get angry if the signal level is not good! The communication among users happen thanks to the base stations which are spread all over the countries. Their size could be as small as TV set, and as big as three refrigerators (the big one!). The size depends on the number of clients that each base station is supposed to handle; the more are the clients the bigger the base station is. By the way modularity approaches is used, so multiple base station can be used in parallel to serve more customers. Currently the technology is pushing to have the so called nano and pico cell which are supposed to be spread like seeds on the ground. Their size can be as small as a radio set

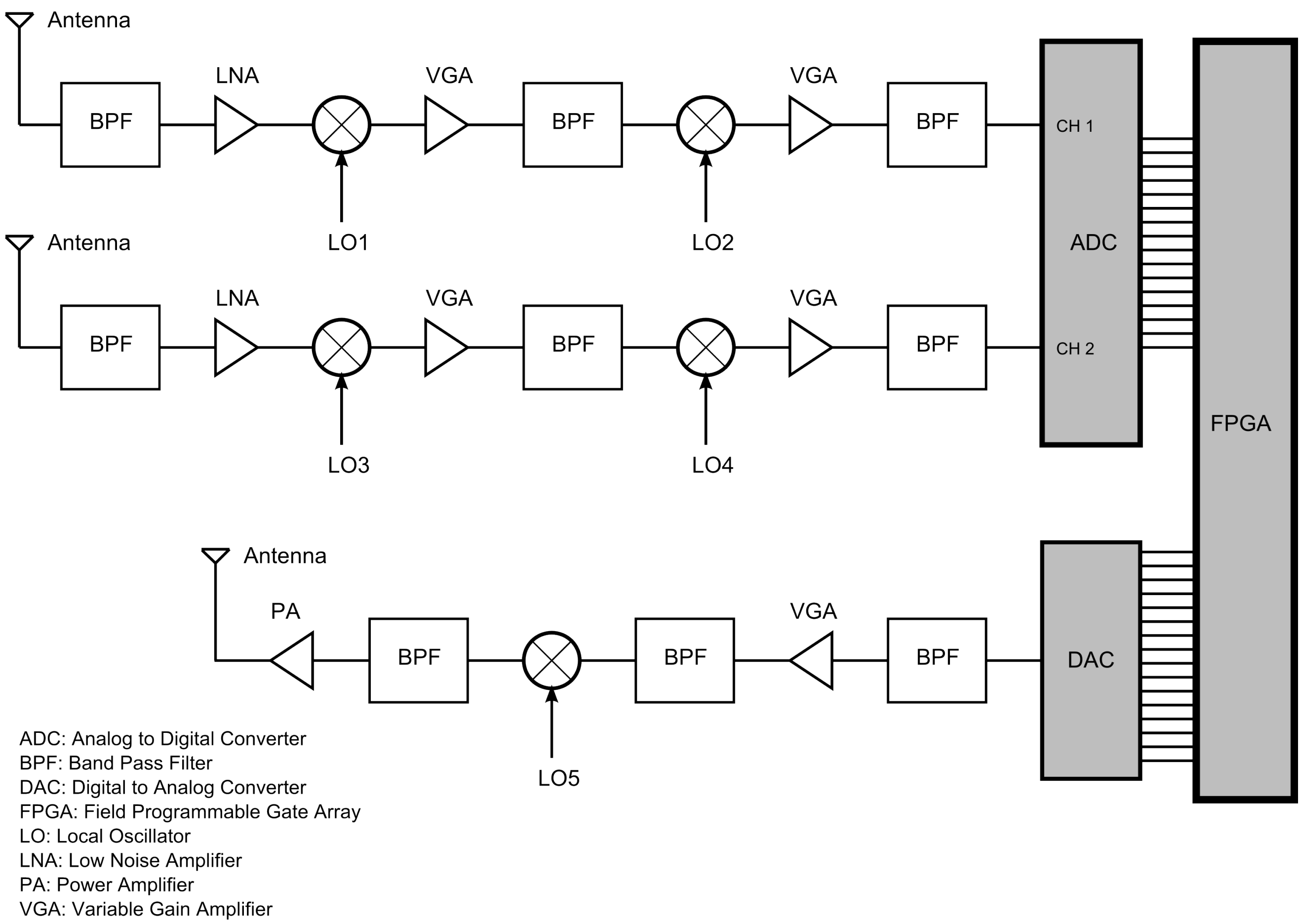

Figure 5: *Base station front-end block diagram, shoving the two input paths (primary channel and diversity channel) and the output path. Clock distribution, interfaces, signal control loops and power modules are not depicted.*

(the small one!) so potentially everyone could have a personal base station at home. It's easy to understand that the quality of service you could get is quite high, allowing downstream and upstream speed as never before. Furthermore having millions of pico cell let also the service provider to reach each area of the country without problems, letting the customer forget about the low level signal area!

By the way each new technology step has its debate and having pico cell at home has raised a new one about radiation problems, but we will not talk about this topic and the ideas to overcome the problem. Well, how it's possible to get base stations as small as a radio set? Decreasing the base station size and increasing at the same time the quality of the service that they are able to provide, it's possible thanks to the technology progress that is reducing components size and increasing integration. By the way also other approaches has made the reduction of the BOM (Bill of Material) possible, reducing, in general, both cost and size; using undersampling technique is one of it. In Figure 5 is shown the block diagram of a typical base station system, focusing the attention on the front end side. The system would be more complex if the clock distribution, interfaces, signal control loops and power modules would have been considered out.

It's possible to see that there are two receiver paths, one is the main and the other is the secondary one, also known as diversity channel; besides these, there is a transmitter path. The diversity channel is used to let the base station receives the signal from two different locations and reconstruct the signal using the information from both channels. This let the receiver increase the quality of the received signal, which is in general quite weak, since the transmitter are the cell phones. The transmission path is having just one antenna since there are less power constrains. After this brief overview we do know why the base stations normally have three antennas.

The LNA amplifier (Low Noise Amplifier) is used to increase the signal strength keeping the noise figure as low as possible. The VGA amplifiers (Variable Gain Amplifier) are used to amplify the signal with the right gain to get the maximum ADC performance. As probably known by other books, the ADCs have the maximum performance if the input signal is full scale amplitude. This means, the VGA should amplify the input signal to get a full scale signal; the gain which is required it depends by the signal strength, which is related to the distance and signal path between the cell phone and the base station. Normally more than one VGA is used to get more flexibility along the front end path. What just explained has actually done a simplification of considering just one user. The current ADC dynamic performances, besides undersampling also allow multi channels to be received at the same time, so getting the best of the ADC performance could be a challenging task; just consider the case of dealing with users that could be at the same time 100 feet from the base station and as far as half a mile. Dealing with both users and providing them a good service is quite challenging.

As you can see the receiver paths are using two mixers each, with different local oscillator, this is actually a simplified block diagram because many base stations do use even three mixers. Thanks to the undersampling technique and the high dynamic performance of high speed ADC, the third mixer can be easily removed and the second one could be also removed if high performance ADC is used. System specifications and constrains would anyway guide the system design. Another benefit of using undersampling is that the FPGA or the DSP, used to deal with the samples coming out from the ADC, can handle fewer samples, letting them compute multi channels (multi users) without problems. On the other hand, having removed one mixer and the circuitry around it, also allows reducing the power consumption.

Conclusion

A good understating of the theory is always required to make a good design. Exploiting undersampling technique let the designer reduce the system size and the power consumption making receivers such as base stations being cheaper and smaller. This will led in the next years having more base stations spread as seeds all over the countries, making the cell phone services better and cheaper.

Bibliography

[1] www.LaurTec.com : Electronic web page where you can find the most updated document version and further documents.