Mathematical basics of bandlimited sampling and aliasing

Modern applications often require that we sample analog signals, convert them to digital form, perform operations on them, and reconstruct them as analog signals. The important question is how to sample and reconstruct an analog signal while preserving the full information of the original.

By Vladimir Vitchev

To begin, we are concerned exclusively about bandlimited signals. The reasons are both mathematical and physical, as we discuss later. A signal is said to be bandlimited if the amplitude of its spectrum goes to zero for all frequencies beyond some threshold called the cutoff frequency. For one such signal (*g(t)* in Figure 1), the spectrum is zero for frequencies above a. In that case, the value a is also the bandwidth (BW) for this baseband signal. (The bandwidth of a baseband signal is defined only for positive frequencies because negative frequencies have no meaning in the physical world.)

The next step is to sample *g(t)*. We can express that operation in mathematical form by multiplying *g(t)* by a train of delta functions separated by the interval *T*. By multiplying *g(t)* with a delta function we select only the value of *g(t)* corresponding to the instant at which the delta function occurs; the product equals zero for all other times. This is analogous to sampling *g(t)* with a frequency $f_{SAMPLING} = 1/T$. This operation is expressed in Equation 1, and the new sampled signal is called *s(t)* (Eq. 1):

$$s(t) = g(t)\sum_{n=-\infty}^{\infty}\delta(t-nT)$$

The next step is to find the spectrum of the sampled signal *s(t)*. We do that by taking its Fourier transform (Eq. 2):

$$S(f) = F\{s(t)\} = \int_{-\infty}^{\infty} s(t)e^{-i2\pi ft}dt$$

Taking Fourier transforms

Evaluation of the above integral is cumbersome. To simplify it, we note that *s(t)* is the effective multiplication of *g(t)* with a train of impulses. We also note that multiplication in the time domain corresponds to convolution in the frequency domain. (For proof of that assertion, consult any text on Fourier transformations.) Thus, we can express *S(f)* as (Eq.3):

$$S(f) = G(f) * F\left\{\sum_{n=-\infty}^{\infty}\delta(t-nT)\right\}$$

Furthermore, note that the asterisk in Equation 3 denotes convolution, not multiplication. Since we know the spectrum of the original signal *G(f)*, we need find only the Fourier transform of the train of impulses. To do so we recognize that the train of impulses is a periodic function and can, therefore, be represented by a Fourier series. Consequently, we write (Eq. 4):

$$\sum_{n=-\infty}^{\infty}\delta(t-nT) = \sum_{n=-\infty}^{\infty} A_n e^{i2\pi\frac{n}{T}t}$$

where the Fourier coefficients are (Eq. 5)

$$A_n = \frac{1}{T}\int_{-T/2}^{T/2}\sum_{n=-\infty}^{\infty}\delta(t-nT)e^{-i2\pi\frac{n}{T}t}dt$$

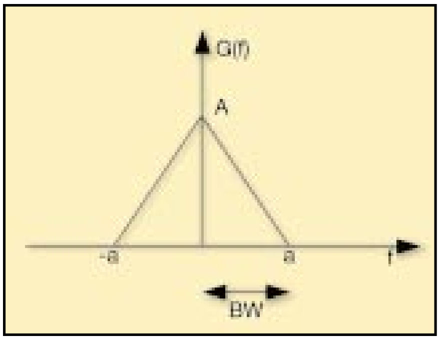

Figure 1. Frequency spectrum of the signal *g(t)*.

The limits of integration for Equation 5 are specified only for one period. That isn't a problem when dealing with the delta function, but to give rigor to the above expressions, note that substitutions can be made: The integral can be replaced with a Fourier integral from minus infinity to infinity, and the periodic train of delta functions can be replaced with a single delta function, which is the basis for the periodic signal. Thus, we can rewrite Equation 5 as (Eq. 6):

$$A_n = \frac{1}{T}\int_{-\infty}^{\infty}\delta(t)e^{-i2\pi\frac{n}{T}t}dt = \frac{1}{T}F\{\delta(t)\}\Big|_{f=\frac{n}{T}} = \frac{1}{T}$$

The train of delta functions then assumes the following simplified expression, which is easily Fourier transformable (Eq. 7):

$$\sum_{n=-\infty}^{\infty}\delta(t-nT) = \frac{1}{T}\sum_{n=-\infty}^{\infty} e^{i2\pi\frac{n}{T}t}$$

Recognizing that a signal can be synthesized from its Fourier transform (Eq. 8):

$$f(t) = \int_{-\infty}^{\infty} F(f)e^{i2\pi ft}df$$

and also that (Eq. 9):

$$F^{-1}\{\delta(f-f_0)\} = \int_{-\infty}^{\infty}\delta(f-f_0)e^{i2\pi ft}df = e^{i2\pi f_0 t}$$

we can finally write (Eq. 10):

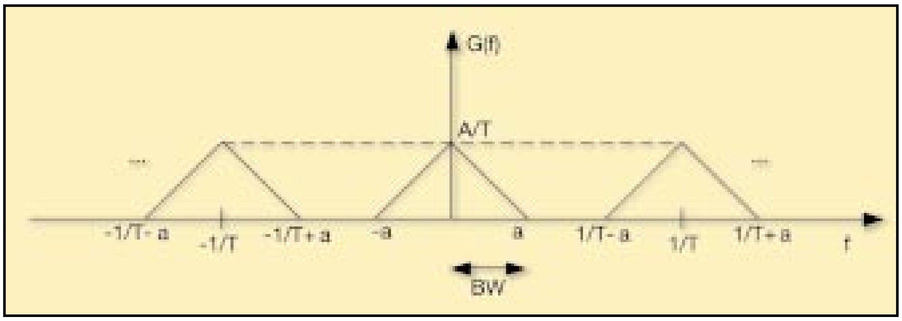

Figure 2. Frequency spectrum of the sampled signal *s(t)*.

$$e^{i2\pi\frac{n}{T}t} = \int_{-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right) e^{i2\pi ft} df$$

$$F\left\{\sum_{n=-\infty}^{\infty} \delta(t-nT)\right\} = \frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\,(f - nf_{SAMPLING})$$

Having obtained this result, we again focus our attention on the sampled baseband signal. We can now express its Fourier transform as follows (Eq. 11):

$$S(f) = G(f) * \frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right)$$

The convolution of two signals *A(f)* and *B(f)* is defined as (Eq. 12):

$$A(f) * B(f) = \int_{-\infty}^{\infty} A(f')B(f-f')df' = \int_{-\infty}^{\infty} B(f')A(f-f')df'$$

and we can express *S(f)* as (Eq. 13):

$$S(f) = G(f) * \frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right) = \int_{-\infty}^{\infty} G(f')\frac{1}{T}\sum_{n=-\infty}^{\infty} \delta\left(f - f' - \frac{n}{T}\right) df' =$$

$$= \frac{1}{T}\sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} G(f')\delta\left(f - f' - \frac{n}{T}\right) df' =$$

$$= \frac{1}{T}\sum_{n=-\infty}^{\infty} G\left(f - \frac{n}{T}\right) = \frac{1}{T}\sum_{n=-\infty}^{\infty} G\,(f - nf_{SAMPLING})$$

Equation 13, commonly called the sampling theorem, is the result we have been working toward. It shows that sampling in the time domain at intervals of *T* seconds replicates the spectrum of our unsampled signal every $1/T$ cycles per second. That result allows us to clearly and intuitively answer the question asked at the beginning: How do we sample in a way that preserves the full information of the original signal? Before answering that question, we present our result in a graphical form (Figure 2).

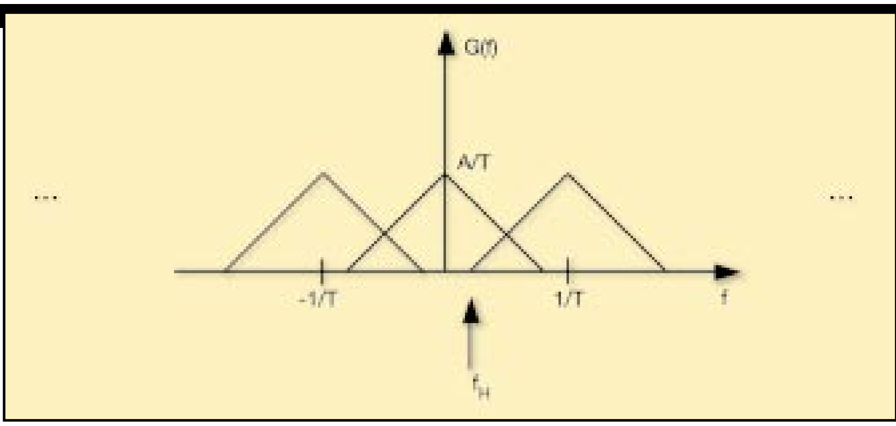

Figure 3. Aliasing.

Aliasing

To preserve all information in the unsampled baseband signal, we must ensure that the spectrum "islands" do not overlap when replicating the spectrum. If they do (a phenomenon called aliasing), we can no longer extract the original signal from the samples. Aliasing allows higher frequencies to disguise themselves as lower frequencies, as can be seen in Figure 3. To avoid aliasing, you must preserve the following condition: $1/T \geqslant 2a$, or $1/T \geqslant 2BW$. This result can be expressed in terms of the sampling frequency (Eq. 14):

$$f_{SAMPLING} \geq 2BW$$

Thus, the minimum sampling frequency

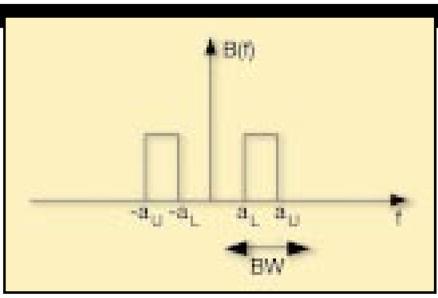

Figure 4. Bandpass signal.

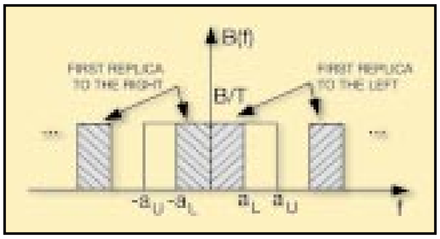

Figure 5. Spectrum of sampled bandpass signal.

necessary for sampling without aliasing is 2*BW*. This result is generally known as the Nyquist criterion.

Figure 3 depicts a sampled signal suffering from aliasing. Note that the high-frequency component f_H appears at a much lower frequency. You can recover a signal from its sampled version by using a lowpass filter to isolate the original spectrum and by cutting (attenuating) everything else. Thus, extracting the signal with a lowpass filter of cutoff frequency a does not eliminate the aliased high frequency but allows it to corrupt the signal of interest.

With that in mind, consider a special class of bandlimited signals known as bandpass signals. A bandpass signal is characterized by a bandwidth not bounded by zero at its lower end. To illustrate, the bandpass signal shown in Figure 4 has signal energy between the frequencies a_L and a_U, and its bandwidth is defined as a_U - a_L. Thus, the main difference between bandpass and baseband signals is in their definition of bandwidth: The bandwidth of a baseband signal equals its highest frequency, and the bandwidth of a bandpass signal is the difference between its upper- and lower-bound frequencies.

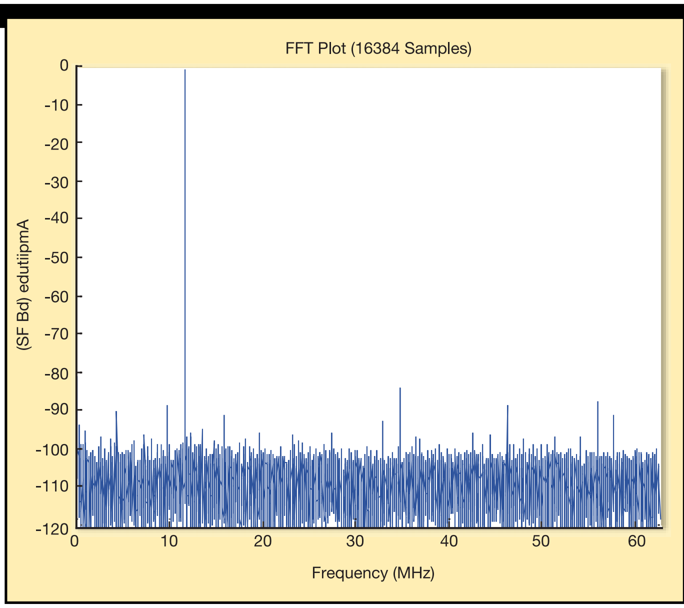

Figure 6. Spectrum of signal sampled with MAX19541, f_{SAMPLE} = 125 MHz, f_{IN} = 11.5284 MHz.

We know that by sampling this signal we replicate its spectrum at intervals of 1/*T*. Because that spectrum includes a substantial zero-amplitude band between zero and the signal's lower frequency bound, the actual signal bandwidth is smaller than a_U. We can, therefore, get away with smaller shifts in the frequency domain, which allows a sampling frequency lower than that required for a signal whose spectrum occupies all frequencies from zero to a_U. For example, assume a signal bandwidth of a_U/2. To satisfy the Nyquist criterion, our sampling frequency equals a_U, producing the sampled-signal spectrum of Figure 5.

You can see that this sampling produces no aliasing, so we could extract the original signal (from the samples) if we had a perfect bandpass filter. It is important to note in this example the difference between a baseband signal and a bandpass signal. For baseband signals, the bandwidth, and hence the sampling frequency, depend solely on the highest frequency present. For bandpass signals, bandwidth is usually smaller than the highest frequency.

Recovering sampled signal

These characteristics determine the method for recovering the sampled signal: Consider a baseband and a bandpass signal, each with the same value of maximum frequency. The bandpass signal permits a lower sampling frequency only if the method of recovery includes a bandpass filter that isolates the original signal spectrum (the white rectangles in Figure 5). A lowpass filter (used for baseband recovery) cannot recover the original bandpass signal because it includes the

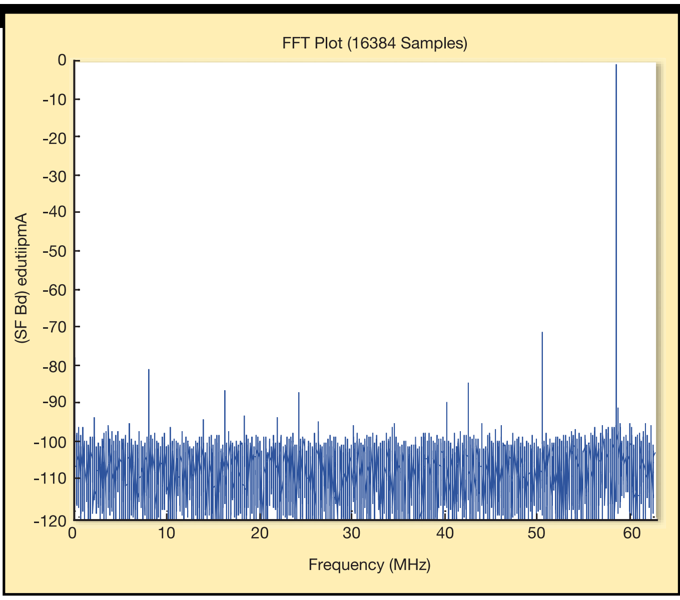

Figure 7. Spectrum of signal sampled with MAX19541, f_{SAMPLE} = 125 MHz, f_{IN} = 183.4856 MHz.

shaded areas shown in Figure 5. Thus, if you use a lowpass filter to recover the bandpass signal in Figure 5, you must sample at $2a_U$ to avoid aliasing.

Thus, bandlimited signals can be sampled and fully recovered only when observing the Nyquist criterion. For bandpass signals, the Nyquist criterion will ensure no aliasing only when the recovery of the signal is done with a bandpass filter. Otherwise, a higher sampling frequency will be required. This knowledge is important when choosing the sampling frequency of a digital-to-analog converter (DAC) or analog-to-digital converter (ADC).

One last consideration is our assumption of bandlimited signals. Mathematically, a signal can never be truly bandlimited. A law of Fourier transformations says that if a signal is finite in time, its spectrum extends to infinite frequency, and if its bandwidth is finite, its duration is infinite in time. Clearly, we can't have a time-domain signal of infinite duration, so we can never have a truly bandlimited signal. Most practical signals, however, concentrate most of their energy in a definite portion of the spectrum. The analysis above is effective for such signals.

An easy and convenient way of illustrating the disguising of higher frequencies as lower frequencies that is inherent in aliasing is by sampling sinusoidal signals. Pure sinusoidal signals have spectrums consisting only of spikes (delta functions) at the respective frequency, and aliasing with pure tones is seen as the spike moving from one location to another. This other location is referred to as the image and is in reality the aliased signal.

The results presented below were taken with the newly introduced MAX19541, 125 Msps, 12-bits ADC from Maxim. Figure 6 shows the spectrum at the converter's output for an input frequency f_{IN} = 11.5284 MHz. The main spike occurs exactly at this frequency. A number of other spikes are harmonics introduced by the non-linearities of the converter, but they are irrelevant to our discussion. The sampling frequency f_{SAMPLE} = 125 MHz is more than twice the input frequency as required by the Nyquist criterion and, therefore, no aliasing occurs. Next, let's see what will happen to the location of the main spike if we increase the input frequency to f_{IN} = 183.4856 MHz. This input frequency is higher than $f_{SAMPLE}/2$ and hence we expect aliasing to occur. The resulting spectrum given in Figure 7 shows that the main spike is now located at 58.48 MHz, and this is the aliased signal. In other words, an image has appeared at 58.48 MHz when in fact our input signal did not contain this frequency. Note that in both figures we have plotted the spectrum only up to the Nyquist frequency. The reason for this is that the spectrum is periodic and this portion contains all of the essential information. RFD

An easy and convenient way of illustrating the disguising of higher frequencies as lower frequencies that is inherent in aliasing is by sampling sinusoidal signals.

ABOUT THE AUTHOR

Vladimir Vitchev is an applications engineer with Maxim Integrated Products Inc., Sunnyvale, Calif. He obtained a bachelor's degree in electrical engineering from San Jose State University in 2002.