Selecting Mixed-Signal Components for Digital Communications Systems—Part V

Aliases, images, and spurs

by Dave Robertson

Part I (Analog Dialogue 30-3) provided an introduction to the concept of channel capacity, and its dependence on bandwidth and SNR; part II (30-4) briefly summarized different types of modulation schemes; part III (31-1) discussed different approaches to sharing the communications channel, including some of the problems associated with signal strength variability. Part IV (31-2) examined some of the architectural trade-offs used in digital communications receivers, including the problems with frequency translation and the factors contributing to dynamic range requirements. This final installment considers issues relating to the interface between continuous-time and sampled data, and discusses sources of spurious signals, particularly in the transmit path.

Digital communications systems must usually meet specifications and constraints in both the time domain (e.g., settling time) and the frequency domain (e.g., signal-to-noise ratio). As an added complication, designers of systems that operate across the boundary of continuous time and discrete time (sampled) signals must contend with aliasing and imaging problems. Virtually all digital communications systems fall into this class, and sampled-data constraints can have a significant impact on system performance. In most digital communications systems, the continuous-time-to-discrete-time interface occurs in the digital-to-analog (DAC) and analog-to-digital (ADC) conversion process, which is the interface between the digital and analog domains. The nature of this interface requires clear understanding, since the level-sensitive artifacts associated with conversion between digital and analog domains (e.g., quantization) are often confused with the time-sensitive problems of conversion between discrete time and continuous time (e.g., aliasing). The two phenomena are different, and the subtle distinctions can be important in designing and debugging systems. (Note: all digital signals must inherently be discrete-time, but analog signal processing, though generally continuous-time, may also be in discrete time—for example, with switched-capacitor circuits.)

The Nyquist theorem expresses the fundamental limitation in trying to represent a continuous-time signal with discrete samples. Basically, data with a sample rate of F_s samples per second can effectively represent a signal of bandwidth up to $F_s/2$ Hz. Sampling signals with greater bandwidth produces *aliasing*: signal content at frequencies greater than $F_s/2$ is folded, or aliased, back into the $F_s/2$ band. This can create serious problems: once the data has been sampled, there is no way to determine which signal components are from the desired band and which are aliased.

Most digital communications systems deal with band-limited signals, either because of fundamental channel bandwidths (as in an ADSL twisted-pair modem) or regulatory constraints (as with radio broadcasting and cellular telephony). In many cases, the signal bandwidth is very carefully defined as part of the standard for the application; for example, the GSM standard for cellular telephony defines a signal bandwidth of about 200 kHz, IS-95 cellular telephony uses a bandwidth of 1.25 MHz, and a DMT-ADSL twisted-pair modem utilizes a bandwidth of 1.1 MHz . In each case, the Nyquist criterion can be used to establish the *minimum* acceptable data rate to unambiguously represent these signals: 400 kHz, 2.5 MHz, and 2.2 MHz, respectively. Filtering must be used carefully to eliminate signal content outside of this desired bandwidth. The analog filter preceding an ADC is usually referred to as an *anti-alias filter*, since its function is to attenuate signals beyond the Nyquist bandwidth prior to the sampling action of the A/D converter. An equivalent filtering function follows a D/A converter, often referred to as a *smoothing filter*, or *reconstruction filter*. This continuous-time analog filter attenuates the unwanted frequency images that occur at the output of the D/A converter.

At first glance, the requirements of an anti-alias filter are fairly straightforward: the passband must of course accurately pass the desired input signals. The stopband must attenuate any interferer outside the passband sufficiently that its residue (remnant *after* the filter) will not hurt the system performance when aliased into the passband after sampling by the A/D converter. Actual design of anti-alias filters can be very challenging. If out-of-band interferers are both very strong and very near the pass frequency of the desired signal, the requirements for filter stopband and narrowness of the transition band can be quite severe. Severe filter requirements call for high-order filters using topologies that feature aggressive filter roll-off. Unfortunately, topologies of filters having such characteristics (e.g., Chebychev) typically place costly requirements on component match and tend to introduce phase distortion at the edge of the passband, jeopardizing signal recovery.

Designers must also be aware of distortion requirements for anti-alias filters: in general, the pass-band distortion of the analog anti-alias filters should be at least as good as the A/D converter (since any out-of-band harmonics introduced will be aliased). Even if strong interferers are not present, *noise* must be considered in anti-alias filter design. Out-of-band noise is aliased back into the baseband, just like out-of-band interferers. For example, if the filter preceding the converter has a bandwidth of twice the Nyquist band, signal-to-noise (SNR) will be degraded by 3 dB (assuming white noise), while a bandwidth of 4× Nyquist would introduce a degradation of 6 dB. Of course, if SNR is more than adequate, wide-band noise may not be a dominant constraint.

Aliasing has a frequency translation aspect, which can be exploited to advantage through the technique of *undersampling*. To understand undersampling, one must consider the definition of the Nyquist constraint carefully. Note that sampling a signal of *bandwidth*, $F_s/2$, requires a minimum sample rate $\geq F_s$. This $F_s/2$ bandwidth can theoretically be located anywhere in the frequency spectrum [e.g., NF_s to $(N+1/2)F_s$], not simply from dc to $F_s/2$. The aliasing action, like a mixer, can be used to translate an RF or IF frequency down to the baseband. Essentially, signals in the bands $NF_s<\text{signal}<(N+1/2)F_s$ will be translated down intact, signals in the bands $(N-1/2)F_s<\text{signal}<NF_s$ will be translated "flipped" in frequency (see Figure 1) This "flipping" action is identical to the effect seen in high-side injection mixing, and needs to be considered carefully if aliasing is to be used as part of the signal processing. The anti-alias filter in a conventional baseband system is a low-pass filter. In undersampling systems, the anti-alias filter must be a bandpass function.

Undersampling offers several more challenges for the A/D converter designer: the higher speed input signals not only require wider input bandwidth on the A/D converter's sample-and-hold (SHA) circuit; they also impose tighter requirements on the jitter performance of the A/D converter and its sampling clock. To illustrate, compare a baseband system sampling a 100-kHz sine-wave signal and an IF undersampling system sampling a 100-MHz sine-wave signal. In the baseband system, a jitter error of 100 ps produces a maximum signal error of 0.003% of full scale (peak-to-peak)—probably of no concern. In the IF undersampling case, the same 100-ps error produces a maximum signal error of 3% of full scale.

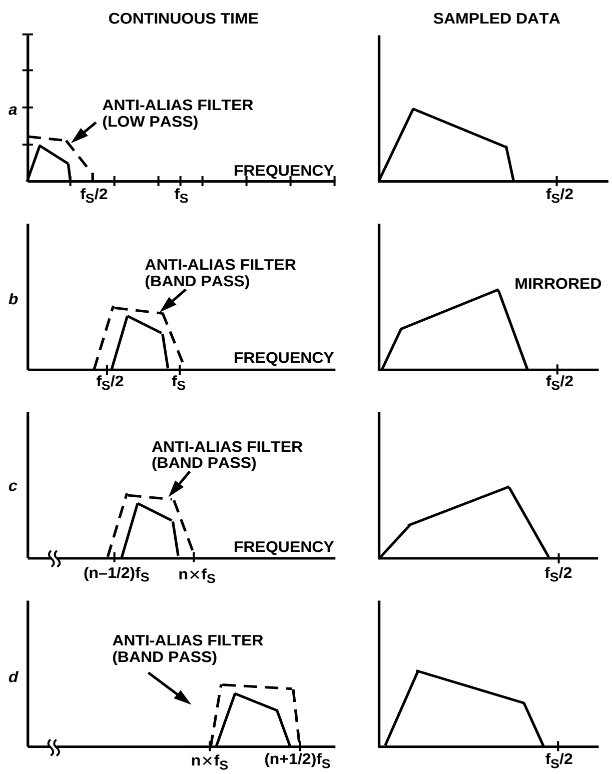

Figure 1. Aliasing, and frequency translation through undersampling.

Oversampling is not quite the opposite of undersampling (in fact, it is possible to have a system that is simultaneously oversampling and undersampling). Oversampling involves sampling the desired signal at a rate greater than that suggested by the Nyquist criterion: for example, sampling a 200-kHz signal at 1.6 MHz, rather than the minimum 400 kHz required. The oversampling ratio is defined:

$$OSR = sample\ rate/(2 \times input\ bandwidth)$$

Oversampling offers several attractive advantages (Figure 2). The higher sampling rate may significantly ease the transition band requirements of the anti-alias filter. In the example above, sampling a 200-kHz bandwidth signal at 400 kHz requires a "perfect" brick-wall anti-alias filter, since interferers at 201 kHz will alias in-band to 199 kHz. (Since "perfect" filters are impossible, most systems employ some degree of oversampling, or rely on system specifications to provide frequency guard-bands, which rule out interferers at immediately adjacent frequencies). On the other hand, sampling at 1.6 MHz moves the first critical alias frequency out to 1.4 MHz, allowing up to 1.2 MHz of transition band for the anti-alias filter.

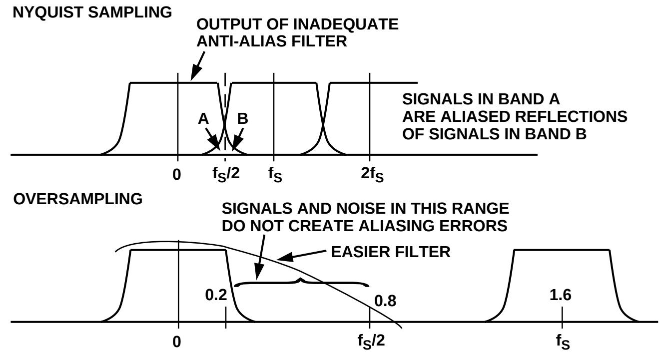

Figure 2. Oversampling makes filtering easier.

Of course, if interferers at frequencies close to 200 kHz are very strong compared to the desired signal, additional dynamic range will be required in the converter to allow it to capture both signals without clipping (see part IV, *Analog Dialogue* 31-2, for a discussion of dynamic range issues.) After conversion, oversampled data may be passed directly to a digital demodulator, or *decimated* to a data rate closer to Nyquist. Decimation involves reducing the digital sampling rate through a digital filtering operation analogous to the analog anti-aliasing filter. A well-designed digital decimation filter provides the additional advantage of reducing the quantization noise from the A/D conversion. For a conventional A/D converter, a *conversion gain* correspnding to a 3-dB reduction in quantization noise is realized for every octave (factor-of-two) decimation. Using the 1.6-MHz sample rate for oversampling as above, and decimating down to the Nyquist rate of 400 kHz, we can realize up to 6 dB in SNR gain (two octaves).

Noise-shaping converters, such as sigma-delta modulators, are a special case of oversampling converters. The sampling rate of the modulator is its high-speed clock rate, and the antialiasing filter can be quite simple. Sigma delta modulators use feedback circuitry to shape the frequency content of quantization noise, pushing it to frequencies away from the signal band of interest, where it can be filtered away. This is possible only in an oversampled system, since by definition oversampled systems provide frequency space beyond the signal band of interest. Where conventional converters allow for a 3-dB/octave conversion gain through decimation, sigma-delta converters can provide 9-, 15-, 21- or more dB/octave gain, depending on the nature of the modulator design (high-order loops, or cascade architectures, provide more-aggressive performance gains).

In a conventional converter, quantization noise is often approximated as "white"—spread evenly across the frequency spectrum. For an N-bit converter, the full-scale signal-to-quantization noise ratio (SQNR) will be ($6.02\ N + 1.76$) dB over the bandwidth from 0 to $F_s/2$. The "white" noise approximation works reasonably well for most cases, but trouble can arise when the clock and single-tone analog frequency are related through simple integer ratios—for example, when the analog input is exactly 1/4 the clock rate. In such cases, the quantization noise tends to "clump" into spurs, a considerable departure from white noise.

While much has been written in recent years about anti-aliasing and undersampling operations for A/D converters, corresponding filter problems at the output of D/A converters have enjoyed far less visibility. In the case of a D/A converter, it is not unpredictable interferers that are a concern, but the very predictable frequency images of the DAC output signal. For a better understanding of the DAC image phenomenon, Figure 3(a,b) illustrates an ideal sine wave and DAC output in both the time and frequency domains. It is important to realize that these frequency images are *not* the result of amplitude quantization: they exist even with a "perfect" high-resolution DAC. The cause of the images is the fact that the D/A converter output exactly matches the desired signal only *once* during each clock cycle. During the rest of the clock cycle, the DAC output and ideal signal differ, creating error energy. The corresponding frequency plot for this time-domain error appears as a set of Fourier-series image frequencies (c). For an output signal at frequency F_{out} synthesized with a DAC updated at F_{clock}, images appear at $NF_{clock} \pm F_{out}$. The amplitude of these images rolls off with increasing frequency according to

$$\frac{\sin \pi(F_{out}/F_{clock})}{\pi(F_{out}/F_{clock})}$$

leaving "nulls" of very weak image energy around the integer multiples of the clock frequency. Most DAC outputs will feature some degree of clock feedthrough, which may exhibit itself as spectral energy at multiples of the clock. This produces a frequency spectrum like the one shown in Figure 4.

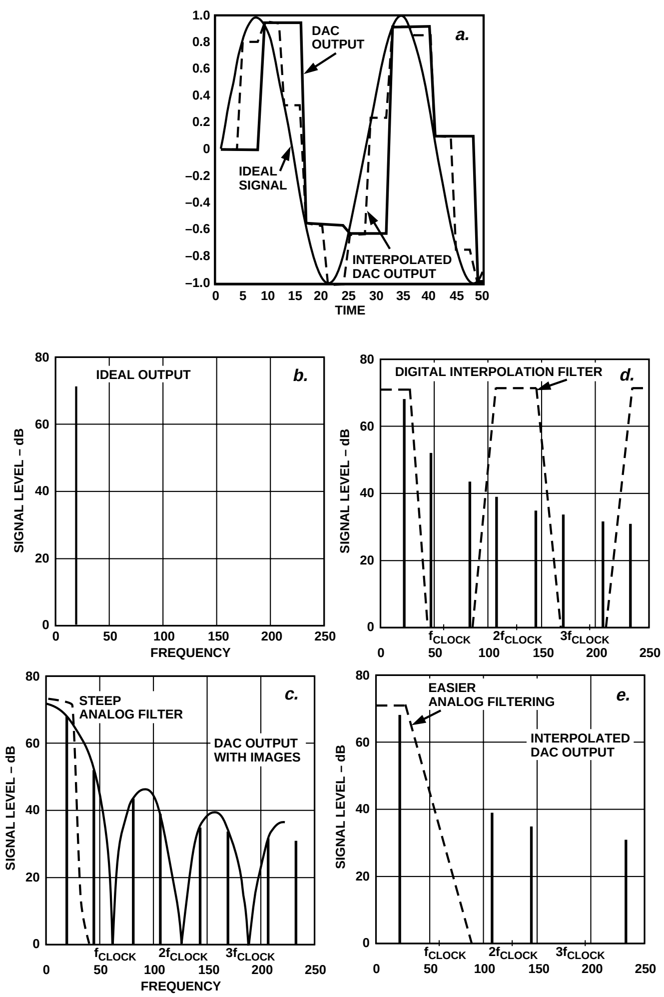

Figure 3. Time domain and frequency domain representation of continuous time and discrete sampled sine wave, and an interpolated discrete sampled sine wave.

The task of the DAC reconstruction filter is to pass the highest desired output frequency, F_{outmax}, and block the lowest image frequency, located at $F_{clock} - F_{outmax}$, implying a smoothing filter transition band of $F_{clock} - 2F_{outmax}$.

This suggests that as one tries to synthesize signals close to the Nyquist limit ($F_{outmax} = F_{clock}/2$), the filter transition gets impossibly steep. To keep the filter problem tractable, many designers use the rule of thumb that the DAC clock should be at least three times the maximum desired output frequency. In addition to the filter difficulties, higher frequency outputs may become noticeably attenuated by the sinx/x envelope: a signal at $F_{clock}/3$ is attenuated by 1.65 dB, a signal at $F_{clock}/2$ is attenuated by 3.92 dB.

Oversampling can ameliorate the D/A filter problem, just as it helps in the ADC case. (More so, in fact, since one need not worry about the strong-interferer problem.) The D/A requires an *interpolation filter*. A digital interpolation filter increases the effective data rate of the D/A by generating intermediate digital samples of the desired signal, as shown in Figure 3(a). The frequency-domain results are shown in (d,e): in this case 2× interpolation has suppressed the DAC output's first two images, increasing the available transition bandwidth for the reconstruction filter from $F_{clock} - 2F_{outmax}$ to $2F_{clock} - 2F_{outmax}$. This allows simplification of the filter and may allow more-conservative pole placement—to reduce the passband phase distortion problems that are the frequent side effects of analog filters. Digital interpolation filters may be implemented with programmable DSP, with ASICs, even by integration with the D/A converter (e.g., AD9761, AD9774). Just as with analog filters, critical performance considerations for the interpolation filters are passband flatness, stop-band rejection (how much are the images suppressed?) and narrowness of the transition band (how much of the theoretical Nyquist bandwidth ($F_{clock}/2$) is allowed in the passband?)

DACs can be used in undersampling applications, but with less efficacy than are ADCs. Instead of using a low-pass reconstruction filter to *reject* unwanted images, a bandpass reconstruction filter can be used to *select* one of the images (instead of the fundamental). This is analogous to the ADC undersampling, but with a few complications. As Figure 3 shows, the image amplitudes are actually points on a sinx/x envelope in the frequency domain. The decreasing amplitude of sinx/x with frequency suggests that the higher frequency images will be attenuated, and the amount of attenuation may vary greatly depending on where the output frequency is located with respect to multiples of the clock frequency. The sinx/x envelope is the result of the DAC's "zero-order-hold" effect (the DAC output remains fixed at target output for most of clock cycle). This is advantageous for baseband DACs, but for an undersampling application, a "return-to-zero" DAC that outputs ideal impulses would not suffer from attenuation at the higher frequencies. Since ideal impulses are physically impractical, actual return-to-zero DACs will have some rollof of their frequency-domain envelopes. This effect can be pre-compensated with digital filtering, but degradation of DAC dynamic performance at higher output frequencies generally limits the attractiveness of DAC undersampling approaches.

Frequency-domain images are but one of the many sources of spurious energy in a DAC output spectrum. While the images discussed above exist even when the D/A converter is itself "perfect", most of the other sources of spurious energy are the result of D/A converter non-idealities. In communications applications, the transmitter signal processing must ensure that these spurious outputs fall below specified levels to ensure that they do not create interference with other signals in the communications medium. Several specifications can be used to measure the dynamic performance of D/A converters in the frequency domain (see Figure 4):

- *Spurious-free dynamic range* (SFDR)—the difference in signal strength (dB) between the desired signal (could be single tone or multi-tone) and the highest spurious signal in the band being measured (Figure 4). Often, the strongest spurious response is one of the harmonics of the desired output signal. In some applications, the SFDR may be specified over a very narrow range that does not include any harmonics. For narrowband transmitters, where the DAC is processing a signal that looks similar to a single strong tone, SFDR is often the primary spec of interest.
- *Total harmonic distortion* (THD)—while SFDR indicates the strength of the highest single spur in a measured band, THD adds the energy of all the harmonic spurs (say, the first 8).
- *Two-tone intermodulation distortion* (IMD)—if the D/A converter has nonlinearities, it will produce a mixing action between synthesized signals. For example, if a nonlinear DAC tries to synthesize signals at 1.1 and 1.2 MHz, second-order intermodulation products will be generated at 100 kHz (difference frequency) and 2.3 MHz (sum frequency). Third-order intermodulation products will be generated at 1.3 MHz ($2 \times 1.2 - 1.1$) and 1.0 MHz ($2 \times 1.1 - 1.2$). The application determines which intermodulation products present the greatest problems, but the third-order products are generally more troublesome, because their frequencies tend to be very close to those of the original signals.
- *Signal-to-noise-plus-distortion* (SINAD)—THD measures just the unwanted harmonic energy. SINAD measures all the non-signal based energy in the specified portion of the spectrum, including thermal noise, quantization noise, harmonic spurs, and non-harmonically related spurious signals. CDMA (code-division, multiple-access) systems, for example, are concerned with the total noise energy in a specified bandwidth: SINAD is a more-accurate figure of merit for these applications. SINAD is probably the most difficult measurement to make, since many spectrum analyzers don't have low-enough input noise. The most straightforward way to measure a DAC's SINAD is with an ADC of significantly superior performance.

These specifications, or others derived from them, represent the primary measures of a DAC's performance in signal-synthesis applications. Besides these, there are a number of conventional DAC specifications, many associated with video DACs or other applications, that are still prevalent on DAC data sheets. These include integral nonlinearity (INL), differential nonlinearity (DNL), glitch energy (more accurately, glitch *impulse*), settling time, differential gain and differential phase. While there may be some correlation between these time-domain specifications and the true dynamic measures, the time-domain specs aren't as good at predicting dynamic performance.

Even when looking at dynamic characteristics, such as SFDR and SINAD, it is very important to keep in mind the specific nature of the signal to be synthesized. Simple modulation approaches like QPSK tend to produce strong narrowband signals. The DAC's SFDR performance recreating a single tone near full scale will probably be a good indicator of the part's suitability for the application. On the other hand, modern systems often feature signals with much different characteristics, such as simultaneously synthesized multiple tones (for wideband radios or discrete-multi-tone (DMT) modulation schemes) and direct sequence spread-spectrum modulations (such as CDMA). These more-complicated signals, which tend to spend much more time in the vicinity of the DAC's mid- and lower-scale transitions, are sensitive to different aspects of D/A converter performance than systems synthesizing strong single-tone sine waves. Since simulation models are not yet sophisticated enough to properly capture the subtleties of these differences, the safest approach is to characterize the DAC under conditions that closely mimic the end application. Such requirements for characterization over a large variety of conditions accounts for the growth in the size and richness of the datasheets for D/A converters.

For Further Reading:

For detailed discussion of discrete time artifacts and the Nyquist Theorem: Oppenheim, Alan V. and Schaeffer, Ronald W, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1989.

For more details on sigma-delta signal processing and noise shaping: Norsworthy, Steven R, Schreier; Richard; Temes, Gabor C., *Delta-Sigma Data Converters: Theory, Design, and Simulation*. New York: IEEE Press, 1997.

For more details on DAC spectral phenomena: Hendriks, Paul, "Specifying Communication DACs", *IEEE Spectrum* magazine, July, 1997, pages 58-69.

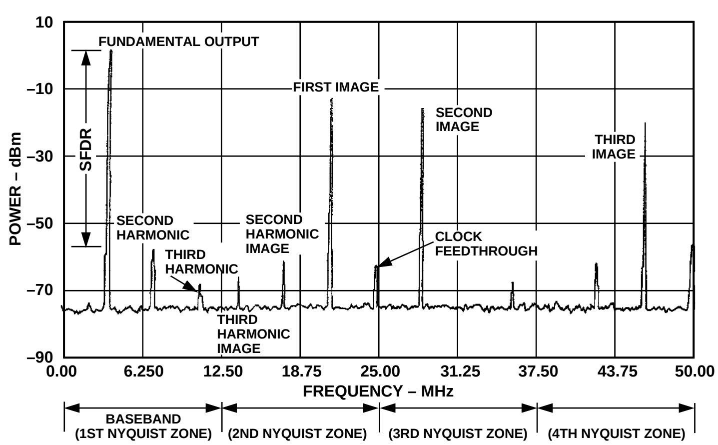

Figure 4. Different error effects in the output spectrum of a DAC.