To appear in *IEEE Potentials*

Signals and Systems II

Part II: Interpolation, decimation, complex signals, and Nyquist signaling

Jeffrey O. Coleman
Naval Research Laboratory, Radar Division

This six-part series is a mini-course, focused on system concepts, that is aimed at the gap between Signals and Systems and the usual first DSP course.

Part I discussed motivation, philosophy, and notation. This second article in the series is about oversampling in D/A conversion and the basics of decimation, complex signals, and Nyquist signaling. Figures are numbered in one sequence across the entire series, and gaps appear in their numbering in some individual articles. Figures are posted for instructional use on the author's website:
http://alum.mit.edu/www/jeffc

PART II

Oversampling for reconstruction

Fig. 4 from Part I, reproduced here, depicts the operation of a system for reconstructing a signal from its samples. In that system the analog lowpass filter must have a narrow transition region in order to pass and shape (to counter the sinc of the D/A) the desired signal spectrum and reject the unwanted spectral replicas created by sampling. Could this filtering be done digitally instead?

Not without a change in sample rate, because the periodic frequency response of a digital filter operating at the sample rate of the Fig. 4 system would cause it to treat desired and undesired spectral replicas equally. To make their treatment unequal, the period of the frequency response must be increased through *oversampling*. This does not permit digital filtering to replace analog filtering, because the periodicity of the digital filter's frequency response will inevitably cause it to pass some undesired signal components. An analog filter will still be required. But the two in combination are often the best option: a digital filter can contribute in-band precision and a sharper transition than possible with inexpensive analog filters, and an analog filter can provide asymptotic rolloff that is impossible for a digital filter.

In the "$2\times$ oversampling" system on the left in Fig. 6, interpolation by two and digital filtering have been inserted prior to D/A conversion at twice the original rate. The digital filter rejects half the replicas, including the difficult-to-suppress replicas adjacent to the desired spectral content, and the analog filter rejects the rest. Compared to the filter in Fig. 4, the much wider transition regions of this Fig. 6 analog filter make its design far less demanding. In addition, the digital filter can be designed to compensate for the undesired signal shaping by the D/A sinc and for the inevitable passband imperfections, in both magnitude and phase responses, of the nominal analog-filter response.

Just such $2\times$ oversampling was used in the signal-reconstruction systems of many early CD players, but as DSP became less expensive, it did not take long for $4\times$ oversampling to appear in order to further ease transition-band requirements on the analog filter. Higher oversampling ratios also allowed wider analog-filter passbands and thereby lowered the passband distortion they contributed to signals. At that time the digital filters were generally not designed with sufficiently sophisticated techniques to allow them to compensate for such distortion.

Figure 6 shows two such systems. In each case, the interpolation steps have been notationally combined with the filter stages that follow. Indeed, such interpolation-filtering sequences are also typically realized in combination for computational efficiency.

In the system on the right, the interpolation-filter step of the middle system is broken into two interpolation-filter steps to save computation in the realization. The first digital filter is a *shaping filter*, as it largely determines the shape of the combined response, the product of the several responses. The second digital filter is termed a *masking filter*, as its design function is to mask out entire spectral replicas, though it may incidentally contribute a little shaping as well.

Delta-sigma D/A conversion

CD manufacturers bragged loudly to clueless consumers about the "$8\times$ oversampling" and even "$16\times$ oversampling" of CD players. Such creative marketing! Decades earlier, four cylinders in engines became six, then eight, and occasionally even twelve. The original six transistors of 1960's portable radios became eight and ten transistors before the counting became silent. Should we care whether a modern radio uses a thousand transistors or a hundred thousand in its IC circuitry? Those numbers matter not at all to the consumer, and neither do the oversampling details. But it does sound cool!

The D/A conversion systems discussed thus far are generally termed *Nyquist* systems because the sampling rates used exceed the theoretical minimum alias-free sample rate, the Nyquist rate, by only a modest factor. This factor, the *oversampling ratio*, is small. Today the term *oversampled conversion* actually refers most often to *delta-sigma* conversion, which is built around a *delta-sigma modulator* and a high oversampling ratio. Delta-sigma, sigma-delta, $\Delta\Sigma$, and $\Sigma\Delta$ are interchangeable adjectives with identical meanings.

In its most basic form, a $\Delta\Sigma$ modulator uses feedback around a quantizer to replace the signal input with a one-bit output—its amplitude takes just two values—in such a way that the large quantization error introduced ends up almost entirely in a different part of the spectrum from the signal. This scheme is sometimes called *noise-shaped conversion* because of this shaping of the noise spectrum. The process is simplest when each spectral period contains quite a lot of signal-free space in which to put this quantization noise: in other words, when the system features a high oversampling ratio. The system of Fig. 7 uses $64\times$ oversampling just to create such a spectral home for that error. The analog output filter easily does the rest.

When such two-level $\Delta\Sigma$ conversion first appeared in CD players the marketing catch phrase was "single-bit conversion." The motivation for using such a scheme is simple: at the modest conversion rates involved, it is quite simple to make a good one-bit D/A converter. One might even say it is hard to make a truly bad one. This is because errors in the high and low D/A output levels are equivalent to errors in their average and difference, and these are in turn equivalent to just a DC offset and a gain error, both of which are relatively harmless compared to the nonlinear distortion a conventional D/A can introduce through hardware imperfections.

Noise-shaped D/A conversion dominated the CD-player market for many years, but as of this writing, the availability of newer, high-performance D/A converters has led to many high-performance players again using Nyquist conversion.

Complex signals

All the operations that we have applied to real signals can be applied to complex signals as well. In the frequency domain they can be multiplied by frequency responses or convolved with impulses to accomplish analog or digital filtering, D/A conversion, or sampling. Complex signals can be analog (more properly we'd say continuous time) or discrete time (d.t.), and signals of the latter form can be interpolated. A real analog signal is conveniently realized as one voltage, say, and a complex analog signal $z(t)$ is conveniently realized as two: $V_{\text{ref}}\,\mathrm{Re}\{z(t)\}$ and $V_{\text{ref}}\,\mathrm{Im}\{z(t)\}$.

This approach must be used carefully however, because V_{ref} unfortunately appears here twice, and the impossibility of making their values match as precisely in the realization hardware as is assumed in analysis, for example in sampling and D/A conversion, can lead to serious problems in some systems. A small time offset between the two component signals is a common problem also. Complex d.t. signals realized in DSP work better, as realizing complex impulse areas numerically in a computational system is no more error-prone than realizing real ones. So in system design we prefer our complex signals in d.t. form.

Some operations are unique to complex signals, and one of them, complex conjugation, is key to both our notation for complex signals and to nearly all of their meaningful applications.

The spectrum of the conjugate

When a signal is conjugated in the time domain, what becomes of it in the frequency domain? The math is simple. Conjugating the Fourier-transform equation relating $Z(f)$ and $z(t)$,

$$\begin{aligned} Z^*(f) &= \left(\int z(t)\, e^{-j2\pi f t}\, dt \right)^* \\ &= \int z^*(t)\, e^{j2\pi f t}\, dt, \end{aligned}$$

and we see that the integral would again be a Fourier transform if the sign of the exponent were only different. But this is simple: just replace f with $-f$ everywhere to obtain

$$Z^*(-f) = \int z^*(t)\, e^{-j2\pi f t}\, dt$$

and see that the Fourier transform of $z^*(t)$ is $Z^*(-f)$, the *conjugate mirroring* of $Z(f)$.

This is a slight abuse of the English language, as we really mean here the "mirror image" and not the act of mirroring itself. But as we will see soon, the term "image" has another, standard use with which we must not conflict.

Denoting conjugate mirroring in spectral sketches is a little clumsy, but we boldly proceed nevertheless with the notation of Fig. 8. Dashed lines in signal spectra denote important conjugate-mirrored counterparts to signal components that can be found elsewhere in the diagram with similar shapes but from solid lines.

Conjugate symmetry

A real signal is equal to its own conjugate, and so its transform must be equal to its own conjugate mirroring. We refer to this latter equality when we say that a spectrum is *conjugate symmetric*. In respect of this, we sketch conjugate-symmetric spectra for signals that are purely real but always sketch asymmetric spectra, as for example in Fig. 8, to represent signals that are not required to be purely real.

A fine point arises here. What does the conjugate of a spectral sketch look like? In fact, having no complex ink, we have always used real sketches to represent spectra that are in general complex. A conjugate-symmetric spectrum that is also real, as our real ink unfairly forces ours to be, is in fact an even function, so we have always used even spectral sketches as stand-ins for the conjugate-symmetric spectra that should represent real signals and for the conjugate-symmetric frequency responses of filters with real impulse responses.

Signal spectra should generally be sketched as symmetric only where complex signals would be impossible or inappropriate. Most of the systems we have designed here for real signals would work fine with complex input signals as well, for example the systems in Figs. 4 and 6. Such signals might be naturally complex or be some pair of reals, such as the left and right channels of music in stereo or the two-dimensional information giving color hue and saturation in an old-style analog television signal.

But in such systems the complex-signal version reveals nothing that the real system does not, and it is not generally appropriate to show signals as complex in these cases just to make the trivial point that two copies of a system could be built in parallel. There is more than that to the intelligent use of complex signals. We begin quite simply, however, by examining a key identity in a complex-signal context for generality.

The most noble identity

Suppose we wish to filter a bandlimited signal and then sample the result. Because digital filtering is more precise and repeatable than analog filtering, in Fig. 9 we show a periodic frequency response with the needed shaping characteristics contained within a single period. As shown this plan is unreasonable, as a d.t. impulse response can be applied in a practical way to an analog signal only in very limited contexts. To realize the benefits of digital filtering, filtering with d.t. impulse responses must be applied only to d.t. inputs with compatible rates. The d.t. impulse response must either be at the input rate or at an integral multiple thereof, the latter if input interpolation is used.

In the Fig. 9 example then, it would be much more convenient if the filtering and sampling operations could be reversed without affecting the output! We might suppress the "(f)" dependence of the variables in the notation for brevity and ask "when does $(X \times H) * G = (X * G) \times H$ hold, if ever?"

To answer this question somewhat generally, consider the two systems driven by a common input in Fig. 10. They apply identical operations but in opposite orders. On the left, spectral copies in the output have been shifted in frequency, scaled by individual impulse areas $\{a_k\}$, and then "all" shaped by frequency response $H(f)$. On the right spectral shaping comes first, so it is the product $X(f)H(f)$ that is shifted, and output copy k takes the form $a_k X(f - \text{shift}_k)\, H(f - \text{shift}_k)$. The two system outputs are thus identical except that copy k is shaped by $H(f)$ in one and by $H(f - \text{shift}_k)$ in the other. The results are therefore identical if every shift is by a multiple of the period of $H(f)$.

This conclusion is summarized in Fig. 11 under the new name *the most noble identity*, because it is a generalization of the usual noble identity for decimation, one of two noble identities discussed in multirate-DSP texts. It holds for either a finite or an infinite number of arbitrary complex impulse areas, so it is quite general. In Fig. 11 and in many figures to come, red arrows denote a system transformation that reorders two processing steps.

The other textbook noble identity, for interpolation, says nothing about signals and is just a simple statement about their realizations.

Decimation

The desired interchange of operations in Fig. 9 is indeed permitted by the most noble identity. But consider a second example, in Fig. 12, in which the new identity fails to apply because the period of the frequency response is too large. A standard "trick" applies: split the periodic impulse spectrum into a convolution of two pieces, one that is like the first except with a wider impulse spacing and a second comprising a finite set of impulses that "fills the gaps" when the two are convolved. The associativity of convolution lets us convolve the signal with these in sequence, and the period of the first is chosen so that the most noble identity will apply. One mystery remains: the final operation is a time-domain multiplication of a d.t. input with some analog waveform whose Fourier transform is the two impulses shown. What is this analog waveform?

The new operation that resulted from splitting the convolution in the last example is one that occurs routinely in signal processing and with any number of spectral impulses, not just two, so here we generalize. On the right in Fig. 13, a periodic input spectrum is convolved with N unit impulses spaced to give the spectral result one Nth the period of the input. This corresponds to the system shown, a multiplier with an input (above), an output (below), and an analog waveform (left) with which the input is multiplied.

That analog waveform can be derived exactly using an inverse Fourier transform. It turns out to be a *Dirichlet kernel*, which is sort of a periodic version of a sinc function, multiplied by a complex exponential that realizes a frequency shift. But that is not important here, as we don't need the whole waveform but only its values at the time instants at which it multiplies impulses.

We can find those values without deriving the whole waveform. In the frequency domain the waveform's output period is N times smaller than its input period, so in the time domain the output impulse spacing is N times larger than the input impulse spacing. Apparently, the analog waveform has value zero at the times of those input impulses that do not survive.

And just as we have many times now used our basic understanding of a d.t. signal, that periodicity in the frequency domain corresponds to a time-domain signal comprising impulses at certain times, we can here use the dual, that frequency-domain impulses at multiples of some fundamental frequency correspond to a periodic signal in the time domain. This is, of course, an ordinary statement about Fourier series.

So our analog waveform is periodic, and its period exactly matches the impulse spacing of the system's time-domain output. Apparently those input impulses that do survive are all multiplied by the same value. One of those survivors is at the time origin, and the analog waveform's value there is just the integral of its Fourier transform, here the total area N of its frequency-domain impulses. So, in the time domain the overall system multiplies every Nth input impulse by N and discards the rest.

Realizing this operation with the minimum input and output sample rates, as shown by the ticks, changes sample normalization by dividing by N. This just cancels the just determined scaling by N of the nonzeroed impulse areas, so the net effect of the realization is to keep every Nth input sample and discard the rest. This realization, often denoted $\downarrow N$, is termed *decimation by* N. In realizing systems, we often group decimation with some preceding operation that its realization simplifies. This is to emphasize that in realization we do not discard samples; we just do not bother to create them.

Nyquist signaling

The most noble identity was used above to rearrange the order of operations for the sake of practicality. But it is also sometimes useful for system-equivalence arguments that are less about design than about analysis and insight. In either case it is entirely typical that we use the identity to move frequency-domain convolution with impulses to an earlier point in the system, rather than the reverse.

In the sampling and reconstruction systems of Figs. 4 and 6 we began with analog signals, created d.t. signals from them, and then re-created the original analog signals. The only spectral convolution with impulses was at the beginning, so moving it earlier was never an option. But let us now begin with a d.t. signal and convert it to analog and back, as illustrated on the left in Fig. 14. Getting the filtering right was the key to successful reconstruction of an analog signal from its samples, and it is the key here as well. But now spectral convolution comes at the end of the process.

The system of Fig. 14 amounts to an abstract representation of a simple modem, one focusing on the filtering. The complex d.t. input typically has areas or samples chosen from some finite set or *constellation* of possible values according to data to be transmitted through an analog medium, perhaps a telephone line. The sample rate of the modem data impulses is termed the modem's *signaling rate*. The second line on the left in Fig. 14 is the frequency response of the cascade of the sinc response of our usual D/A converter with any digital filtering that precedes it, any analog filtering that follows it, and any frequency response of the medium itself, indeed everything up to the resampling operation. If the resampled output is equal, give or take a gain (and a delay, more generally), to the original input, the filtering is said to be *Nyquist* and is part of a *Nyquist signaling system*.

What is the nature of Nyquist filtering? In Fig. 14 the noble identity is used to show that it does not matter which operation is done first, that here $(X \times H) * G = X \times (H * G)$. It is as if the input is passed through an "equivalent" digital filter with a periodic frequency response that is just the sum of shifted copies of the actual filter's frequency response, where shifts are by all integral multiples of the signaling rate. If this so-called *Nyquist sum* is constant (more generally the $e^{-j2\pi f\tau}$ frequency response of a delay by some amount τ) everywhere, the actual filtering is Nyquist.

Part III will continue with discussions of analytic signals and linear data modulation.

Read more about it

Oppenheim and Shafer contains classic discussions of conjugate symmetry and of decimation. Vaidyanathan goes much deeper into multirate systems. Crols and Steyaert discuss analog systems for processing complex signals realized as voltage pairs. The last paper below, an early, limited precursor to this series, contains the earliest presentation of the most noble identity.

A. V. Oppenheim and R. Shafer, *Digital Signal Processing*. Prentice Hall: *http://www.phptr.com/*, 1993.

P. P. Vaidyanathan, *Multirate Systems And Filter Banks*. Prentice Hall: *http://www.phptr.com/*, 1993.

J. Crols and M. S. J. Steyaert, "Low-IF topologies for high-performance analog front ends of fully integrated receivers," *IEEE Trans. Circuits and Systems II*, vol. 45, no. 3, Mar. 1998.

J. O. Coleman, "Multi-rate DSP before discrete-time signals and systems," in *Proc. First IEEE Workshop on Signal Processing Education (SPE 2000)*, Hunt TX, Oct. 2000.

About the author

Jeffrey O. Coleman (S'75–M'79–SM'99) joined the Radar Division of the Naval Research Laboratory (NRL) in Washington DC in 1978 then left it in 1985 for graduate studies, for a stint with The Boeing Company, and for a faculty position at Michigan Technological University from which he returned to NRL in 1997. His 1975/1979/1991 SBEE/MSEE/PhD degrees are from the Massachusetts Institute of Technology, Johns Hopkins University, and the University of Washington respectively, and his research is on theory and design methods in DSP. More: http://alum.mit.edu/www/jeffc

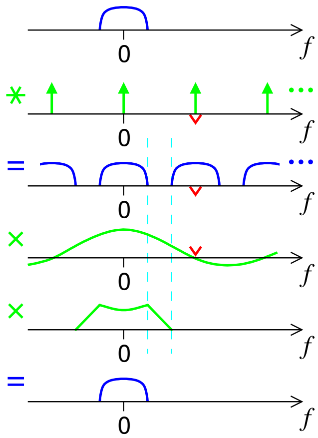

Fig. 4 (from Part I) Sampling and reconstruction of the signal sampled.

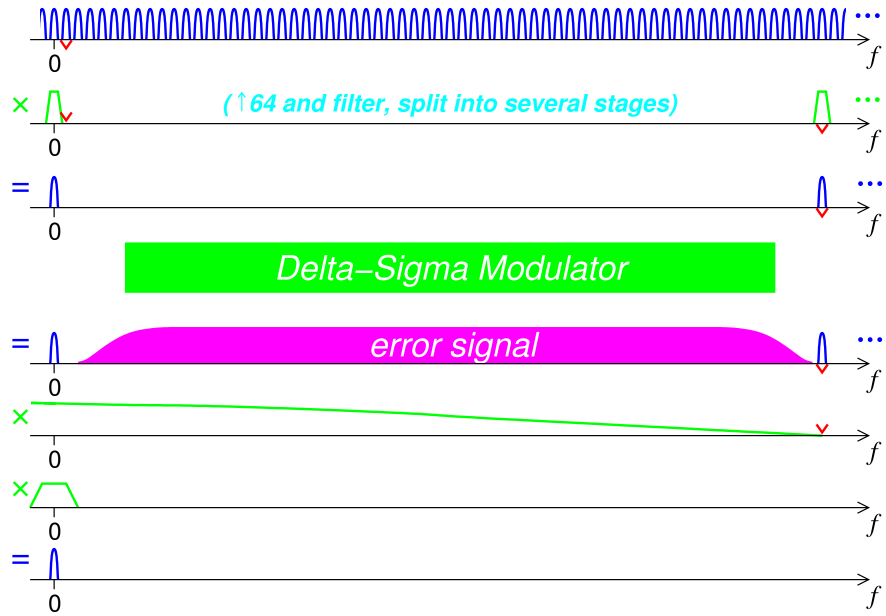

Fig. 7 A delta-sigma single-bit reconstruction system. The green box represents operations that together realize a “delta-sigma modulator,” the details of which are not shown.

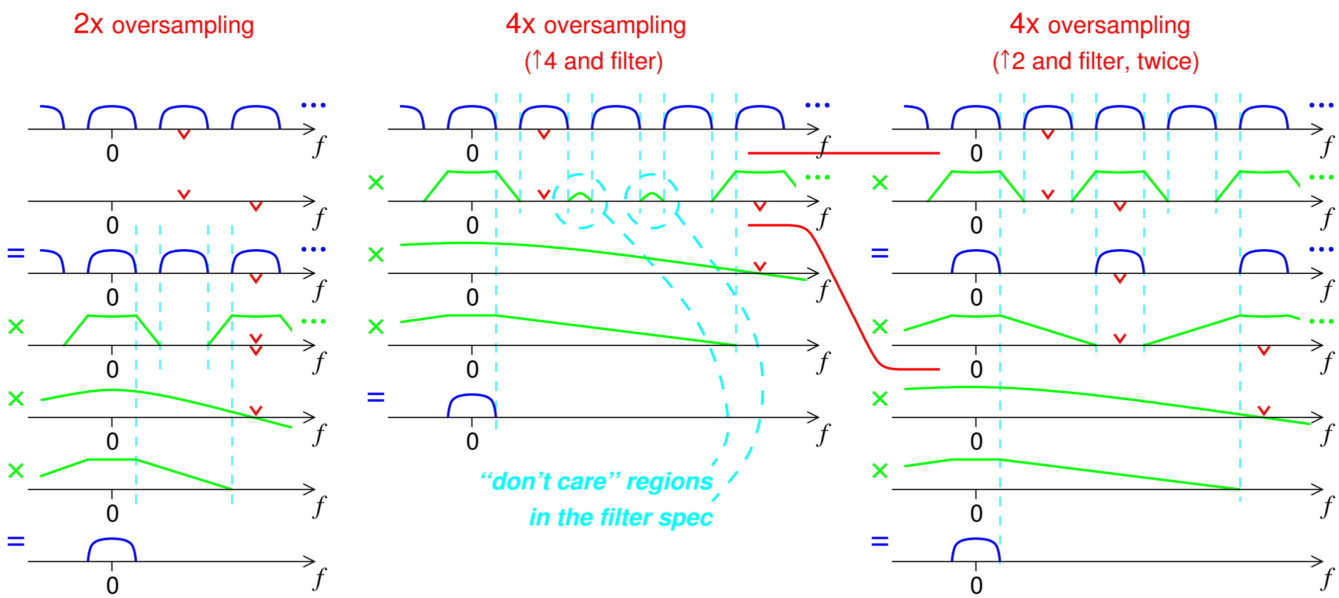

Fig. 6 Three oversampling-based systems for reconstruction of a sampled signal.

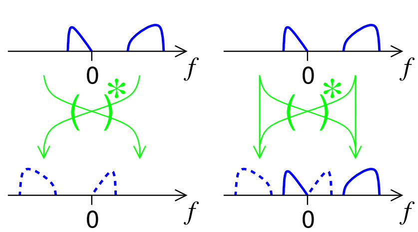

Fig. 8 Notation for and spectral effects of taking the conjugate (left) and adding the conjugate (right).

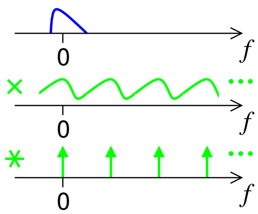

Fig. 9 Can the operations on this complex signal, spectral shaping with a complex filter and sampling, be reversed in order here for practicality?

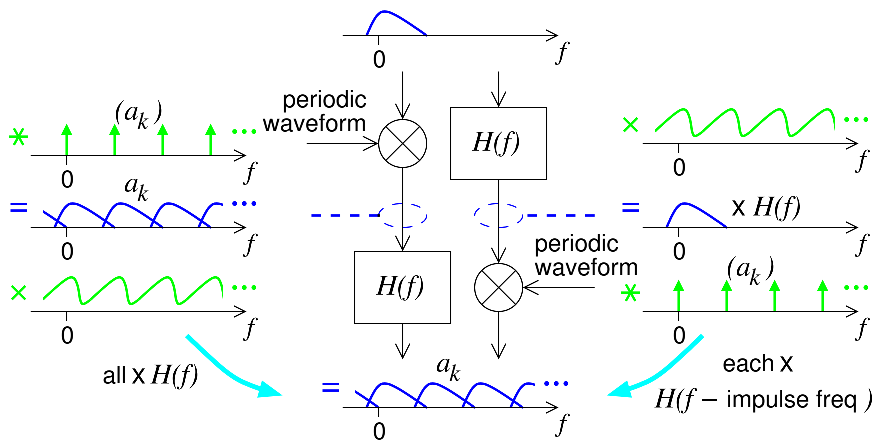

Fig. 10 A general illustration of the most noble identity.

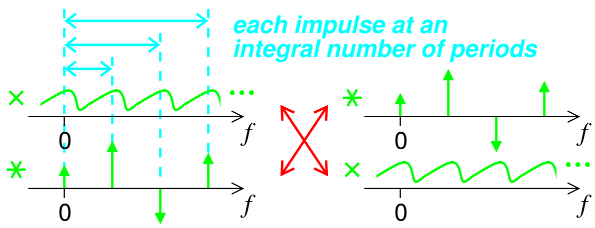

Fig. 11 The most noble identity: these operations commute if the frequency response is invariant to shifts by all impulse frequencies.

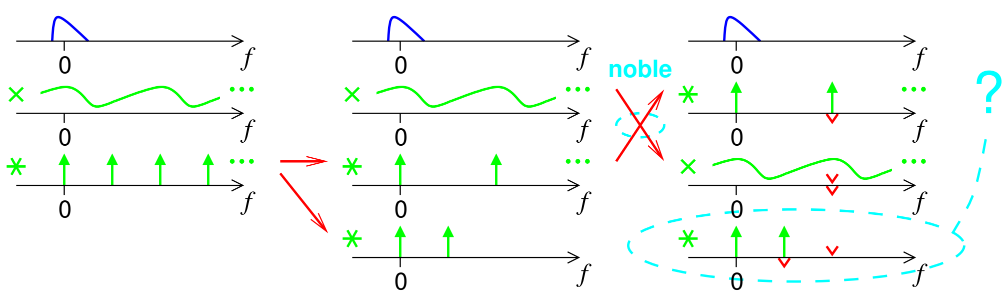

Fig. 12 Reordering a spectral-shaping and sampling system for practicality.

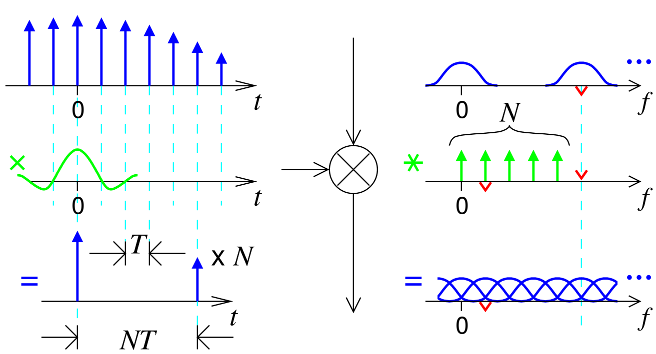

Fig. 13 Decimation multiplies a d.t. signal by a continuous waveform to discard signal impulses and replicate its spectrum.

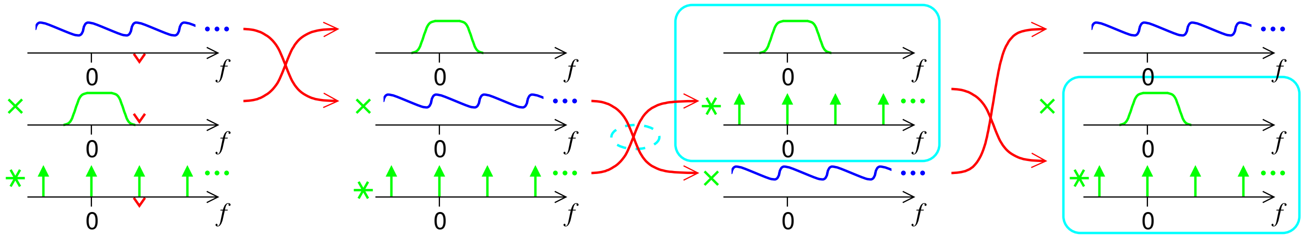

Fig. 14 The ***most noble identity*** **(center) and product reordering (other red arrows) permit the top to bottom order of operations to be altered according to the** ***parenthetical groupings*** **and lead, on the right, to the Nyquist criterion.**