

Chapter 2

Sampling

2.1 Sampling in Time and Frequency

Sampling is a common technical process with the aim to represent a continuous-time signal by a sequence of samples. A movie consists out of a sequences of photographs (the samples), a newspaper photograph has been grated in little dots in two dimensions, a television broadcast consists out of a sequence of half pictures, etc. The sampling function can be implemented in many ways. In a photo camera the chemical substance on the film is exposed during the aperture time. In modern camera's the image sensor performs this function and allows light to generate charge during a short time. In all sampling realizations, a switch mechanism followed by some form of storage is required. In an electronic circuit a sample pulse defines the sampling moments and controls a switch (relays, bipolar, MOS, and avalanche device). There are two electronics storage media available: currents in a coil and voltages on a capacitor. The practical use of a switched coil is in the ignition circuit of a combustion engine, but is here outside of the scope. The most widely applied sampling circuit in microelectronics consists of a switch and a capacitor.

The analysis of the sampling process starts with a mathematical view. The sampling pulse determines the value of a signal on a predetermined frame of time moments. The sampling frequency f_s defines this frame and determines the sampling moments as:

$$t = \frac{n}{f_s} = nT_s, \quad n = -\infty, \dots -3, -2, -1, 0, 1, 2, 3, \dots \infty \quad (2.1)$$

Between the sampling moments there is a time period of length T_s , where mathematically speaking no value is defined.¹ In practice this time period is used to perform operations on the sample sequence. These various operations (summation,

¹“No value is defined” does not imply that the value is zero! There is simply no value.

multiplication, and delay) are described in the theory of time-discrete signal processing, e.g., [16, 17] and allow to implement filtering functions. In the present context the value of T_s is considered constant, resulting in a uniform sampling pattern. Generalized non-uniform sampling theory requires extensive mathematical tools and the reader is referred to the specialized literature.

Sampling transforms a time-continuous signal in a time-discrete signal and can be applied on all types of band-limited signals. In electronics, sampling of analog time-continuous signals into analog time-discrete signals is most common. Also time-continuous digital signals (like pulse-width modulated signals) and sampled signals themselves (as found in image sensors) can be (re-)sampled.

The mathematical description of the sampling process uses the “Dirac” function. This function $\delta(t)$ is a strange² mathematical construct as it is only defined within the context of an integral. The Dirac function requires that the result of the integral equals the value of the integral function at the position of the running variable as given by the Dirac function’s argument.

$$\int_{t=-\infty}^{\infty} f(t)\delta(t-t_0) dt = f(t_0) \quad (2.2)$$

The dimension of the Dirac function is the inverse of the dimension of the running variable. A more popular, but not exact, description states that the integral over a Dirac function approximates the value “1”:

$$\delta(t) = \begin{cases} 0, & -\infty < t \leq 0 \\ \frac{1}{\epsilon}, & 0 < t < \epsilon \\ 0, & \epsilon \leq t < \infty \end{cases} \Rightarrow \int_{t=-\infty}^{\infty} \delta(t) dt = 1 \quad (2.3)$$

with $\epsilon \rightarrow 0$.

A sequence of Dirac pulses mutually separated by a time period T_s defines the time frame needed for sampling:

$$\sum_{n=-\infty}^{n=\infty} \delta(t - nT_s)$$

This repetitive sequence of pulses with a mutual time spacing of T_s can be equated to a discrete Fourier series. The discrete Fourier transform (DFT) has sinusoidal frequency components with a base frequency $f_s = 1/T_s$ and repeats at all integer multiples of f_s . The amplitude factor for each frequency component at frequency kf_s is C_k . Equating both series yields:

$$\sum_{n=-\infty}^{n=\infty} \delta(t - nT_s) = \sum_{k=-\infty}^{\infty} C_k e^{jk2\pi f_s t} \quad (2.4)$$

²Strange in the sense that many normal mathematical operations cannot be performed, e.g., $\delta^2(t)$ does not exist.

As the Dirac sequence is periodic over one period T_s , the coefficients C_k of the resulting discrete Fourier series are found by multiplying the series with $e^{-jk2\pi f_s t}$ and integrating over one period.

$$C_k = \frac{1}{T_s} \int_{t=-T_s/2}^{T_s/2} \sum_{n=-\infty}^{\infty} \delta(t - nT_s) e^{-jk2\pi f_s t} dt \quad (2.5)$$

Within the integration interval there is only one Dirac pulse active at $t = 0$, so the complicated formula reduces to:

$$C_k = \frac{1}{T_s} \int_{t=-T_s/2}^{T_s/2} \delta(t) e^{-jk2\pi f_s t} dt = \frac{1}{T_s} e^{-jk2\pi f_s \times 0} = \frac{1}{T_s} \quad (2.6)$$

Now the substitution of C_k results in the mathematical description of the DFT from the sequence of Dirac pulses in the time domain.

$$\sum_{n=-\infty}^{\infty} \delta(t - nT_s) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} e^{jk2\pi f_s t} = \frac{1}{T_s} \left(1 + \sum_{k=1}^{\infty} 2 \cos(k2\pi f_s t) \right) \quad (2.7)$$

Note that both terms are time-domain functions. The right-hand side is a summation of simple sine waves that can also be obtained from a frequency domain description using Dirac functions. This sum of Dirac functions in the discrete frequency domain is the counterpart of the time-domain Dirac sequence.

$$\sum_{n=-\infty}^{\infty} \delta(t - nT_s) \Leftrightarrow \sum_{k=-\infty}^{\infty} \delta(f - kf_s) \quad (2.8)$$

The infinite sequence of short time pulses corresponds to an infinite sequence of frequency components at integer multiples of the sampling rate.

2.1.1 Sampling Signals

In Fig. 2.1 a signal³ $A(t)$ is sampled at a rate f_s . This signal corresponds in the frequency domain to $\mathbf{A}(f) = \mathbf{A}(\omega/2\pi)$ with a bandwidth from $f = 0$ Hz to $f = BW$. The straight-forward Fourier transform is defined as⁴:

³For clarity this chapter uses for time-domain signals normal print, while their spectral equivalents are in bold face. The suffix s refers to sampled sequences.

⁴The Fourier transform definition and its inverse require that a factor $1/2\pi$ is added somewhere. Physicists love symmetry and add $1/\sqrt{2\pi}$ in front of the transform and its inverse. Engineers mostly shift everything to the inverse transform, see Eq. 2.15. More attention is needed when a

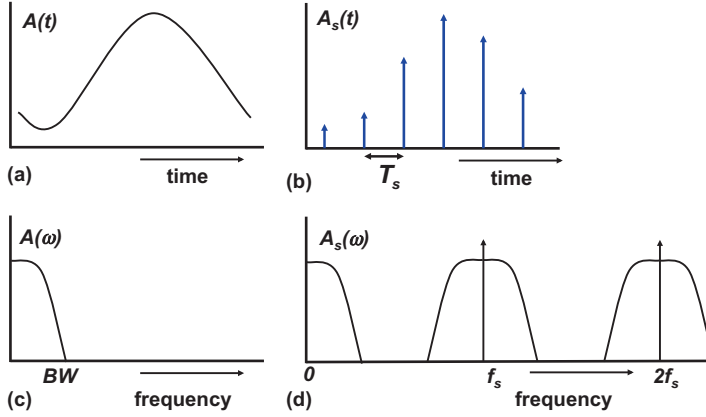


Fig. 2.1 Sampling an analog signal (a) in the time-continuous domain results in a series of analog signal samples (b). In the frequency domain the time-continuous signal (c) is folded around the sampling frequency and its multiples (d)

$$\mathbf{A}(f) = \int_{t=-\infty}^{\infty} A(t) e^{-j2\pi ft} dt \quad (2.9)$$

Mathematically sampling is performed by multiplying the time-continuous function $A(t)$ of Fig. 2.1a with the sequence of Dirac pulses, resulting in a time-discrete signal in Fig. 2.1b. The product of the time-continuous function and the Dirac sequence is defined for those time moments equal to the multiples of the sampling period T_s

$$A_s(t) = \int_{\tau=-\infty}^{\infty} A(t - \tau) \times \sum_{n=-\infty}^{n=\infty} \delta(\tau - nT_s) d\tau \quad (2.10)$$

A useful property of the Fourier transform is that a multiplication of time-domain functions corresponds to a convolution of their frequency counterparts in the Fourier domain.

$$\mathbf{A}_s(f) = \int_{\chi=-\infty}^{\infty} \mathbf{A}(f - \chi) \times \sum_{k=-\infty}^{k=\infty} \delta(\chi - kf_s) d\chi \quad (2.11)$$

Evaluation of this integral is easy as the result of an integral with a Dirac function is the integral function evaluated at the values where the Dirac function is active. In this case this means that for $k = 0$ the sampled data spectrum equals the original time-continuous spectrum: $\mathbf{A}_s(f, k = 0) = \mathbf{A}(f)$. For $k = \pm 1$ the result is $\mathbf{A}_s(f, k = \pm 1) = \mathbf{A}(f_s \pm f)$. Or in other words: the spectrum is mirrored around the first

single-sided engineering type Fourier transform is used instead of a mathematically more accurate double sided.

multiple of the sample rate. The same goes for $k = 2, 3, \dots$ and negative k . Another property of the Fourier transform is applied: for a real function in the time domain, the Fourier result is symmetrical around 0: $\mathbf{A}(f) = \mathbf{A}(-f)$.

Consequently $\mathbf{A}_s(f)$ consists of a sum of copies of the time-continuous spectrum each with a frequency shift of $k \times f_s$. The total spectrum \mathbf{A}_s can be written as:

$$\mathbf{A}_s(f) = \sum_{k=-\infty}^{\infty} \mathbf{A}(f - kf_s) \quad (2.12)$$

The original time-continuous signal $A(t)$ is connected to only one spectrum band in the frequency domain $\mathbf{A}(f)$. By sampling this signal with a sequence of Dirac pulses with a repetition rate (f_s) a number of replicas of the original spectral band $\mathbf{A}(f)$ are created on either side of each multiple of the sampling rate f_s . Figure 2.1c, d depicts the time-continuous signal and the sampled signal in the frequency domain. In the frequency domain of the sampled data signal, next to the original signal, also the upper bands are present.

The idea that from one spectrum an infinite set of spectra is created seems to contradict the law on the conservation of energy. If all spectra were mutually uncorrelated and could be converted in a physical quantity, there would indeed be a contradiction. However, in the reconstruction from a mathematical sequence of Dirac pulses to a physical quantity, there is an inevitable filtering operation, limiting the energy.

An important consequence of the previous sampling theory is that two frequency components in the time-continuous domain that have an equal frequency distance to arbitrary multiples of the sampling frequency will end up on the same frequency location in the sampled data band. Figure 2.2 shows three different sine wave signals that all result in the same sampled data signal (dots). Different signals in the time-continuous domain can have the same representation in the time-discrete domain.

A time-continuous signal close to $(m \times f_s)$ will result in replica around $((k \pm m) \times f_s)$. In case $k = m$ the signals around $(m \times f_s)$ will appear near DC, shifting the original signal band to low frequencies. This phenomena not only finds an application in down-mixing of communication signals, Sect. 2.3.2, but also means that unwanted or unexpected signals will follow the same path and show up in the wanted bandwidth. In the Sect. 2.2.2 countermeasures are discussed.

Example 2.1. The input stage of a sampling system is generating distortion. Plot the input signal of 3.3 MHz and its 4 harmonics in the band between 0 and a sampling frequency of 100 Ms/s. Change the sampling rate to 10 Ms/s and plot again.

Solution. Figure 2.3 shows on the left side the sampled spectrum at 100 Ms/s and on the right side the 10 Ms/s variant. Signal components will appear at frequencies: $i \times f_{in} \pm j \times f_s$, where $i = 1 \dots$ number of harmonics and $j = 0 \dots \infty$. Therefore in the last graph there are components at the frequencies shown in Table 2.1.

Note that within each $f_s/2$ range there are exactly five components, corresponding to each of the original tones.

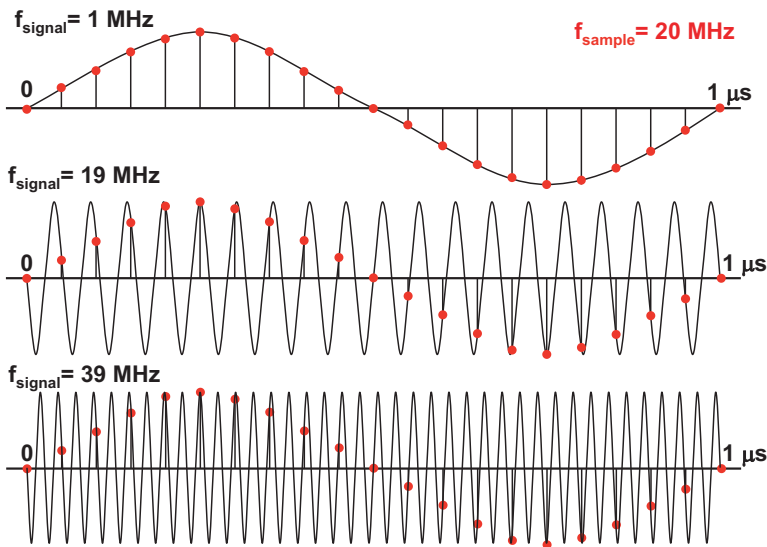


Fig. 2.2 Sampling three time-continuous signals: 1, 19, and 39 MHz sine waves result after sampling with 20 Ms/s in the same sampled data sequence (*dots*)

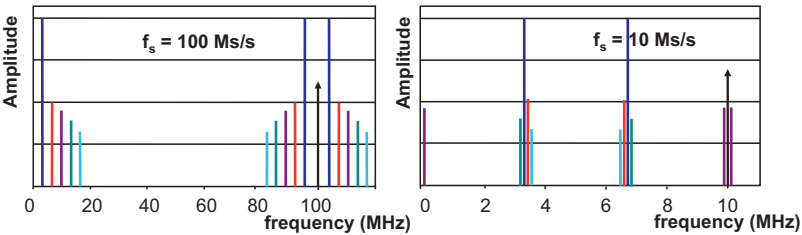


Fig. 2.3 The sample rate in the *left* plot is 100 Ms/s and in the right plot 10 Ms/s

Table 2.1 Frequency components of a distorted 3.3 MHz sinusoid sampled at 10 Ms/s

0.1 MHz	$f_s - 3f_{in}$
3.2 MHz	$f_s - 4f_{in}$
3.3 MHz	f_{in}
3.4 MHz	$f_s - 2f_{in}$
3.5 MHz	$f_s - 5f_{in}$
6.5 MHz	$2f_s - 5f_{in}$
6.6 MHz	$2f_{in}$
6.7 MHz	$f_s - f_{in}$
6.8 MHz	$2f_s - 4f_{in}$
9.9 MHz	$3f_{in}$

2.2 Sampling Limits

2.2.1 Nyquist Criterion

Figure 2.4 shows a signal in the time domain with higher frequency components than the signal in Fig. 2.1. The samples of this signal are valid values of the signal at that sample moments, however, it is not possible to reconstruct uniquely the signal based on these values. A likely reconstruction would be the dotted line, a signal that largely differs from the original.

If the bandwidth in the time-continuous domain increases, the mirror bands around the multiples of the sample frequency will follow. Figure 2.5 shows that this will lead to overlap of signal bandwidths after sampling, and mixing of data. This phenomenon is called “aliasing.” The closest upper band directly adjacent to the original band is called: “the alias band.” If the original signal mixes with its alias, its information contents are corrupted.

This limitation of sampling signals is known as the “Nyquist” criterion. Indicated already in a paper by Harry Nyquist [18, Appendix 2-a], it was Claude E. Shannon

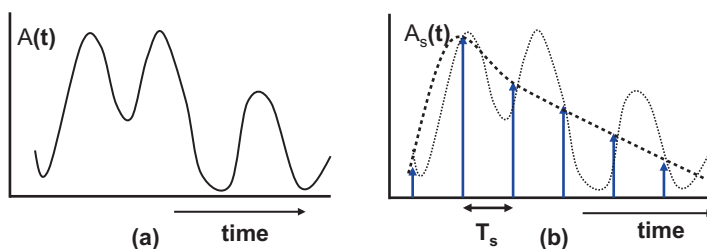


Fig. 2.4 The time-continuous signal contains higher frequency components (a) and does not satisfy the Nyquist criterion. The sample series in the time domain (b) allow multiple reconstructions of the original time-continuous signal

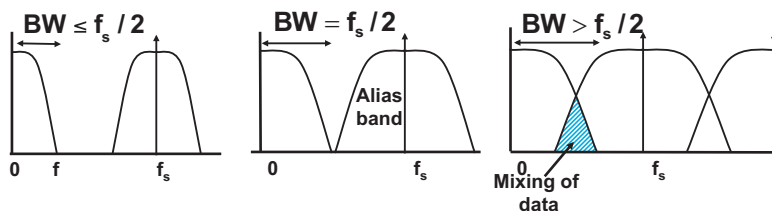


Fig. 2.5 The time-continuous bandwidth can be increased until half of the sample rate is reached

who extended his mathematical theory of communication [19] in 1949 with a paper dealing with communication in the presence of noise. In that paper [20] the Nyquist criterion, also known as Nyquist theorem, is formulated as⁵:

“If a function contains no frequencies higher than BW cycles per second, it is completely determined by giving its ordinates at a series of points spaced $1/2BW$ seconds apart.”

This Nyquist criterion says that if the sample rate is more than twice the highest frequency in a bandwidth, there is a theoretical manner to uniquely reconstruct the signal. This criterion imposes a simple mathematical relation between a bandwidth BW and the minimum sample rate f_s :

$$f_s > 2BW \quad (2.13)$$

The Nyquist sample rate is often defined as $f_{s,ny} = 2 \times BW$ and the Nyquist bandwidth is the bandwidth $BW = f_{s,ny}/2$. The Nyquist frequency is the highest frequency in the Nyquist bandwidth.⁶ This criterion is derived assuming ideal filters and an infinite time period to reconstruct the signal. In practical circumstances designers will use additional margins to avoid having to meet these constraints. An interesting discussion on present insights in the mathematical aspects of the Nyquist criterion was published by Unser [21].

The Nyquist criterion specifies that the useable bandwidth is limited to half of the sample rate. But the Nyquist criterion does not define where the bandwidth is located in the time-continuous spectrum. The only constraint on the position of this limited bandwidth is that this bandwidth does not include any multiple of half of the sample rate. That would lead to overlap in the sampled spectrum. There is no need to specify the bandwidth starting at 0 Hz. For example, if it is known that the original signal in Fig. 2.2 is in the bandwidth between 10 and 20 MHz, the samples can be reconstructed to yield the originating 19 MHz time-continuous sine wave. A bandwidth, located in the spectrum at a higher frequency than the sample rate, can therefore also be properly sampled. The sampling operation generates copies around all multiples of the sample rate, including near DC. This is in some communication systems used to down-modulate or down-sample signals, see Sect. 2.3.1.

Example 2.2. A 10 kHz sine wave is distorted with components at 20, 30, 40, and 50 kHz, and sampled with 44 ks/s. Draw the spectrum.

Solution. In the left part of Fig. 2.6 shows the input spectrum. The right part shows the result after sampling. The tones from the original spectrum are in bold lines, the results of the folding and mirroring around the 44 ks/s sample rate are drawn

⁵Other originators for this criterion are named as E. Whittaker and V.A. Kotelnikov. Landau proved in 1967 for non-baseband and non-uniform sampling that the average sample rate must be twice the occupied bandwidth.

⁶Precise mathematicians will now argue that a signal with frequency $f = f_{s,ny}/2$ cannot be reconstructed, so they should read here: $f = f_{s,ny}/2 - \Delta f$, where Δf goes to zero.

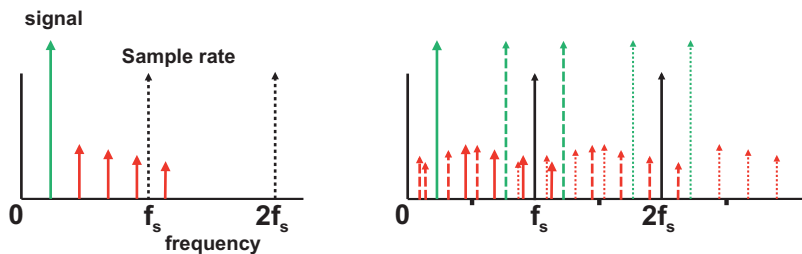


Fig. 2.6 The sample rate in the upper plot is 44 ks/s

in dashed lines and the components originating from the second multiple at 88 ks/s are shown in thinner dashed lines. The 50 kHz component results after sampling in a (-6) kHz frequency. This component is shifted to the positive frequency domain, while keeping in mind that it differs in phase. In every interval of $f_s/2$ width there is exactly one copy of each originating component. So there is a simple check on the correctness and completeness of the spectrum: make sure the number of components exactly matching the number at the input.

Note here, that the discrete tones as they are, have zero bandwidth. And as long as folding of one tone on top of another is prevented, or is considered irrelevant, many forms of sampling can be applied.

Example 2.3. A sampling system with distortion is excited with a single-tone sine wave. An unexpected tone appears in the output spectrum. How do you find out what its origin is?

Solution. In a sampled data system it is obvious that tones can be generated by distortion of the sine wave in combination with aliasing through sampling. The first check is made by varying the input frequency by a small frequency offset Δf_{in} . If the tone in the output spectrum varies by a multiple of the offset $i \times \Delta f_{in}$, the i -th harmonic of the input tone is involved. Now the same procedure is repeated for the sampling rate, identifying $j \times \Delta f_s$ as the originator. The integers i, j in the formula $i \times f_{in} \pm j \times f_s$ are known and the evaluation of this formula should point at the unexpected tone position.

There are of course many more possible scenarios, e.g., suppose there is a reaction on varying the sampling rate, but not on input signal variations. Probably an external frequency is entering the system and gets mixed down in the sampling chain.

2.2.2 Alias Filter

The Nyquist criterion requires that all input signals are band-limited in order to prevent mixing up of modulated signal components. This requirement is even more

stringent for signals that are up-front unwanted: various noise contributions, tones, distortion, etc. These signals are also modulated by the (integer multiples of) the sample rate. As this sampling process stretches out into high-frequency ranges, even RF signals can cause low frequency disturbance after modulation with hundreds $\times f_s$. In a correctly designed analog-to-digital conversion system the bandwidth of the incoming signal is limited by means of an “alias-filter,” so that no mixing of out-of-band frequency components can take place, see Fig. 2.7. An analog-to-digital converter is therefore preceded by a band-limiting filter, see Fig. 2.8. This filter prevents the components outside the desired frequency range to be sampled and to mix up with the wanted signals.

In practical system design it is recommended to choose a higher sample rate than prescribed by the Nyquist criterion. The fraction of frequency spacing between the extremes of the base and its alias with respect to the sample rate determines the number of poles needed in the anti-alias filter. A filter will suppress signals at a rate of 6 dB per octave per filter pole, Fig. 2.9.

Sharp band-limiting filters require many accurately tuned poles. Additional amplification is needed, and therefore these filters tend to become expensive and hard-to-handle in a production environment. On the other hand, there are some good reasons not to choose for an arbitrary high sample rate: the required capacity for storing the digital data will increase linear with the sample rate, as well as the power needed for any subsequent data processing.

Anti-alias filters are active or passive time-continuous filters. Time-discrete filters, such as switched-capacitor filters, sample the signal themselves and require consequently some alias filters. An additional function of the anti-alias filter can be the suppression of unpredictable interference in the system. Note that interference

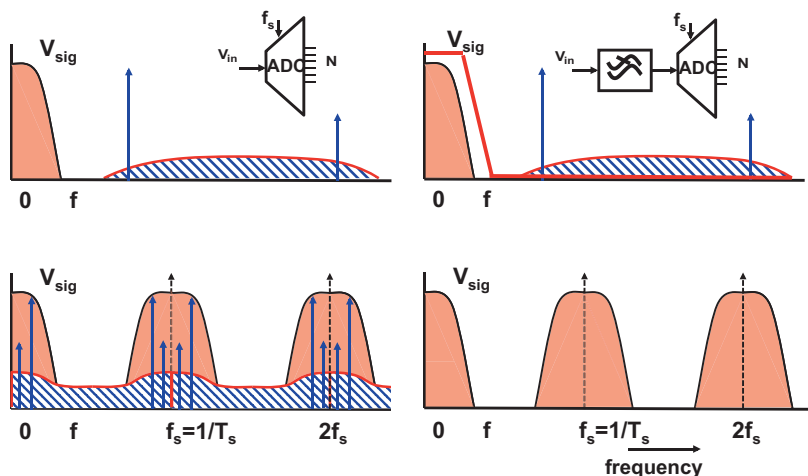


Fig. 2.7 *Left:* sampling of an unfiltered signal leads to lots of unwanted components in the signal. *Right:* an alias filter prevents disturbing signals, tones, or spurs to enter the signal band

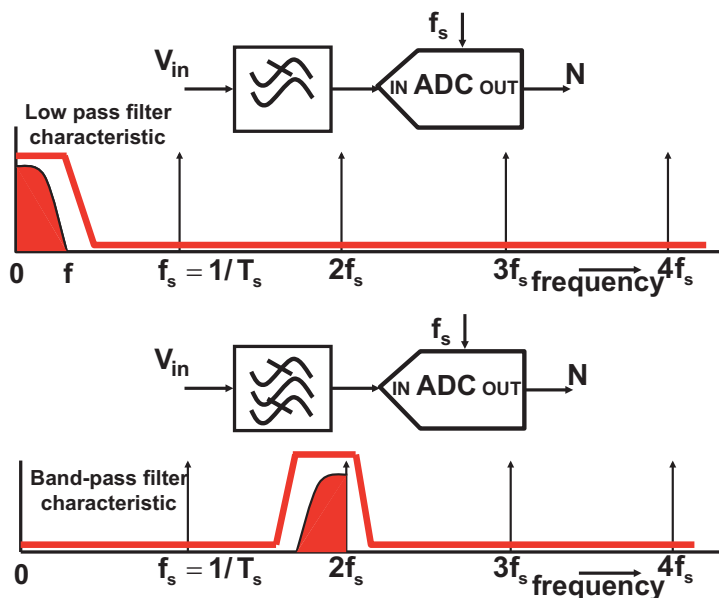


Fig. 2.8 A low-pass filter (*upper*) or bandpass filter (*lower*) is used to avoid unwanted components near other instances of the sample rate

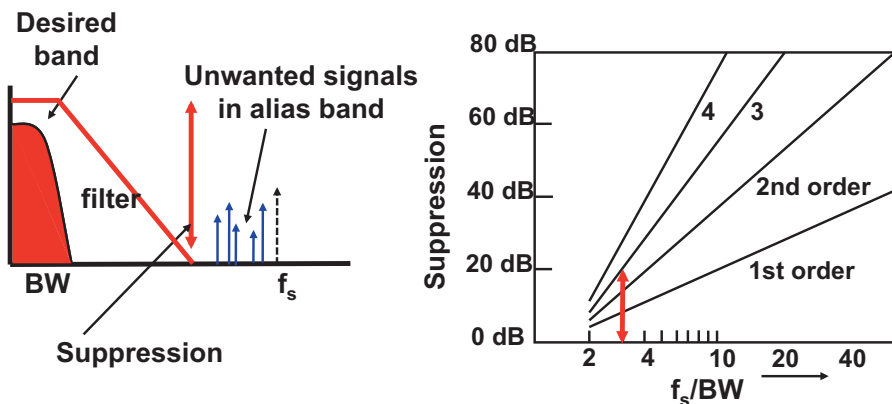


Fig. 2.9 The attainable suppression of the anti-alias filter depends on the number of poles in the filter and the ratio of the bandwidth to the sample rate

does not necessarily enter the system via its input terminal, the designer should have an equal interest in suppressing any interference on supply and bias lines.

Some systems are band-limited by construction. In a radio, the IF filters of a heterodyne radio architecture may serve as anti-alias filters, and in a sensor system, the sensor may be band-limited.

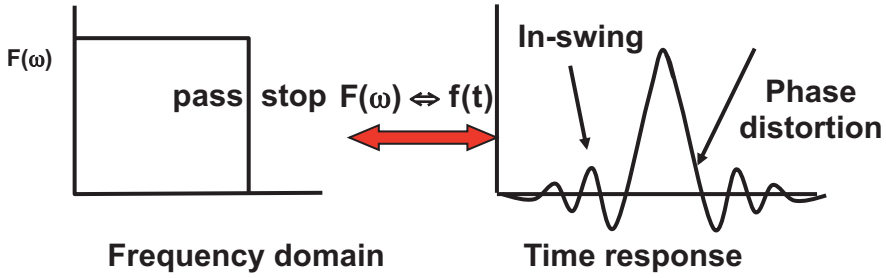


Fig. 2.10 The mathematical description of the relation between frequency domain and time domain implies that a sharp-limited frequency response generates a ringing response in the time domain

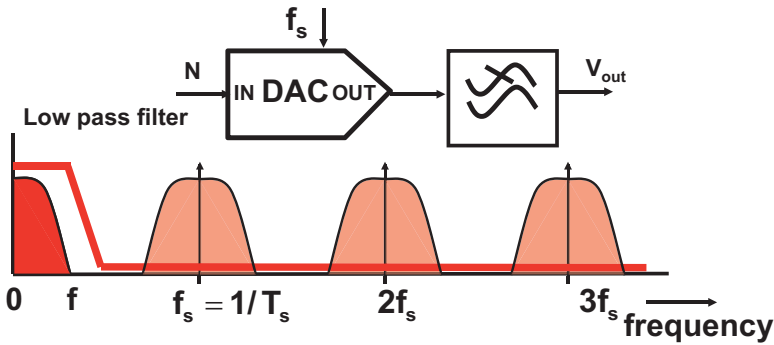


Fig. 2.11 Before using the result of the reconstruction the higher harmonic bands must be removed. In this example it is assumed that the reconstruction provides no filtering

The definition of a filter for alias suppression requires to look at pass-band, rejection but also at signal ringing. Figure 2.10 shows a brick-wall filter characteristic.

$$\mathbf{F}(\omega) = \begin{cases} 0, & \omega \leq 0 \\ 1, & 0 < \omega \leq \omega_{BW} \\ 0, & \omega > \omega_{BW} \end{cases} \quad (2.14)$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{F}(\omega) e^{j\omega t} d\omega \quad (2.15)$$

$$f(t) = \frac{2 \sin(\omega_{BW} t)}{2\pi t} \quad (2.16)$$

The impulse response is a $\sin(x)/x$ function with ringings on both sides of the pulse. These ringings will be triggered by transitions in the signal and are disastrous in many applications. On a television screen a vertical stripe would on either side be accompanied by shadows. In an audio system these ringings lead to phase distortion, or in more musical terms: blurring of the instrument's position. These practical considerations often lead to a sample rate of at least $2.5\text{--}3\times$ the bandwidth.

Alias filtering at the input of a conversion chain is necessary to remove unwanted components in the spectrum that may fold back into the signal after sampling. Also at the output side of the conversion chain an alias filter can be necessary as the sampled data format contains high-frequency components, Fig. 2.11. During reconstruction, see Sect. 2.4, some filtering will occur, but mostly additional filtering is needed to avoid problems. If these components are processed in a non-linear fashion, unwanted signals can be produced. Also other failures may occur. For example, in an audio chain, the speakers are dimensioned assuming that (by far) most audio energy is in the low frequency range. If too much alias products exist, the tweeters can be harmed.

Example 2.4. A bandwidth of 2 MHz is sampled at 10 MHz. Determine the order of an anti-alias filter build as a cascade of equal first order filters, suppressing alias components by 35 dB.

Solution. Aliasing will occur due to signals in bands around multiples of the sampling rate. In this case all signals between $f_s - BW = 8$ and $f_s + BW = 12$ MHz will appear after sampling in the desired signal band. The task will be to design a filter that passes a bandwidth of 2 MHz, but suppresses at 8 MHz. The transfer expression is

$$H(\omega) = \left| \frac{1}{1 + (\omega\tau)^2} \right|^{n/2}$$

where n is the filter order and assuming that $\omega\tau = 1$ for a frequency of 2 MHz, then a 3th order filter is chosen. This filter will attenuate the signal at 2 MHz by 9 dB. If only 3 dB attenuation is allowed, the filter order must be increased to 7. It is obvious that doubling the sample rate eases this trade-off dramatically.

Example 2.5. Comment on the choice of the sample rate in the CD audio standard.

Solution. An example of a critical alias filter is found in the compact-disc music recording format. Here a sample rate of 44.1 ks/s⁷ is used for a desired signal bandwidth of 20 kHz. This combination leaves only a small transition band between 20 and 24.1 kHz to suppress the alias band by some 90 dB. The expensive filter required to achieve this suppression needs some 11–13 poles. Moreover such a filter will create a non-linear phase behavior at the high baseband frequencies. Phase distortions are time distortions ($\Delta\text{phase} = \text{signal frequency} \times \Delta\text{time}$) and have a strong audible effect. Fortunately the use of “oversampling” allows to separate baseband and alias band sufficiently, see Sect. 10.1.

⁷The only storage in the early days of CDs were video recorders. The 44.1 ks/s sample rate was chosen such that the audio signal exactly fits to a video recorder format (25 fields of 588 lines with 3 samples per line) of 44.1 ks/s.

2.2.3 Getting Around Nyquist?

An implicit assumption for the Nyquist criterion is that the bandwidth of interest is filled with relevant information. This is not necessarily true in all systems. In communication systems like Wifi or GSM, only a few channels in the allocated band will be used at any moment in time. Video signals are by their nature sampled signals: a sequence of images consisting of sequences of lines. The spectral energies are concentrated around multiples of the video line frequency. The intermediate frequency bands are empty. These systems show “sparsity” in the frequency domain.

In a radar or ultra-sound system a pulse is generated and transmitted. The only relevant information for the system is the moment the reflected pulse is received. This is an example of time sparsity.

A sparse signal in a relatively wide bandwidth can be reconstructed after sampling by a non-uniform sampling sequence. Such a sequence can be generated by a high-frequency random generator. The information from the few active carriers is spread out over the band and theoretically it is possible to design algorithms that recover this information. A first intuitive approach is to assume a high uniform sampling pattern, from which only a few selected samples are used. In a higher sense the Nyquist criterion is still valid: the total amount of relevant bandwidth (Landau bandwidth) is still less than half of the effective sample rate.

Compressive sensing or compressive sampling [22] multiplies the signal with a high-rate random sequence, which is easier to implement in the analog domain than sampling. The relevant signals are again spread out over a large bandwidth. After bandwidth-limiting, a reconstruction (“L1” minimization) is possible if the random sequence is known and the domain in which the signal is monitored, is sparse. The theory is promising but requires heavy post-processing. Whether a real advantage can be obtained remains to be proven.⁸

Example 2.6. Two sine wave signals at 3.2 and 4.8 MHz each modulated with a 0.1 MHz bandwidth signal are sampled at 1.1 Ms/s. Is the Nyquist criterion violated?

Solution. No, the total band occupied with relevant signal is 0.4 MHz, while the Nyquist bandwidth is 0.55 MHz. The sample rate must be carefully chosen not to mix things. Here the sampled bandwidths will span 0–0.2 MHz and 0.3–0.5 MHz.

⁸Keep track of: dsp.rice.edu/cs for an overview of all compressive sampling developments.

2.3 Modulation and Chopping

2.3.1 Modulation

Sampling of signals resembles modulation of signals. In both cases the operation results in the creation of frequency shifted bands of the original signal. A modulator multiplies the baseband signal with a sine wave, resulting in a pair of upper bands around the carrier frequency, see Fig. 2.12. Mathematically modulation is the multiplication of a signal with a pure sine wave of radial frequency ω_{local} :

$$G_{mix}(t) = A(t) \times \sin(\omega_{local}t) \quad (2.17)$$

In the simple case of $A(t) = A \sin(\omega t)$:

$$A \sin(\omega t) \times \sin(\omega_{local}t) = \frac{A}{2} \cos((\omega_{local} - \omega)t) - \frac{A}{2} \cos((\omega_{local} + \omega)t) \quad (2.18)$$

In the result there are no components left at the input frequencies. Only two distinct frequencies remain. This modulation technique is the basis for the first radio transmission standard: amplitude modulation.

If $A(t)$ is a band-limited spectrum composed of many sinusoidal signals, mixing results in two frequency bands:

$$G_{mix}(t) = A(t) \times \sin(\omega_{local}t)$$

$$G_{mix}(\omega) = \frac{1}{2}A((\omega_{local} - \omega) - \frac{1}{2}A(\omega_{local} + \omega)$$

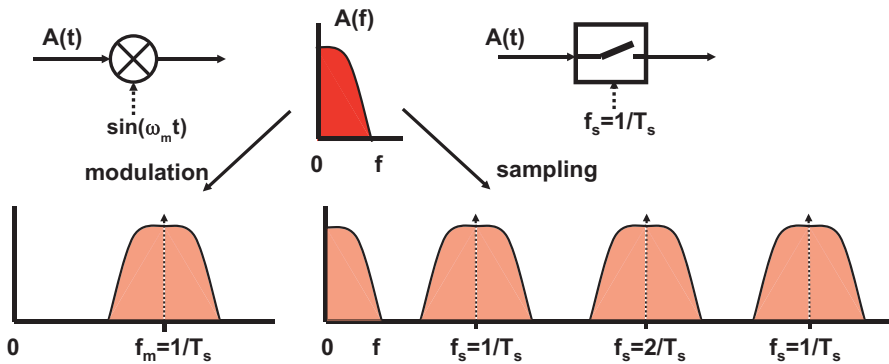


Fig. 2.12 Modulation and sampling of signals. Ideally the modulation and sampling frequencies disappear from the resulting spectrum. Here they are indicated for reference

The modulated bands appear as mirrored copies of each other around ω_{local} . Often one band is desired and the other band is referred to as the “mirror image.”

If the modulation principle is repeated the original sine wave can be recovered:

$$\begin{aligned}
 G_{mix-down}(t) &= G_{mix}(t) \times \sin(\omega_{local}t) \\
 \left(\frac{A}{2} \cos((\omega_{local} - \omega)t) - \frac{A}{2} \cos((\omega_{local} + \omega)t) \right) \times \sin(\omega_{local}t) &= \\
 \frac{A}{2} \sin(\omega t) - \frac{A}{4} \sin(2\omega_{local}t + \omega t) + \frac{A}{4} \sin(2\omega_{local}t - \omega t) & \quad (2.19)
 \end{aligned}$$

The original component is accompanied by a pair of frequencies around $2\omega_{local}$. With a low-pass filter these components are removed.

In contrast to modulation, sampling results in upper bands around every multiple of the sample rate. The sequence of Dirac pulses is equivalent to a summation of sine waves with frequencies at multiples of the sample rate.

$$\begin{aligned}
 \text{Time domain } \sum_{n=-\infty}^{n=\infty} \delta(t - nT_s) &= \frac{1}{T_s} \sum_{k=-\infty}^{\infty} e^{jk2\pi f_s t} = \frac{1}{T_s} + \frac{2}{T_s} \sum_{k=1}^{\infty} \cos(k2\pi f_s t) \\
 \text{Frequency domain } \mathbf{D}_s(\omega) &= \frac{2\pi}{T_s} \sum_{k=-\infty}^{k=\infty} \delta(\omega - \frac{2\pi k}{T_s}) \quad (2.20)
 \end{aligned}$$

This sequence of Dirac functions in the frequency domain translates back to sine waves in the time domain with frequencies that are integer multiples of the sample rate. Therefore sampling can be viewed as a summation of modulations. The intrinsic similarity between sampling and modulation is used in various system architectures: an example is found in down-mixing of radio frequency signals.

A particular aspect of sampling and mixing is called “self-mixing.” A mixer can be seen as a device with two (mathematically) equivalent input ports. If a fraction of the signal on one port leaks into the other port, self-mixing will occur. If this leakage is described as $\alpha \sin(\omega_{local}t)$, the resulting output component will contain terms of the form: $\alpha/2 + \sin(2\omega_{local}t)/2$. In practical circuits mostly the large-amplitude local-oscillator frequency will leak into the low-amplitude port or the sample frequency is injected on the input node. A noticeable DC component is the result that can easily be mistaken for a circuit offset.

2.3.2 Down-Sampling

In the description of signals in the previous paragraphs, implicitly the band of interest was assumed to be a baseband signal, starting at 0 Hz with a bandwidth $f = BW$. This is the situation that exists in most data-acquisition systems. The mirror bands will appear around the sample rate and its harmonics. The choice for

this location of the band of interest is by no means obligatory and certainly not imposed by the Nyquist criterion. A band of interest located at a higher frequency, or even beyond the sample rate, can be sampled equally well. The signal band can be regarded as being sampled by the closest multiple of the sample rate. This band is again copied to all integer multiples of the sample rate, including “0 Hz”. This process is called “under-sampling” or “down-sampling.”⁹ If there are components of the signal lying on equal frequency spacings above and below a multiple of the sample rate, both of these will be sampled into the same frequency region. The consequence is an overlap of signals and must be avoided.

Deliberate forms of down-sampling are used in radio-communication applications, where down-sampling is used as a way to perform demodulation, see Fig. 2.13.

Unwanted forms of down-sampling occur if undesired signals are present in the signal band. Examples are

- Harmonic distortion products of the baseband signal.
- Thermal noise in the entire input band, see Sect. 2.5.1.
- Interference signals from other parts of the equipment or antenna.

Example 2.7. Set up a down-sampling scheme for an FM-radio receiver.

Solution. In an FM-radio the IF signal of 100 kHz is located at a carrier frequency of 10.7 MHz. This signal can be down-modulated and sampled at the same time by a 5.35 Ms/s signal, Fig. 2.13.

Essentially any sample rate that fulfills $f_s = 10.7/i$ Ms/s, where i is a positive integer, will convert the modulated band from 10.7 MHz to DC.

Two dominant considerations play a role in the choice of the sample rate. A low sample rate results in a low power consumption of the succeeding digital circuitry and less memory if storage is needed. A high sample rate creates a wide empty frequency range and allow easy and cheap alias filtering. Sometimes the sample rate

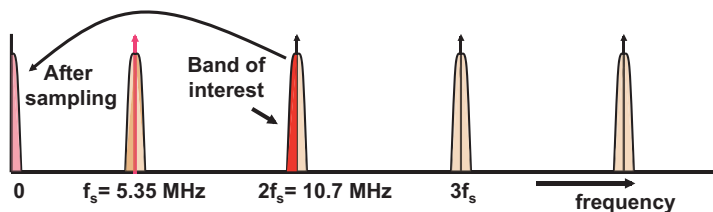


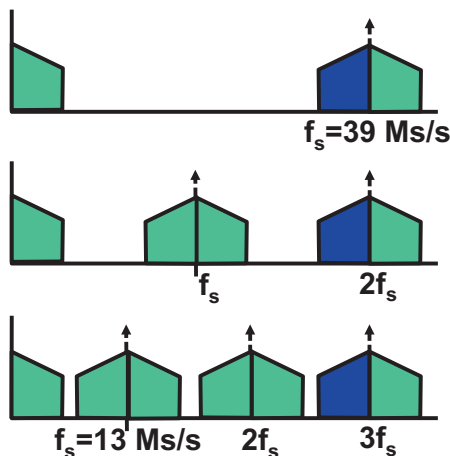
Fig. 2.13 Demodulation and down-sampling of an IF-FM signal at 10.7 MHz by a 5.35 Ms/s sample pulse

⁹In this book the term down-sampling is used for sampling an analog signal with the purpose to perform a frequency shift of the band of interest. Subsampling in Sect. 2.3.3 removes samples in a predetermined manner from an existing sample stream, but does not change the signal band.

can be chosen in such a manner that undesired input frequencies end up in an unused part of $f = 0, \dots, f_s/2$.

Example 2.8. An IF-television signal occupies a bandwidth from 33 to 39 MHz. Propose a sampling frequency that maps the 39 MHz component on DC. Consider that the power consumption of the following digital circuit is proportional to the sampling frequency and must be low.

Fig. 2.14 Three solutions for sampling a bandwidth between 33 and 39 MHz



Solution. A sampling rate of 78 Ms/s misses the point in the Nyquist criterion, as the bandwidth is only 6 MHz and not 39 MHz. The Nyquist rate is 12 Ms/s so alternatives are shown in Fig. 2.14. A sample rate of 39 Ms/s will work, but causes a lot of digital power.

A sample rate of 19.5 Ms/s is a viable alternative to 39 Ms/s as it halves the digital power but leaves enough frequency space for alias filtering.

And a sample rate of 13 Ms/s which will leave only a 1 MHz frequency range for alias filtering. This might be an expensive solution when the alias has to be suppressed.

2.3.3 Subsampling and Decimation

In some applications reducing the sample rate is a necessity. In sigma-delta conversion, decimation or subsampling is a necessity to translate the high-speed bit-stream signal in normal samples. Another example is in measuring the performance of a very high-speed sample system. This requires a test-setup with even better specifications at those frequencies. Here too, subsampling reduces the sampling frequency and allows to measure accurately, without the need for extreme performance.

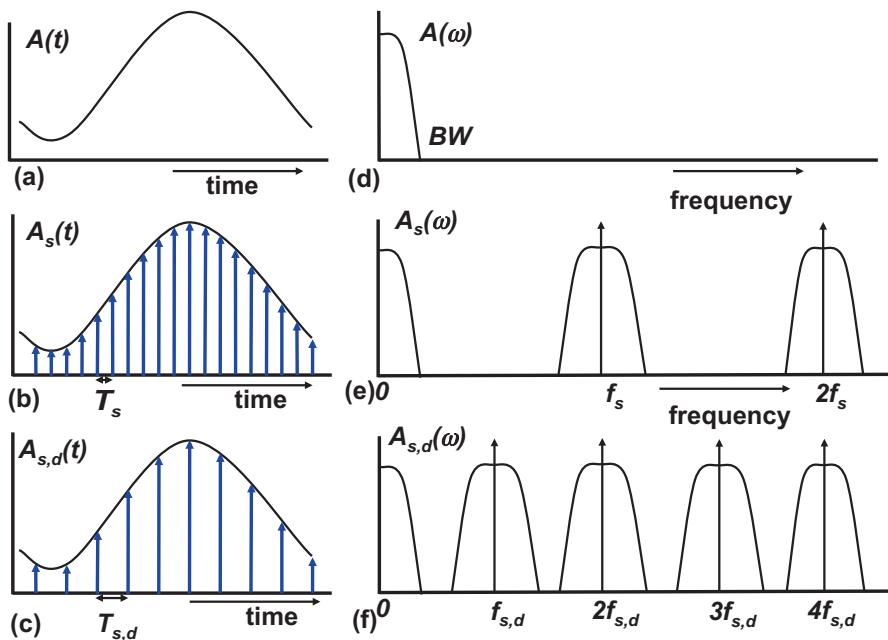


Fig. 2.15 Decimation or subsampling of a sampled signal. *Upper:* a time-continuous signal in the time (left) and frequency domain (right). *Middle:* Time and frequency representation after sampling. *Lower:* Time and frequency representation after subsampling by a factor of two

Figure 2.15 shows the basic process of subsampling or decimation. In Fig. 2.15a, b, d, e the time sequence and frequency representation of a signal sampled at f_s are shown. If this signal is subsampled by an integer factor¹⁰ M (in this figure $M = 2$) every M -th sample is kept and the unused samples are removed, see Fig. 2.15c, f.

The procedure in Fig. 2.15 can be used because the bandwidth of the original signal is less than half of the new sample rate. Thereby this signal fulfills the Nyquist criterion for the new sample rate f_s/M . In cases where this is not the case, the bandwidth must be sufficiently reduced, before subsampling is applied. If this is omitted, serious aliasing will occur with loss of information.

2.3.4 Chopping

Chopping is a technique used for improving accuracy by modulating error-sensitive signals to frequency bands where the signal processing is free of errors, see also

¹⁰Subsampling by a rational factor (a division of two integers) or an irrational factor requires to calculate the signal at each new sample moment by interpolation of the existing samples. This technique is often applied in image processing and is used to combine data from sources with asynchronous clocks. Some fast-running hardware is needed to carry out the interpolation.

Sect. 7.6. In Fig. 2.16 first the signal is modulated to a higher frequency band by multiplication with a chopping signal $f_{chop}(t)$. After signal processing, the signal is modulated back by multiplying again with $f_{chop}(t)$. The technique works well with a sine wave or a block wave as modulator as $f_{chop}^2(t)$ contains a DC-term and for the rest only frequency components far above the band of interest. Chopping can also be used to move unwanted signals out of the band of interest. For example, alternating between DC-current sources (also known as dynamic element matching) will move mismatch and the $1/f$ noise to higher bands.

In differential circuits, chopping is implemented easily by alternating between the differential branches. Mathematically this corresponds to a multiplication with a block wave with amplitude $+1, -1$. This block wave can be decomposed into a series of sine waves:

$$f_{chop}(t) = \sum_{n=1,3,5,\dots}^{\infty} \frac{4 \sin(n\pi/2)}{n\pi} \cos(\omega_{chop} t) \quad (2.21)$$

Now $f_{chop}^2(t) = 1$ and a perfect restoration after chopping back is possible. Note that a signal $f_{chop}(t)$ composed of any sequence of $+1, -1$ transitions, at fixed frequency or at arbitrary time moments, shows this property and can be used for chopping purposes. As Fig. 2.17 shows, chopping with a block wave can be done with lower frequencies than the bandwidth of the input signal. Chopping does not compress the signal into one single value, and consequently there is no direct impact of the Nyquist criterion on chopping. On the other hand, chopping is a form of modulation and so any unwanted signals entering the chopping chain between the two modulators may cause problems and alias filtering is a remedy.

The spectrum of a block wave fixed-frequency chopped signal will be composed of a series of modulated spectra around odd multiples of the chopping frequency:

$$A_{chop}(\omega) = \sum_{n=1,3,5,\dots}^{\infty} \frac{4 \sin(n\pi/2)}{n\pi} A(n\omega_{chop} \pm \omega)$$

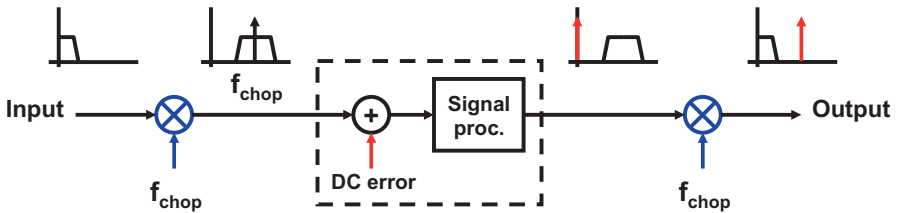


Fig. 2.16 A simple chopping chain: the input signal must be protected against the unwanted DC-term. Two chopping modulators move the signal band to a high frequency and back, thereby avoiding the DC term. This configuration was a standard technique in precision electronics of the 1950s

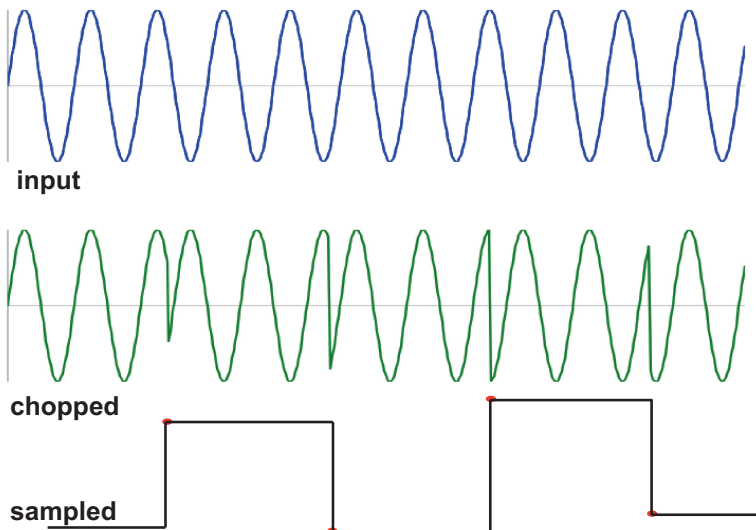


Fig. 2.17 Chopping (*middle*) and sampling (*below*) differ fundamentally in their information contents

E.g. chopping a spectrum from 0 to 1 MHz with a block wave (+1, −1) of 10 MHz, will remove the spectrum near DC and generate mirror bands at 9–11, 29–31, 49–51 MHz, etc.

The higher bands of the chopped signal should not be removed. This would cause imperfections after chopping back. Any removed components can be regarded as a negative addition of signals to a perfectly chopped spectrum. These components will be treated as new input signals for the chopping back operation. So products of these components with the signal of Eq. 2.21 will appear. The removed parts of the spectrum $A(n\omega_{chop} \pm \omega)$ will be modulated by the n -th harmonic of Eq. 2.21, resulting in an amplitude contribution at the position of the original signal with a relative strength of $1/n^2$.

Example 2.9. A 135 MHz sine wave is sampled in a 150 Ms/s sampling system. Which frequency components will be in the sampled data spectrum? Is it possible to discriminate the result of this sampling process from sampling a 15 MHz sine wave?

Solution. If an input signal at frequency f_i is sampled by a sample rate f_s then the sample data spectrum will contain the lowest frequency from the series: $f_i, (f_s - f_i), (f_s + f_i), (2f_s - f_i), (2f_s + f_i), \dots (nf_s - f_i), (nf_s + f_i), \dots$ where $n = 0, 1, 2, \dots, \infty$. In this case the second term delivers a 15 MHz component.

If directly a 15 MHz sine wave was sampled the sampled data sequence would be similar and even perfectly identical provided that the mutual phase shift is correct. In perfect conditions there is no way to tell from which time-continuous signal (in this case 15 or 135 MHz) this sequence originates. Continued in Example 2.16 on page 40.

2.4 Reconstruction of Sampled Data

The sequence of samples (after analog-to-digital conversion and any form of digital signal processing) that arrives at the input of a digital-to-analog converter is a set of numerical values corresponding to the frame of sample moments. A spectral analysis would result in an ideal sampled data spectrum, where all copies of the signal band at multiples of the sample rate are equivalent. In the time domain the value of the signal between the sample moments is (mathematically spoken) not defined.

This stream of samples must at some instant be reconverted in the time-continuous domain. The first question is what to do with the lacking definition of a signal in between the samples. The most common implementation to deal with this problem is to simply use the value of the signal at the sample moment and to keep it for the entire sample period. Figure 2.18 (left) shows this “zero-order hold” mode. A more sophisticated mechanism interpolates between two values as in Fig. 2.18 (middle). An elegant form of interpolation uses higher-order or spline-fit algorithms, Fig. 2.18 (right).

In most digital-to-analog converters a zero-order hold function is sufficient because the succeeding analog filters perform the interpolation. Moreover a zero-order hold operation is often for free as the digital input signal is stored during the sample period in a set of data latches. The conversion mechanism (ladders or current sources) simply converts at any moment whatever value the data latches hold. This option has several additional advantages. Whenever the output signal of a digital-to-analog converter contains glitches, an explicit sample-and-hold circuit will remove the glitches and improve the quality of the conversion. In case of algorithmic digital-to-analog converters the output signal has to be constructed during the sampling period (see, e.g., Sect. 7.4.4). Then a hold circuit shields any incomplete conversion results and prevents them to appear at the output.

Holding the signal during a period $T_h \leq T_s$ changes the shape of the signals passing through a zero-order hold operation. Holding of the signal creates a signal transfer function. The impulse response of the hold transfer function is found by considering that the Dirac sequence is multiplied by a function consisting of a constant term “1” over the hold period T_h .

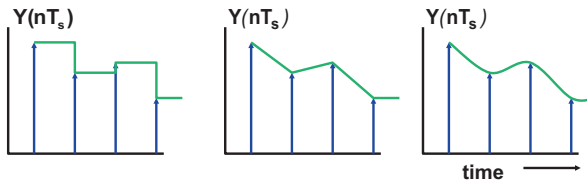


Fig. 2.18 A Dirac sequence can be reconstructed via a zero-order hold (*left*), a first-order interpolation (*middle*) or higher-order reconstruction algorithms (*right*)

$$h(t) = \begin{cases} 1, & 0 < t < T_h \\ 0, & \text{elsewhere} \end{cases} \quad (2.22)$$

The frequency domain transfer function $H(\omega)$ of a zero-order hold function (implemented in Chap. 3 as a sample-and-hold circuit) is calculated via the Fourier transform. The result of this transform has the dimension time.¹¹ In order to obtain a dimensionless transfer function, a normalization to T_s is introduced:

$$H(\omega) = \int_{t=0}^{t=\infty} h(t) \times e^{-j\omega t} dt = \int_{t=0}^{t=T_h} 1 \times e^{-j\omega t} dt = \frac{\sin(\pi f T_h)}{\pi f} e^{-j\omega T_h/2} \Leftrightarrow \frac{\sin(\pi f T_h)}{\pi f T_s} e^{-j\omega T_h/2} \quad (2.23)$$

Figure 2.19 shows the time and frequency response of the zero-order hold function for various values of the hold time T_h . The mathematical formulation of the amplitude function is often summarized to “ $\sin(x)/x$ ” behavior. Some authors use “ $\text{sinc}(x)$ ”. The integral of the function $\sin(x)/x$ belongs to the mathematical class of Dirichlet integrals, with as property:

$$\int_{x=0}^{\infty} \frac{\sin(x)}{x} dx = \pi/2 \quad (2.24)$$

The last term in Eq. 2.23 is $e^{-j\omega T_h/2}$ which represents a delay in the time domain. This delay $T_h/2$ is introduced as the value of the signal that was first concentrated in the sample moment is now distributed over the entire hold period. The average value moves from the sampling moment (the edge of the clock pulse) to the middle of the hold period.

A zero response occurs at frequencies equal to multiples of the inverse of the hold time. Obviously signals at those frequencies complete one or more complete periods in the hold time and exactly average out. For short hold periods approximating a Dirac function, this zero moves to infinity and the transfer of the sample-and-hold circuit is flat over a large frequency range. If T_h becomes equal to the sample period T_s the transfer function shows a zero at the sample rate and its multiples.

The amplitude response in the frequency domain is a representation of the average energy over (theoretically) infinite time. In the time domain sample values and consequently zero-order hold values can occur with amplitudes equal to the maximum analog input amplitude. A signal close to half of the sampling rate can show in one time period a small amplitude while achieving a value close to the full range input signal at another time instance depending on the phase relation of the signal and the sample rate. Still this signal has over infinite time an averaged

¹¹Formally the result of a Fourier transform reflects the intensity of a process or signal at a frequency. Therefore the result has the dimension “events per Hz” or “Volt per Hertz.”

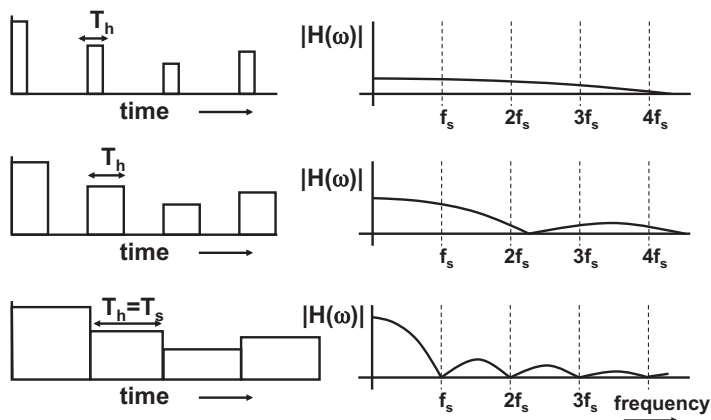


Fig. 2.19 The hold time determines the filter characteristics of a sample-and-hold function

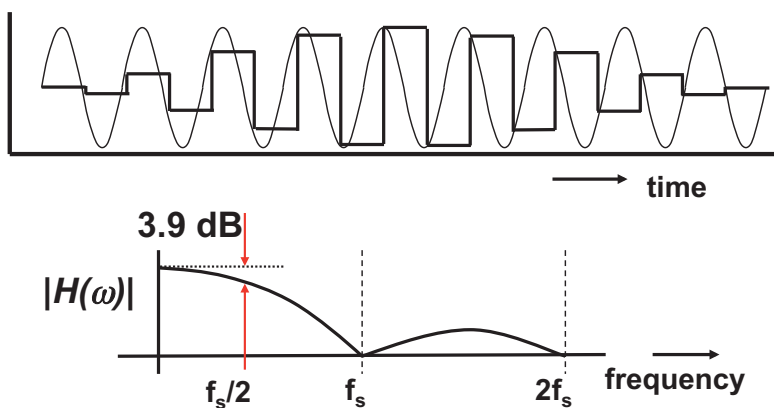


Fig. 2.20 Time and frequency response of a sample-and-hold signal close to half of the sampling rate

attenuation of 3.9 dB. In that sense the attenuation in Fig. 2.20 is different from a frequency transfer function of, e.g., an R-C network, where the attenuation at a certain frequency is independent of the phase.

Example 2.10. Can the $\sin(x)/x$ frequency behavior of a zero-order hold circuit be compensated by a high-pass analog filter?

Solution. In the frequency domain the amplitude loss can (partially) be compensated by means of a first-order high-pass filter, e.g., a voltage divider of two resistors, where the top resistor is shunted with a capacitor. The transfer characteristics are

$$|H_{\text{high-pass}}(f)| = \sqrt{1 + 4\pi^2 f^2 \tau^2} = 1 + 2\pi^2 f^2 \tau^2 - \pi^4 f^4 \tau^4 / 2 + \dots$$

$$|H_{\text{zero-order}}(f)| = \frac{\sin(\pi f T_s)}{\pi f T_s} = 1 - \pi^2 f^2 T_s^2 / 3! + \pi^4 f^4 T_s^4 / 5! - \dots$$

With $\tau = T_s / \sqrt{12}$ both functions will compensate for low frequencies, as both second terms add up to zero. However, beyond $f_s/2$ the time-continuous nature of the high-pass filter and the zero-order hold function will no longer match. Moreover the frequency response is an average over infinite time, and the instantaneous time-domain response will show at certain phase relations large excursions. Finally in such a setup the high-frequency noise will be amplified.

Example 2.11. Derive the transfer function for a first-order hold function in Fig. 2.18 (middle).

Solution. The transfer is now:

$$y(t) = x(n-1)T_s + (x(nT_s) - x(n-1)T_s) \frac{t}{T_s}, \quad nT_s \leq t \leq (n+1)T_s$$

If the time shift $e^{j\omega T_s}$ is ignored, the Fourier transform leads to a frequency domain representation for the transfer function:

$$\int_{t=0}^{t=T_s} e^{-j\omega T_s} e^{-j\omega t} dt + \int_{t=0}^{t=T_s} (1 - e^{-j\omega T_s}) \frac{t}{T_s} \times e^{-j\omega t} dt = T_s \left(\frac{1 - e^{j\omega T_s}}{j\omega T_s} \right)^2 \quad (2.25)$$

Rearranging the terms, extracting the time delay $e^{-j\omega T_s/2}$ and normalizing with T_s yields for the amplitude function:

$$|H(\omega)| = \left(\frac{\sin(\pi f T_s)}{\pi f T_s} \right)^2 \quad (2.26)$$

The first-order reconstruction leads to a better suppression of higher-order aliases.

2.5 Noise

In the previous section the sampling process was analysed from a mathematical perspective. In microelectronics a sampling circuit is realized with a switch and a capacitor. And with that, the standard problems of physical implementation start.

2.5.1 Sampling of Noise

Figure 2.21 and Table 2.2 show an equivalent schematic of the basic sampling circuit consisting of a switch and a storage capacitor. Compared to the ideal situation two non-ideal elements have been added to the switch: the switch resistance R combining all resistive elements between source and capacitor. The resistor is impaired with thermal noise,¹² consequently a noise source is added e_{noise} whose spectrum reaches far¹³ beyond the sampling rate of the switch.

$$e_{noise} = \sqrt{4kTRBW} \tag{2.27}$$

with Boltzmann’s constant $k = 1.38 \times 10^{-23} \text{ m}^2\text{kg s}^{-2} \text{ K}^{-1}$ and the absolute temperature T in Kelvin. This formulation expresses the noise in the positive frequency domain from 0 to ∞ .

When the switch connects to the capacitor, a low-pass filter is formed by the resistor and the capacitor. The average noise energy on the capacitor is therefore a filtered version of the noise energy supplied by the resistor and is filtered by the complex conjugated transfer function of the RC network. Using standard defined integral tables:

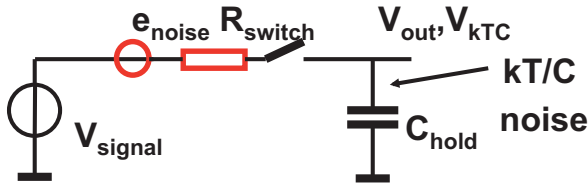


Fig. 2.21 Switched capacitor noise sampling: the series resistances act as a noise source

Table 2.2 kT/C noise for various capacitor values

C_{hold}	$V_{kTC} = \sqrt{kT/C_{hold}}$ at $T = 300^\circ\text{C}$
10 fF	$650 \mu\text{V}_{rms}$
100 fF	$204 \mu\text{V}_{rms}$
1 pF	$65 \mu\text{V}_{rms}$
3 pF	$35 \mu\text{V}_{rms}$
10 pF	$20.4 \mu\text{V}_{rms}$
30 pF	$11.8 \mu\text{V}_{rms}$

¹²Thermal noise in electronics is modeled as a noise source connected to a real impedance. For circuit calculations this works, but impedances are not noisy, electrons with random energy are.

¹³As thermal noise is an atomic phenomenon, its frequency span ends where sub-atomic mechanisms, as described by quantum physics, start. A rule of thumb limits the standard noise spectrum at 1 THz.

$$v_{C,noise}^2 = \int_{f=0}^{f=\infty} \frac{4kTR df}{1 + (2\pi f)^2 R^2 C^2} = \frac{kT}{C} \Rightarrow v_{C,noise} = \sqrt{\frac{kT}{C}} \quad (2.28)$$

The simple and well-known expression for the noise on a capacitor is called¹⁴: “ kT/C -noise.” Comparing this result to the power of the sine wave $v_{signal}(t) = \hat{A} \sin(\omega t)$ over the time period $1/\omega$ results in the signal-to-noise ratio SNR:

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{\hat{A}^2/2}{kT/C} \quad (2.29)$$

The magnitude of the resistor (the origin of the noise) is not part of this first-order expression. On one hand, an increase of the resistor value will increase the noise energy proportionally, however, that same increase in resistor value will reduce the relevant bandwidth also proportionally.

The same result follows from classical thermodynamics. The equipartition theorem says that in thermal equilibrium, the thermal energy is equally distributed over all degrees of freedom. For a capacitor there is only one degree of freedom: its potential. Therefore energy contained in the thermal fluctuation of carriers $Cv_{C,noise}^2/2$, equals the thermal energy for one degree of freedom: $kT/2$. Solving the equation results again in Eq. 2.28. Obviously there is no resistor involved. In simple terms one can say that the thermo-energetic electrons on the capacitor plates will move every now and then to the voltage source and back again due to their thermal energies. So the voltage over the capacitor fluctuates with time. When the sample switch opens the charge situation freezes.

If the noise spectrum is sampled in Fig. 2.22, each multiple of the sampling frequency will modulate the adjacent noise back to the base band, where all the noise bands accumulate. The same happens to all other bands, thereby hugely increasing the impact of noise.

Equation 2.28 holds for the time-continuous case, where the switch is permanently conductive, but holds equally for the sampled situation. Although the signals look completely different in the time domain, both the time-continuous and the sampled noise signal have values taken from a normal distribution with a zero-value mean and a variance $v_{C,noise}^2 = kT/C$.

This kT/C noise can be interpreted as a flat spectrum in the band from DC to $f_s/2$ as long as the RC cut-off frequency largely exceeds the sample rate. The spectral power noise density (power per Hz) of kT/C noise in a sampled system is equal to kT/C over half of the sample rate:

¹⁴Very annoying “T” for absolute temperature as well as for fixed time periods.

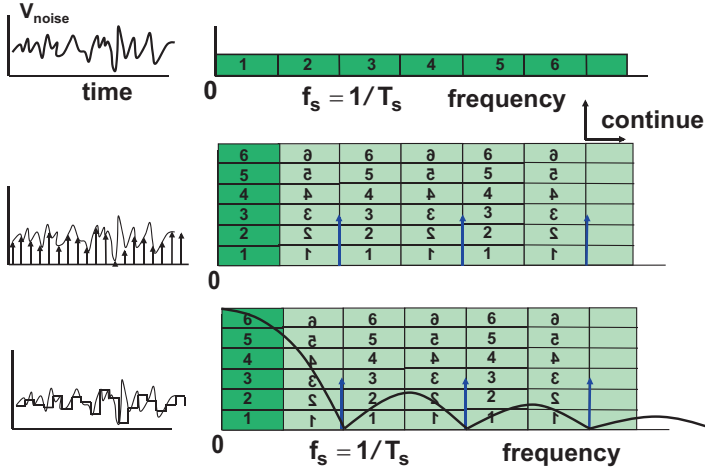
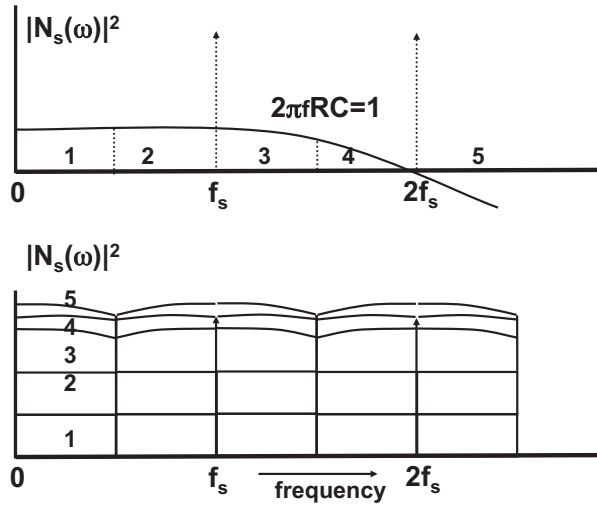


Fig. 2.22 Noise sampling, voltage on the capacitor: *left* in the time domain, *right* in the frequency domain

Fig. 2.23 Band-limited noise is sampled in a similar manner as normal signals. *Top*: this noise has a finite bandwidth, *bottom*: after sampling. The power spectra add up and are mirrored



$$S_{ff,SH} = \frac{2kT}{Cf_s} \quad (2.30)$$

If the RC cut-off frequency is low with respect to the sample rate, the noise bandwidth must be treated in a similar fashion as a normal signal band, see, e.g., Fig. 2.23. The power noise density of the time-continuous network with the same resistor and capacitor having a cut-off frequency of $f_{RC} = 1/2\pi RC$ in its pass-band is

$$S_{ff,RC} = 4kTR = \frac{2kT}{\pi C f_{RC}} \quad (2.31)$$

The comparison of the two noise densities in Eqs. 2.30 and 2.31 shows that in the sampling process the noise density increases by a factor $\pi f_{RC}/f_s$. This factor corresponds to the number of bands that stack up in Fig. 2.23. This considerable increase in noise density causes major problems when designing high-resolution converters.

The switching sequence can influence the total noise accumulated in the circuit. In switched capacitor circuits, every switch cycle will add one portion of kT/C noise. As these noise portions mostly are uncorrelated, they will sum in and root-mean-square way. Root-mean-square is the root of the effective power in a sum of signals. Also in situations where a switch discharges the charge of a capacitor into a fixed voltage or even ground potential, kT/C noise will appear (sometimes referred to as “reset-noise”).

This kT/C noise term presents a lower boundary in choosing the value for a sampling capacitance. An analog-to-digital converter is signal-to-noise limited because of this choice. A circuit with a total sampling capacitance of 1 pF will be limited by a noise voltage floor of $65 \mu V_{rms}$ at room temperature. A larger capacitance value will require IC area, more charging current and will directly impact the power budget.

Example 2.12. An uncorrelated white noise source with a total effective value of 1 mV_{rms} in the band limited to 120 MHz is sampled at 10 Ms/s. What is the noise density of the source? What is the noise density after sampling? What is the rms-value of the noise signal after sampling?

Solution. The effective noise level of 1 mV_{rms} means that the noise has an accumulated power equal to $(1 \text{ mV})^2$ over the impedance. That allows a calculation of the noise density of the noise source: $S_{vv} = (1 \text{ mV})^2/120 \text{ MHz}$. After sampling all noise bands higher than $f_s/2$ are folded back to the baseband. In this case the frequency range between DC and 5 MHz will contain 24 uncorrelated noise bands. The noise density is consequently: $S_{vv,s} = 24 \times (1 \text{ mV})^2/120 \text{ MHz}$. The total noise after sampling is found from integration of the noise density over the band of 5 MHz, yielding again an effective noise level of: 1 mV_{rms} . What about the noise in the band beyond 5 MHz? The noise density in those bands is equally high and real, but during the reconstruction process no more energy can be retrieved than what is available in one band.

Example 2.13. In a process with a nominal supply voltage of 1.2 V a sinusoidal signal of 100 MHz and $500 \text{ mV}_{peak-peak}$ is sampled. A SNR of 72 dB is required. Calculate the sampling capacitor and estimate the circuit power.

Solution. $500 \text{ mV}_{peak-peak}$ corresponds to a root-mean-square “rms” voltage of $500/2\sqrt{2} = 177 \text{ mV}_{rms}$. With a signal-to-noise ratio of $10^{72/20} = 4000$ (corresponding to a 12 bit ADC performance) the kT/C noise must be lower than $177 \text{ mV}_{rms}/4000 = 44 \mu V_{rms}$, and a minimum capacitor of 2.15 pF is needed.

A sinusoidal signal with a frequency of 100 MHz requires a current of $i = \omega \times C \times 500 \text{ mV}_{\text{peak-peak}} = 0.675 \text{ mA}_{\text{peak-peak}}$. This charge on the capacitor has to be supplied from an electronic circuit that allows only current in one direction, a bias current of, e.g., 1 mA can be used. Now the current swings from 162.5 to 837.5 μA . In first order this circuit requirement will consume 1.2 mW.

Example 2.14. A signal is sampled on two parallel connected equal capacitors: C_1, C_2 and $C_2 = C_1$. After sampling the capacitors are stacked in order to double the signal voltage, see Fig. 2.24. Does the signal-to-noise ratio change between the parallel and stacked connection?

Solution. After sampling a voltage V is stored on each capacitor. A noise contribution $v_{\text{noise}} = kT/(C_1 + C_2)$ is added and the signal-to-noise ratio is determined by the rms value of the signal over the noise. This noise gets “frozen” after the sample switch is opened, the same value holds for both capacitors, this is a rare situation where the noise is correlated. In first approximation the stacked capacitor construction will double the signal and its rms value, but the noise contribution on both capacitors doubles too, as these are correlated. The signal-to-noise ratio remains the same.

Ready? Not so fast, time for a second look!

There is another sampling moment that arises when the connection between both capacitors is opened in order to perform the stacking. Until that moment, the electrons with their thermodynamic energy can move freely between the capacitors. When the connection between the capacitors is broken, a new kT/C sampling event happens. This time the noise between the top plates will be equivalent to the noise of the series connection of C_1 and C_2 , which equals $2kT/C_1$ if the capacitors are equal. The voltage over every capacitor is kT/C_1 and both voltages are correlated, but with opposite sign. After stacking these two opposing noise contributions will cancel! And the signal-to-noise ratio will indeed be the same.

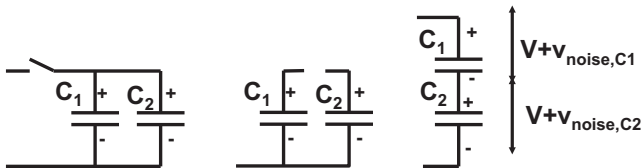


Fig. 2.24 A signal is sampled on two equal capacitors, after which the capacitors are stacked to double the signal

2.6 Jitter

2.6.1 Jitter of the Sampling Pulse

In the previous analysis it is assumed that the sample moments are defined with infinite precision. In practice all signals that define time moments have limited bandwidths, which means that there is no infinitely sharp rising edge. Oscillators, buffers, and amplifiers are all noisy devices [23–26], so consequently they add noise to these edges in Fig. 2.25. If noise changes the switching level of a buffer, the outgoing edge will have a varying delay with respect to the incoming edge. This effect is called: jitter. Jitter causes sample moments to shift from their position, and consequently the sampling circuit will sample the signal at another time moment. Next to noise-like components also signal-related components may influence the clock edge through limited power supply rejection, capacitive coupling, etc. Jitter from noisy sources will result in noise contributions to the signal, jitter from deterministic sources leads to tones (from fixed carriers) or to distortion (if the jitter source is correlated to the signal). Examples of systematic offsets in timing are: skews due to unequal propagation paths of clocks, interference from clock dividers, and clock doubling by means of edge detection. Random “jitter” variations occur not only during the generation of clock signals in noise-sensitive oscillators and PLLs, but also during transportation of timing signals jitter can be added, e.g., in long chains of clock buffers fed by noisy digital power supplies, capacitive coupling, and varying loading. A practical value for jitter on a clock edge coming from a digital CMOS environment¹⁵ is 30–100 ps_{rms}. If an advanced generator is used in combination with bandpass filters, the sampling pulse jitter can be reduced to levels below 100 fs_{rms}. The contributions of dedicated high-power on-chip circuits can be

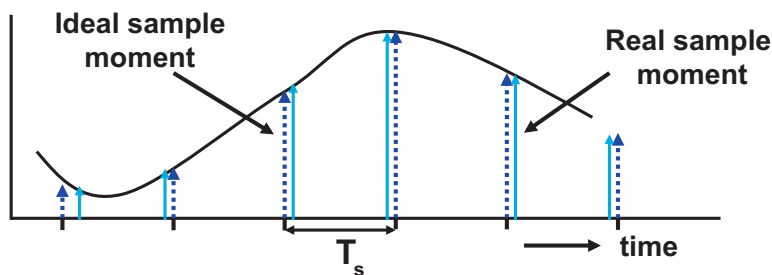
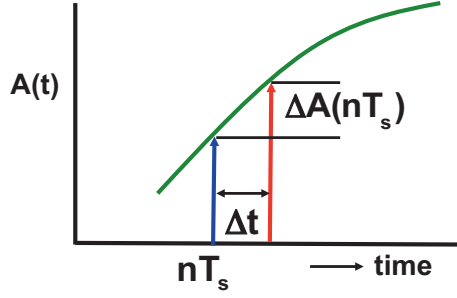


Fig. 2.25 The ideal sampling moments (*dashed*) shift in an arbitrary fashion in time if the sample clock is disturbed by jitter

¹⁵ A peak–peak value is often used for jitter, but peak–peak values for stochastic processes have no significance if the process and the corresponding number of observations are not identified.

Fig. 2.26 The ideal sampling moment is affected by jitter and an amplitude error occurs



brought back to a similar level. For example, Ali et al. [219] reports an overall jitter of 83 fs_{rms}.

The above description specifies the cycle-to-cycle deviation. In some systems a long-term jitter component can be relevant, e.g., for the display of a signal on a screen via a scanning mechanism, the jitter between two samples in the scan direction is determined by cycle-to-cycle jitter, while two samples arranged above each other are given by a long-term jitter. In monitors these samples can be some 1000 clock cycles apart. This jitter is specified over a longer period and requires some extensions of the following analysis. Some more in the discussion of phase-noise in Sect. 2.6.2.

Figure 2.26 shows the effect of shifting a sample moment. If a sinusoidal signal $A(t) = \hat{A} \sin(\omega t)$ with a radial frequency ω is sampled by a sample pulse with jitter, the new amplitude and the amplitude error are estimated as:

$$A(nT_s + \Delta t(t)) = \hat{A} \sin(\omega \times (nT_s + \Delta t(t))) \quad (2.32)$$

$$\Delta A(nT_s) = \frac{d\hat{A} \sin(\omega t)}{dt} \times \Delta t(nT_s) = \omega \hat{A} \cos(\omega nT_s) \Delta t(nT_s) \quad (2.33)$$

The time error is a function of the time itself. The amplitude error is proportional to the slope of the signal $\omega \hat{A}$ and the magnitude of the time error.

If the time error is replaced by the standard deviation σ_t^2 describing the timing jitter variance, the standard deviation of the amplitude σ_A is estimated as:

$$\sigma_A^2(nT_s) = \left(\frac{dA(nT_s)}{dt} \right)^2 \sigma_t^2 = \omega^2 \hat{A}^2 \cos^2(\omega nT_s) \sigma_t^2 \quad (2.34)$$

Averaging this result over all values of nT_s gives a jitter error power of:

$$\sigma_A^2 = \frac{\omega^2 \hat{A}^2 \sigma_t^2}{2} \quad (2.35)$$

When the origin of the jitter is a flat spectrum as for thermal noise, this jitter noise will appear as a flat spectrum between 0 and $f_s/2$ and repeats at every higher band.

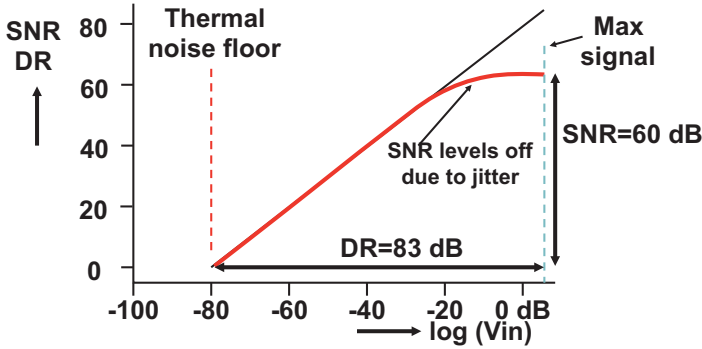


Fig. 2.27 With increasing signal amplitude over a fixed thermal noise level the signal-to-noise ratio increases. However when the amplitude is such that jitter becomes the dominant error, the signal-to-noise ratio flattens

Comparing this result to the power value of the sine wave $\hat{A}^2/2$ over the time period $T = 1/\omega$ results in the signal-to-noise ratio due to jitter:

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{jitter}}} = \frac{\hat{A}^2/2}{\sigma_A^2} = \left(\frac{1}{\omega \sigma_t} \right)^2 = \left(\frac{1}{2\pi f \sigma_t} \right)^2 \quad (2.36)$$

or in deciBel¹⁶ (dB):

$$\text{SNR} = 20^{10} \log \left(\frac{1}{\omega \sigma_t} \right) = 20^{10} \log \left(\frac{1}{2\pi f \sigma_t} \right) \quad (2.37)$$

For sampled signals the above relations hold for the ratio between the signal power and the noise in half of the sampling band: $0 \dots f_s/2$. This simple relation estimates the effect of jitter, assuming no signal dependencies. Nevertheless it is a useful formula to make a first order estimate. For wide-band signals with a uniform power distribution between $0, \dots, f$ [27] gives a $3\times$ higher signal-power to noise ratio or a 4.8 dB more favorable jitter SNR. Note that the jitter power is independent of the sample rate, consequently the jitter power density (power per Hertz) is inversely related to f_s .

The linear dependence of jitter noise to the input frequency and to the signal amplitude often allows a rapid identification of jitter in a time-discrete system. For a given signal frequency the jitter power increases linearly with the amplitude, leading to the flattening of the SNR versus input amplitude curve [28], see Fig. 2.27.

¹⁶For some reason deciBel is spelled with single “l” although it was named after A.G. Bell. Similarly the letter “a” was lost in “Volta.”

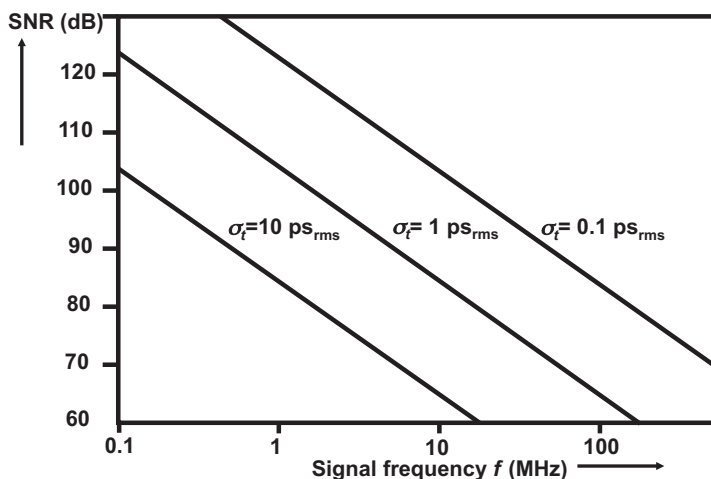


Fig. 2.28 The signal-to-noise ratio depends on the jitter of the sampling signal and the frequency of the time-continuous signal

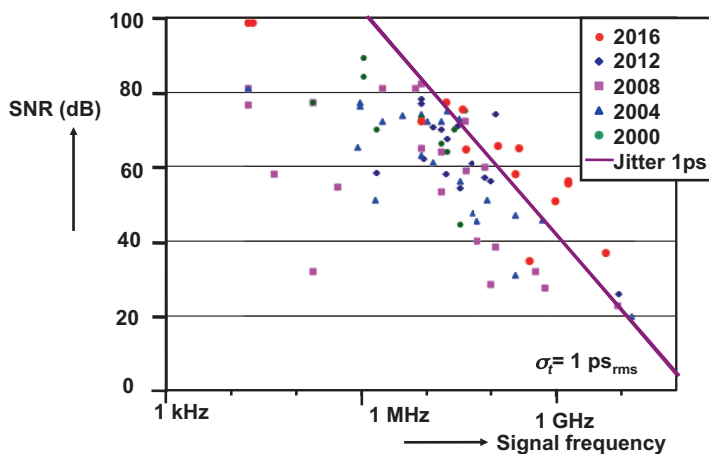


Fig. 2.29 The signal-to-noise ratio versus the signal frequency of analog-to-digital converters reported on the International Solid-State Circuits conferences in the years 2000, 2004, 2008, 2012, and 2016. (from: B. Murmann, “ADC Performance Survey 1997–2016,” Online: <http://web.stanford.edu/~murmann/adcsurvey.html>)

Figure 2.28 shows the signal-to-noise ratio as a function of the input frequency for three values of the standard deviation of the time jitter.

Figure 2.29 compares the jitter performance of analog-to-digital converters published on the International Solid-State Circuits conferences in the years 2000, 2004, 2008, and 2012. It is obvious that a jitter specification better than $\sigma_t < 1$ ps

Table 2.3 Jitter specifications of some commercially available parts

Part	Description	Jitter
“2011”	Quartz 50–170 MHz	3 ps _{rms}
“8002”	Programmable oscillator	25 ps _{rms}
“1028”	MEMS+PLL combi 100 MHz	95 ps _{rms}
“6909”	RC oscillator 20 MHz	0.2 %
“555”	RC oscillator/timer	>50 ns _{rms}

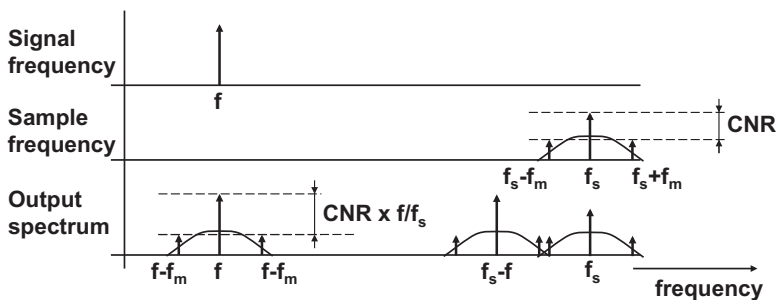


Fig. 2.30 Jitter around the sampling frequency will produce side spectra around the input tone

is a challenge. The comparison of the best converters in every year shows that little progress was made over the last decade.

Table 2.3 indicates some jitter numbers from commercial timing components.

From a spectral point of view, the jitter spectrum modulates the input tone. Therefore the jitter spectrum around the sampling pulse will return around the input frequency as in Fig. 2.30. Translated to a lower frequency the time error due to jitter will produce a proportionally smaller amplitude error. Therefore the carrier-to-noise ratio (CNR) improves.

Equation 2.35 has been derived for a single sine wave as a signal. In communication systems (from ADSL to 5G) multi-tone signals are applied. These signals contain large number of carriers N_c up to 1024. All carriers behave as independent sine waves, and the probability that all carriers are at a maximum is far below the system error level of $10^{-4} - 10^{-6}$, as is typical for these systems. Instead of allowing an individual carrier amplitude of just A/N_c , more commonly the amplitude of the i -th carrier is approximated by Zogakis and Cioffi [29] and Clara and Da Dalt [30]

$$\hat{A}_{c,i} = \frac{\hat{A}}{C_F \sqrt{N_c}}$$

C_F is the so-called crest-factor: the ratio between the maximum signal and the rms value. In ADSL $C_F \approx 5.6$. The jitter error power per carrier is given by Eq. 2.35. Summing the power over all N_c carriers yields

$$\sum_{i=1}^{i=N_c} \sigma_{Ac,i}^2 = \sum_{i=1}^{i=N_c} \frac{\omega_i^2 \hat{A}^2 \sigma_t^2}{C_F^2 N_c} \approx \frac{\omega_{middle}^2 \hat{A}^2 \sigma_t^2}{C_F^2} \quad (2.38)$$

In this coarse approximation ω_{middle} is the frequency of the tone at $i = N_c/2$. Obviously the jitter error power is far lower than in the sine wave case. More accurate analysis is found in [29, 30].

If jitter is caused by delay variations in digital cells as shown in Fig. 2.31, the jitter can also contain signal components and strong spurious components, e.g., linked to periodic processes in the digital domain. These contributions are demodulated similar as in Fig. 2.30 and are the source for spurious components and signal distortion. Therefore digital circuits that generate and propagate the sample pulse must be treated as if these were analog blocks.

Example 2.15. The clock buffer in Fig. 2.31 has edge transition times of 70 ps. How much jitter can be expected if a random noise of 60 mV_{rms} is present on the power supply of 1.2 V.

Solution. Due to voltage changes on the power supply lines, the currents inside the buffer will change, in first order proportional to the voltage change. As a consequence the slope of the transition will vary linearly. The mid-level switching point is now reached after 35 ps delay from the input mid-level passing. A voltage change of $60 \text{ mV}/1.2 \text{ V} = 5\%$ will create a 5% delay variation on the slope and the delay of 35 ps. So the expected jitter is 1.75 ps_{rms} per edge. As the same voltage variation applies to two inverters, the overall jitter is 3.5 ps_{rms} .

Example 2.16. In Example 2.6 the sample sequence is distorted by a random jitter component of 5 ps_{rms} . Is it possible to discriminate in the sampled data domain between a 15 MHz input sine wave or a 135 MHz input sine wave?

Solution. With perfect sampling both signals will result in equivalent wave forms in the sampled data domain. However, the presence of jitter allows to discriminate,

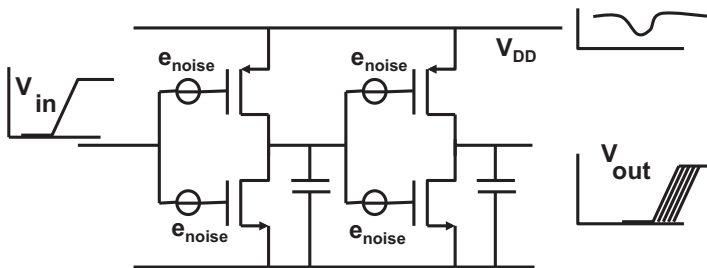


Fig. 2.31 A clock buffer for generating the digital sample signal can add to an ideal sample signal some noise of the buffer transistors. Also fluctuations on the power supply will affect the switching behavior of the buffer, causing uncertainty on the edges and jitter in the sampling

as the resulting SNR for a 15 MHz input signal is $SNR = 1/2\pi f_i \sigma_t = 66.5$ dB, while the SNR for 135 MHz equals 47.5 dB.

Example 2.17. Calculate the jitter due to thermal noise that an inverter with dimensions of NMOST 0.2/0.1 and PMOST 0.5/0.1 in a 90-nm CMOS process adds to an edge of 50 ps (bottom-top).

Solution. Every transistor adds noise that is related to the transconductance in the channel: $i_{noise} = \sqrt{4kTBWg_m}$. If the inverter is at its mid-level point (0.6 V) both transistors will be contributing to a total noise current of: $i_{noise,n+p} = \sqrt{4kTBW(g_{m,n} + g_{m,p})}$. This noise corresponds to an input referred noise voltage of $v_{noise,n+p} = i_{noise,n+p}/(g_{m,n} + g_{m,p}) = \sqrt{4kTBW/(g_{m,n} + g_{m,p})}$. With the help of the parameters in Table 4, an equivalent input noise voltage is found of 0.62 mV_{rms} in a 10 GHz bandwidth. This bandwidth is an approximation based on the observation that an rising edge of 50 ps followed by a similar falling edge limits the maximum frequency of the inverter to 10 GHz. The jitter order of magnitude is estimated as: $\sigma_t/\tau_{edge} = v_{noise,n+p}/V_{DD}$ and $\sigma_t = 25 \text{ fs}_{rms}$.

2.6.2 Phase-Noise and Jitter

For sampling systems the variation in time moments or jitter is an important parameter. Jitter is here described as a random time phenomena. In RF systems the same phenomenon is observed in the frequency domain and is called “phase-noise.” The events in oscillator and PLL spectra, such as in Fig. 2.32, are specified at the offset frequency with respect to the ideal oscillation frequency f_o . A spectrum of the signal from a phase-locked loop or oscillator circuit shows some typical components, see Fig. 2.32:

- White noise in the output (no dependency on the frequency).
- White noise that modulates the oscillator shows up in the power spectrum with a decreasing frequency slope in the oscillation offset frequency. $1/f$ noise generates an even faster decreasing slope in the spectrum with a low offset frequency.

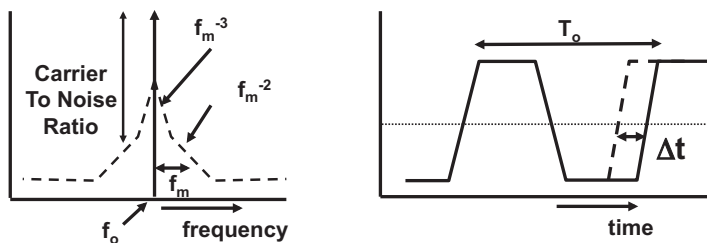


Fig. 2.32 Left: the frequency spectrum of an oscillator, right: time jitter

- PLLs multiply a reference frequency. Often spurious tones are visible on both sides of the generated frequency at an offset frequency equal to the reference frequency.
- Undesired tones entering the PLL via substrate coupling can modulate the output.

An instantaneous phase deviation $\Delta\theta$ offsets the zero-crossings of a sinusoidal signal of frequency ω_o to yield a time deviation. The time variation Δt for a radial frequency ω_o with a corresponding frequency period T_o , and the phase deviation $\Delta\theta = \omega_o \Delta t$ of the same signal, essentially describe the physical phenomena [28, 31]:

$$\frac{\Delta t}{T_o} = \frac{\Delta\theta}{2\pi} \quad (2.39)$$

From this instantaneous relation between time offset and phase offset, a first-order indication of the relation between jitter and phase-noise is obtained. Both originate from the same stochastic source. Now the time-domain offset Δt is replaced by its time-averaged variance: $\sigma_{t,rms}^2$. In order to obtain the phase-error variance $\sigma_{\theta,rms}^2$, the spectral noise density $S_{ff}(f)$ must be integrated over both side lobes to give the total equivalent phase noise power¹⁷:

$$\left(\frac{\sigma_{t,rms}}{T_o}\right)^2 = \left(\frac{\sigma_{\theta,rms}}{2\pi}\right)^2 = \frac{2 \int_{f_{low}}^{f_{high}} S_{ff}(f) df}{(2\pi)^2} \quad (2.40)$$

The span of integration is limited by a lower and higher boundary of the offset frequency. The integration cannot start at $\omega = \omega_o$ Hz due to the singularity in the spectral density. Leaving out the frequencies below 10^{-8} means ignoring 3-years repetitive effects, but more often a lower boundary is chosen in the Hz to kHz range. Obviously one should not expect a 99.9 % prediction level. The choice for f_{low} also depends on whether the cycle-to-cycle jitter is required or longer-term jitter variations. The energy in the low-frequency second-order lobes of the phase spectrum is responsible for the increase of long-term jitter over cycle-to-cycle jitter. In the extreme case of only white phase noise, the contribution of the low-frequency band would be negligible and the long-term jitter would be comparable to the cycle-to-cycle jitter. Translating various forms of phase-noise densities in time jitter clearly requires an assumption of spectral density function for the phase noise [23–26, 31].

Example 2.18. Calculate the jitter from the spectrum in Fig. 2.33.

¹⁷Here a strictly formal derivation requires some 10 pages, please check out specialized literature, e.g., [31].

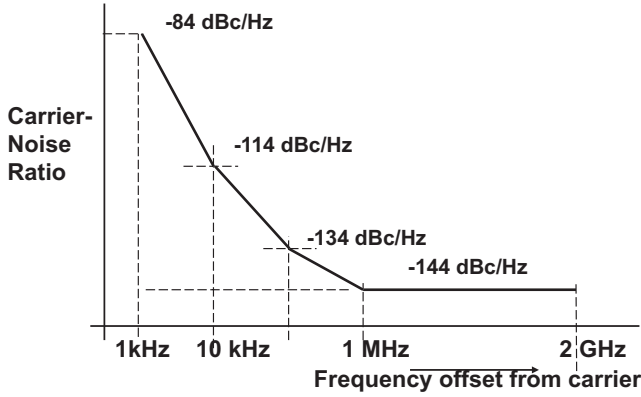


Fig. 2.33 The phase-noise spectrum of an oscillator signal at 2 GHz

Solution. f is the offset frequency from the carrier. The spectrum shows four typical regions: from $f = 1$ kHz to 10 kHz the slope is f^{-3} corresponding to the $1/f$ noise in the oscillator, the next section with a slope of f^{-2} is due to thermal noise in the oscillator. The two right most regions with slope f^{-1} and the floor are due to $1/f$ noise and thermal noise in buffers. This spectrum appears on either side of oscillator frequency f_0 . The level of the floor is given as -144 dBc/Hz. Corresponding to a power of $4 \cdot 10^{-15}$ of the carrier per Hz bandwidth. The curve is approximated with the following equation:

$$S_{ff}(f) = \left[1 + \frac{f_1}{f} + \left(\frac{f_2}{f} \right)^2 + \left(\frac{f_3}{f} \right)^3 \right] S_{floor}(f) \quad (2.41)$$

where $f_1 = 1$ MHz, $f_2 = 100$ kHz, $f_3 = 10$ kHz and the range of interest is limited from 1 kHz to 2 GHz. Formally the phase area under the curve is found by integration and substitution of the frequencies.

$$(f + f_1 \ln(f) + f_2^2/f + f_3^3/2f^2) S_{floor} \Big|_{f=1 \text{ kHz}}^{2 \text{ GHz}} \quad (2.42)$$

A coarse approximation allows to determine the contribution in each section of the curve, which gives an insight where optimization of the circuit is most beneficial:

$$\begin{aligned} & ((2 \cdot 10^9 - 10^6) + f_1 (\ln(10^6) - \ln(10^5)) + f_2^2 (1/10^4 - 1/10^5) \\ & + f_3^3 (1/10^3 - 1/10^4)/2) 4 \cdot 10^{-15} \\ & = (1999 \times 10^6 + 2.3 \times 10^6 + 0.9 \times 10^6 + 450 \times 10^6) 10^{-15} \\ & = 9.8 \times 10^{-6} \end{aligned}$$

With the help of Eq. 2.40 the time jitter is found: 0.35 ps_{rms} . This is a real coarse estimate, e.g., a popular spread sheet called the “Allen Variance” simply adds 3 dB to the noise density to compensate for underestimations. Note that the thermal noise and the $1/f$ noise in the oscillator dominate.

2.6.3 Optical Sampling

The $\sigma_t = 0.1, \dots, 10 \text{ ps}_{rms}$ range for jitter is typical for electronic design and ultimately linked to physical processes such as thermal noise and $1/f$ noise. Mode-locked lasers can generate pulse trains with 200 ps width and a jitter of approximately ten femtoseconds. Building sampling devices triggered by these lasers is a challenge, as the straight-forward solution to capture the laser pulses with diodes would immediately affect the performance. An alternative solution [32] uses GaAs finger structures and attributes $\sigma_t = 80 \text{ fs}_{rms}$ jitter to the sampling process.

2.7 Time-Discrete Filtering

Time-discrete filtering forms a subset of the time-discrete signal processing tool box, see, e.g., [16, 17] and can be found in oversampled digital-to-analog converters, Sect. 10.1, and in sigma-delta modulators, Sect. 10.4. Time-discrete filters play a role in the conversion architecture decisions as well as in the necessary post-processing.

2.7.1 FIR Filters

Sampled signals can easily be delayed in the time-discrete domain. In the analog time-discrete domain, switched capacitors transfer charge packets from one stage into another stage. By means of appropriate switching sequences various time delays are implemented. After amplitude quantization samples can also be delayed in the digital domain via digital delay cells, registers, and memories. Frequency filters in each domain are realized by combining the time-delayed samples with specific weighting factors or multiplying coefficients.

Operations and functions in the discrete-time domain are described in the z -domain. If $f(n) = f(nT_s)$, $n = 0 \dots \infty$ is a sequence of values corresponding to the sample value of $f(t)$ at points in time $t = nT_s$, this sequence can be described in the z -domain as:

$$f(z) = \sum_{n=0}^{n=\infty} f(n)z^{-n}$$

where z is a complex number in polar representation $z = re^{j\omega_z}$, which resembles the Laplace parameter $s = \alpha + j\omega$ with $r \leftrightarrow e^\alpha$, $\omega \leftrightarrow \omega_z$. The important difference is that the z -domain describes a sampled system where the maximum frequency is limited to half of the sample rate. While ω is expressed in rad/sec, $\omega_z \leftrightarrow \omega T_s$ is expressed in radians and abstracts from physical frequencies. The s -plane and the z -plane can be mapped on each other. Due to the polar description the $j\omega$ axis in the s -domain becomes a unity circle in the z -domain, with the DC point at $z = 1e^{j0} = 1$. Poles and zeros in left-side of the s -plane resulting in stable decaying exponential functions in the time domain move to the inner part of the unity circle in the z -domain, Fig. 2.34 (right).

A delay of one basic sample period is transformed into the function z^{-1} . A frequency sweep from 0 to $f_s/2$ results in a circular movement of the z vector in a complex plane from $+1$, via $0 + j$ to -1 . For the frequency range $f_s/2$ to f_s the z vector will turn via the negative imaginary plane and return to $z = 1$. Figure 2.35 shows two integrators described in the z domain. The left structure adds the present sample to the sum of the previous samples. After the next clock the output will equal that sum and a new addition is performed. The right topology does the same, here the sum is directly available. The transfer functions for both structures are

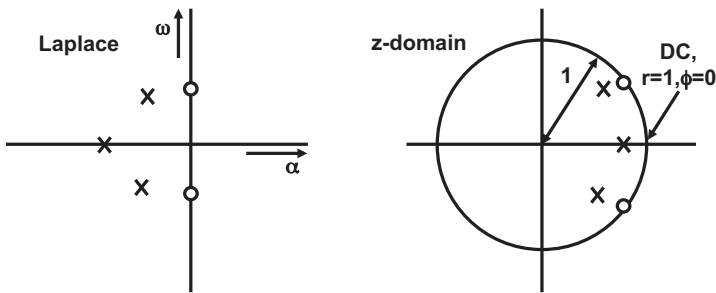


Fig. 2.34 The complex plane for the Laplace transform (s -plane) and the time-discrete plane (z -plane). A real pole, a pair of imaginary poles and a pair imaginary zeros are depicted

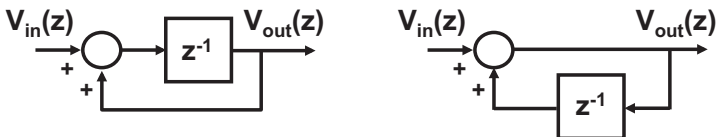


Fig. 2.35 Two integrators in the z -domain

$$H(z) = \frac{z^{-1}}{1 - z^{-1}} \quad H(z) = \frac{1}{1 - z^{-1}}$$

These integrator formulas indicate a mathematical pole at $z = 1$. The transform to the Laplace domain $z \leftrightarrow e^{sT_s}$ shows that $z = 1$ corresponds to $s = 0$ or DC conditions. And indeed a DC signal on an ideal integrator will lead to an unbounded output. Close to $z = 0$ the left integrator has zero output while in right-hand integrator just passes the signal.

The most simple filter in the time-discrete domain is the comb filter. The addition of a time-discrete signal to an m -sample periods delayed signal gives a frequency transfer that can be evaluated in the z -domain:

$$H(z) = 1 \pm z^{-m}$$

This function has zeros at all frequencies where $z^{-m} = \pm 1$, resulting in m zeros distributed over the unity circle. Using the approximation $z \leftrightarrow e^{sT_s}$ results in:

$$\begin{aligned} H(s) &= 1 \pm e^{-smT_s} = e^{-smT_s/2} (e^{+smT_s/2} \pm e^{-smT_s/2}) \\ |H(\omega)| &= 2|\cos(\omega mT_s/2)|, \quad \text{addition} \\ |H(\omega)| &= 2|\sin(\omega mT_s/2)|, \quad \text{subtraction} \end{aligned} \quad (2.43)$$

where the sign at the summation point determines whether the cosine response (with equal signs) or the sine response (with opposite signs) applies, see Fig. 2.36. In this plot the zeros are observed in the frequency domain.

Comb filters are mostly applied in systems where interleaved signals have to be separated. An example is the analog composite video signal, where the frequency carriers with the color information are interleaved between the carriers for the luminance signal.

The comb filter adds signals to their delayed versions. A more general approach uses a delay line where each delayed copy is multiplied with its own weight factor,

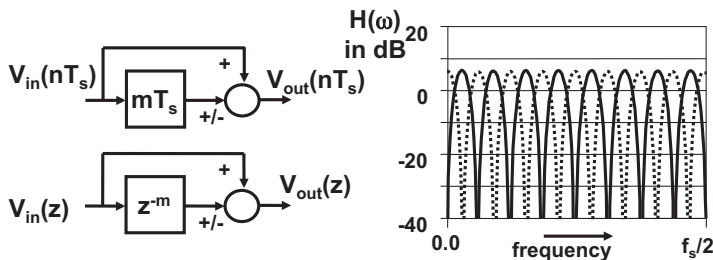


Fig. 2.36 The comb filter as sampled data structure and in the z -domain. The frequency response shows with a *solid line* the sine response (minus-sign at the summation), while the *dotted line* represents the cosine response (plus-sign)

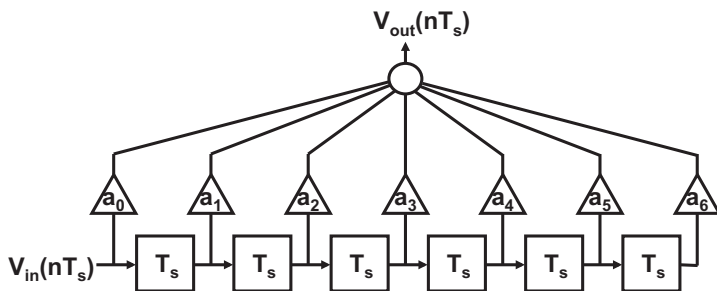


Fig. 2.37 The general structure of a finite impulse filter

see Fig. 2.37. A filter with this structure is known as a “Finite Impulse Response” filter (FIR-filter). The term “finite” means that any input disappears from the filter after passing through the N delay elements. In the summation the signals from the different delay elements can enhance or extinguish each other depending on the periodicity of the signal with respect to the delay time and the multiplication factor. The filter designer can adapt the filter characteristic through these multiplication coefficients or weight factors. Similar to the time-continuous filters the discrete-time transfer function defines the relation between input, filter transfer function, and output in the time domain with a convolution:

$$y(nT_s) = \sum_{k=0}^{k=\infty} h(k)x(nT_s - kT_s) \quad (2.44)$$

Applied to the filter in Fig. 2.37, this gives:

$$V_{out}(nT_s) = \sum_{k=0}^{k=N-1} a_k V_{in}((n-k)T_s) \quad (2.45)$$

An intuitive way of realizing what happens in an FIR filter is to imagine an endless row of delayed samples. Over this row a window defined by the FIR filter length is moved.¹⁸ The z-transform results in a description of the transfer of an FIR filter:

$$\frac{V_{out}(z)}{V_{in}(z)} = H(z) = \sum_{k=0}^{k=N-1} a_k z^{-k} \quad (2.46)$$

In order to transform this transfer function from the discrete time domain to the frequency domain, the term z^{-1} is substituted by $e^{-j\omega T_s}$ which results in:

¹⁸In financing a monthly moving average is a very simple FIR filter with 12 taps and simply “1” as multiplication factor.

$$H(\omega) = \sum_{k=0}^{N-1} a_k e^{-jk\omega T_s} \quad (2.47)$$

This time-continuous approximation is only applicable for a frequency range much smaller than half of the sample rate.

Some important properties of this filter are related to the choice of the weighting factors. Suppose the values of the coefficients are chosen symmetrical with respect to the middle coefficient. Each symmetrical pair will add delayed components with an average delay equal to the middle position. If the delay of each pair equals $NT_s/2$, then the total filter delay will also equal $NT_s/2$. The same arguments holds if the coefficients are not of equal magnitude but have an opposite sign (“anti-symmetrical”). This “linear phase” property results in an equal delay for all (amplified or attenuated) signal components and is relevant if the time-shape of the signal must be maintained, e.g., in quality audio processing.¹⁹

Mathematically the constant delay or linear phase property can be derived from Eq. 2.47 by substitution of the Euler’s relation²⁰:

$$e^{-j\omega T_s} = \cos(\omega T_s) - j \sin(\omega T_s) \quad (2.48)$$

After moving the average delay $NT_s/2$ out of the summation, real and imaginary terms remain:

$$H(\omega) = e^{-j\omega NT_s/2} \sum_{k=0}^{N/2-1} (a_k + a_{N-k}) \cos(k\omega T_s/2) - j(a_k - a_{N-k}) \sin(k\omega T_s/2) \quad (2.49)$$

Without violating the general idea, N has been assumed here to be even. If the coefficients a_k and a_{N-k} are equal as in the symmetrical filter the sine term disappears. The cosine term is removed by having opposite coefficients in an asymmetrical filter. Both filters have a constant delay. Depending on the symmetry and the odd or even number of coefficients the filters have structural properties, e.g., an asymmetrical filter with an even number of coefficients has a zero DC-transfer.

A filter that averages over N samples is designed with coefficients of value $1/N$. A simple transfer characteristic can be determined by hand for a small number of coefficients. More complex filters require an optimization routine. A well-known routine was proposed by McClellan, Parks, and Rabiner (MPR or the “Remez exchange algorithm”) [33]. This routine optimizes the transfer based on a number of filter requirements.

¹⁹The human ear is sensitive to delay variations down to the microsecond range.

²⁰The definition for Euler’s relation is: $e^{j\pi} + 1 = 0$. According to Feynman this is the most beautiful mathematical formula as it relates the most important mathematical constants to each other.

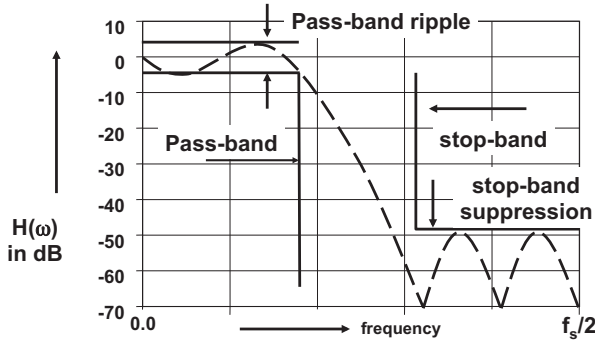


Fig. 2.38 Definition scheme of a filter response

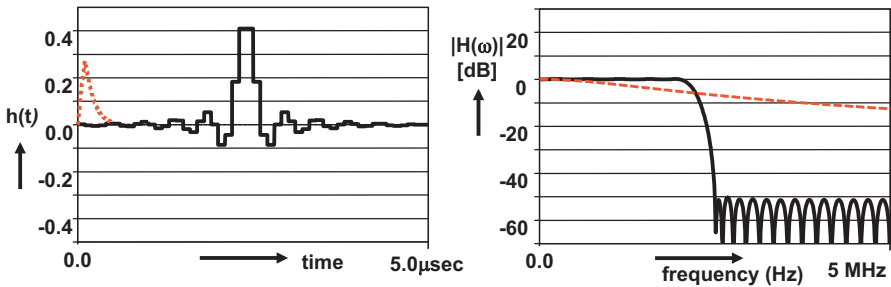


Fig. 2.39 A low-pass FIR filter with 48 coefficients, *left* is the impulse response of the filter and its analog realization (*dashed*). *Right* is the frequency response of both plotted on a linear frequency scale

Figure 2.38 shows a number of terms to define various specification points. Next to that the number of delay elements N , the number of pass and stop bands, and the form of the transition between the bands are required. Some variants of filter design programs allow to compensate alias filters or zero-order hold-effects.

Figure 2.39 shows the impulse response for a somewhat more elaborate filter with 48 coefficients. An impulse response is obtained by shifting a “1” value through the structure preceded and followed by zero samples. An impulse response reveals all filter coefficients. A 2-pole RLC filter transfer function with a quality factor of 0.5 is drawn in dotted lines as a comparison. The delay time is of course much shorter than the 24 cycles of the FIR filter. However the suppression of the digital filter is superior to a simple analog filter²¹ or a 7-tap filter as in Fig. 2.41.

Redesigning this filter with the same 10 Ms/s sample rate and 48 coefficients creates a bandpass filter, Fig. 2.40.

The digital time response highly resembles the ringing of a high-Q analog filter of the same specification. The accuracy in which required filter characteristics can be

²¹ An equivalent analog filter would require 10–12 poles.

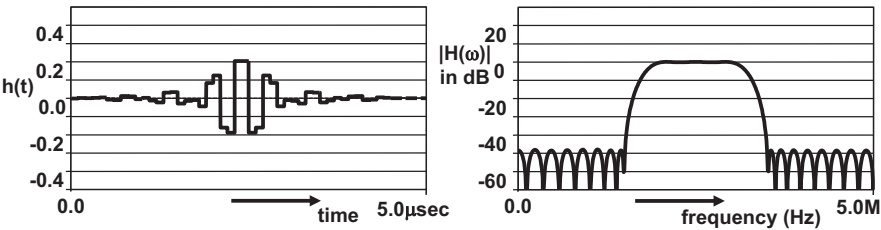


Fig. 2.40 A bandpass FIR filter with 48 coefficients, *left* is the impulse response of the filter and its analog realization. *Right* is the frequency response of both

Table 2.4 Coefficient values
for the low-pass FIR filter of
Fig. 2.41

Coefficient	Value
$a_0 = a_6$	-0.06
$a_1 = a_5$	0.076
$a_2 = a_4$	0.36
a_3	0.52

defined with FIR filters is clearly illustrated here. In practical realizations the price for an accurately defined filter is the large hardware cost of the delay elements, the coefficients and their multipliers, and the associated power consumption.

The FIR filter has been described in this section as a mathematical construction and no relation was made with the physical reality. Some examples of fully analog FIR realizations are found in switched capacitor circuits and charge-coupled devices.²² Most FIR filters are implemented in the digital domain: from IC building blocks to FPGA and software modules. In digital-to-analog conversion the semi-digital filter uses digital delays with analog coefficients, see Sect. 7.3.7.

Example 2.19. Determine with a suitable software tool the coefficients for the structure in Fig. 2.37 to create a low-pass filter.

Solution. If the transition for the low-pass filter is chosen at approximately $f_s/4$ coefficients as in Table 2.4 is found.

Figure 2.41 shows the time response and the frequency transfer function from Fig. 2.37 with the coefficients of Table 2.4. In this example of a time-discrete filter the frequency transfer is symmetrical with respect to half of the sampling rate, which was chosen at 10 Ms/s. The spectrum repeats of course at multiples of the sampling rate.

²²In the period 1970–1980 the charge-coupled device was seen as a promising candidate for storage, image sensing, and signal processing. Analog charge packets are in this multi-gate structure shifted, split and joint along the surface of the semiconductor. Elegant, but not robust enough to survive the digital era.

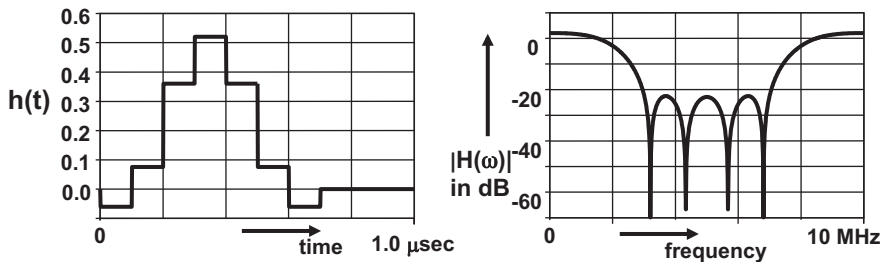


Fig. 2.41 The impulse response and the frequency transfer function of a seven coefficient filter from Fig. 2.37 at a 10 Ms/s sample rate

2.7.2 Half-Band Filters

In order to reduce the complexity of digital FIR filters additional constraints are needed. Introducing the symmetry requirement:

$$H(\omega) + H(\omega_s/2 - \omega) = 1 \quad (2.50)$$

leads to such a complexity reduction. At a frequency $\omega = \omega_s/4$ this constraint results in $H(\omega_s/4) = 0.5$, while the simplest fulfillment of the symmetry requirement around $\omega_s/4$ forces a pass-band, on one side, and a stop band, on the other side, of this quarter sample rate. Consequently these filters are known as “half-band” filters. Substitution of the transfer function for symmetrical filters with an odd number of N coefficients $k = 0, 1, \dots, m, \dots, N-1$ and with the index of the middle coefficient equal to $m = (N-1)/2$, leads to:

$$\begin{aligned} a_m &= 0.5 \\ a_{m+i} &= a_{m-i} = C_i, \quad i = 1, 3, 5, \dots \\ a_{m+i} &= a_{m-i} = 0, \quad i = 2, 4, 6, \dots \end{aligned}$$

Half of the filter coefficients are zero and need no hardware to implement. Optimizing the filter transfer for a minimum deviation of an ideal filter results in a $\sin(x)/x$ approximation:

$$a_{m+i} = \frac{\sin(i\pi/2)}{i\pi}, \quad i = -m, \dots, -2, -1, 0, 1, 2, \dots, m \quad (2.51)$$

Table 2.5 lists the coefficients for four half band filters designed for a pass-band from 0 to $f_s/8$ and a stop band from $3f_s/8$ to $f_s/2$. Figure 2.42 compares these four half-band filter realizations. The filter with the least suppression has three non-zero coefficients increasing to nine for 72 dB suppression.

Table 2.5 The non-zero coefficients for four half band filters in Fig. 2.42 (courtesy: E.E. Janssen)

Coefficients	Suppression (dB)	Ripple (dB)
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.2900$	20	0.8
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.2948$ $a_{m-3} = a_{m+3} = -0.0506$	38	0.1
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.3016$ $a_{m-3} = a_{m+3} = -0.0639$ $a_{m-5} = a_{m+5} = 0.0130$	55	0.014
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.3054$ $a_{m-3} = a_{m+3} = -0.0723$ $a_{m-5} = a_{m+5} = 0.0206$ $a_{m-7} = a_{m+7} = -0.0037$	72	0.002

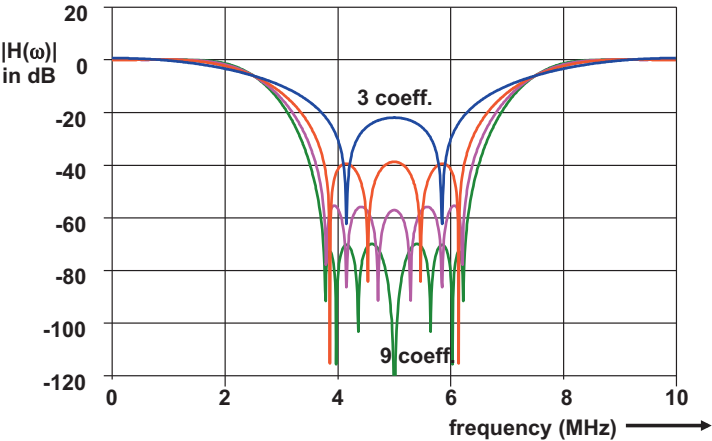


Fig. 2.42 Four half-band filters, with 3, 5, 7, and 9 non-zero coefficients (courtesy: E.E. Janssen)

In order to obtain a small-area implementation the coefficients are rounded integers. With integer filter coefficients no full multiplier circuit is needed but dedicated shift and add circuits create the weighting of the signal samples.

2.7.3 IIR Filters

A drastic solution to the hardware problem of FIR filters is the “infinite impulse response” IIR filter.

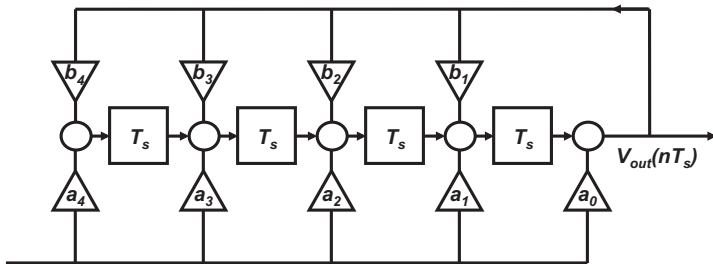


Fig. 2.43 The structure of an infinite impulse response (IIR) filter containing a feedback path from output to the summation nodes

Figure 2.43 shows the general or canonical form of a digital IIR filter. Coefficients a_0 to a_4 perform the same function as in an FIR filter. In addition the coefficients b_1 to b_4 feed the output back into the delay chain. This feedback modifies the FIR transfer. If all coefficients b_k equal zero, again an FIR filter will result.

A similarity to the RLC filter is that in both filter types the signal is circulating in the filter. In the RLC filter the signal swings between the electrical energy in the capacitor and the magnetic energy in the coil. In an IIR filter the signal circulates via the feedback path. The signal frequency in relation to the delay of the loop and the coefficients will determine whether the signal is amplified or attenuated and for how long. The transfer function of an IIR filter is (for the mapping from z -domain to frequency domain the approximation $z = e^{j\omega T_s}$ is applied):

$$H(z) = \frac{\sum_{k=0}^{N-1} a_k z^{-k}}{1 - \sum_{k=1}^{N-1} b_k z^{-k}} \Leftrightarrow H(\omega) = \frac{\sum_{k=0}^{N-1} a_k e^{-jk\omega T_s}}{1 - \sum_{k=1}^{N-1} b_k e^{-jk\omega T_s}} \quad (2.52)$$

The numerator specifies the FIR filter part, while the denominator describes the feedback path. Both are formulated as a polynomial in z^{-1} . For absolute stability (a bounded input results in a bounded output signal) the zeros of the denominator polynomial must be smaller than 1, they reside inside the unity circle of Fig. 2.34. In theory the signal will never fully extinguish in an IIR filter. In practice, a signal that experiences a feedback factor $1 - \Delta$ will pass in the order of $1/\Delta$ times through the filter. A filter with $\Delta \ll 1$ is called a resonator and resembles a high-Q RLC filter. An IIR construction where the denominator has a zero term equal to “1” will oscillate.

A sharp low-pass filter with just four delay elements as in Fig. 2.44 realizes between 2 and 4 MHz a suppression of 40 dB, which is comparable with a seventh order analog filter.

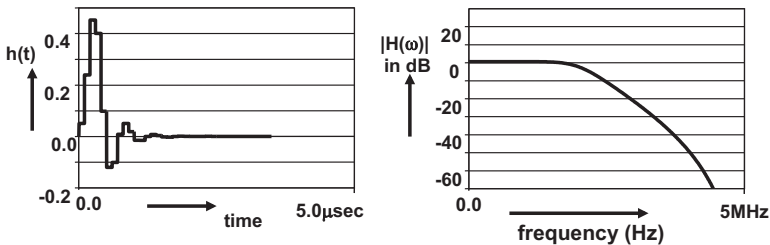


Fig. 2.44 The impulse response of this 4-tap IIR continues beyond four sample periods. The frequency response of this 4-tap filter is much steeper than the response of the 7-tap FIR filter

Table 2.6 Comparison of discrete filter realization techniques

Implementation	Switched capacitor	Semi-digital filter Sect. 7.3.7	Digital hardware
Delay	Analog	Digital	Digital
Coefficients	Analog	Analog	Digital
Most used	As IIR/resonator	As FIR	Both FIR and IIR
Noise	Accumulates in signal range	Only from coefficients	Related to word width
Tolerance	Capacitor matching	Current source matching	Unlimited
Alias-filter	Required	Depends on system requirements	Required
Power	Moderate	Output related	High
Performance	Limited by noise	Limited by noise	Limited by word width

Time-discrete filters can be realized in various implementation technologies. Table 2.6 compares three realization forms of time-discrete filters. The switched capacitor filters are mostly used in medium specification circuits. The realization is practically limited to 40–50 dB signal-to-noise levels.

Exercises

- 2.1. A sinusoidal signal of 33 MHz is distorted and generates second and third harmonics. It is sampled by a 32 Ms/s system. Draw the resulting spectrum.
- 2.2. A signal bandwidth from DC to 5 MHz must be sampled in the presence of an interferer frequency at 20 MHz. Choose an appropriate sampling rate.
- 2.3. An image sensor delivers a sampled-and-held signal at a fixed rate of 12 Ms/s. The succeeding digital signal processor can run at 10 MHz. Give an optimal theoretical solution. What is a (non-optimal) practical solution?
- 2.4. What is a stroboscope? Explain its relation to sampling.

- 2.5.** Must the choice for a chopping frequency obey the Nyquist criterion?
- 2.6.** Set up a circuit where the signal is stored as a current in a coil. What is the meaning of $i_{noise} = \sqrt{kT/L}$?
- 2.7.** In Example 11.6 the available equipment can only measure signals up to 5 MHz. What can be done in order to measure the harmonic distortion of the analog-to-digital converter at roughly 5 GHz?
- 2.8.** The signal energy of the luminance (black-and-white) signal of a television system is concentrated around multiples of the line frequency (15,625 Hz). Although the total television luminance signal is 3 MHz wide, a sample rate of around 4 Ms/s is possible. Give an exact sample rate. Why will this sampling not work for a complete television signal with color components added?
- 2.9.** An audio system produces samples at a rate of 44.1 ks/s. With a maximum audio signal of -6 dB of the full-scale between 10 and 20 kHz, propose an alias filter that will reduce the frequency components over 20 kHz to a level below -90 dB.
- 2.10.** How much SNR can be obtained if a signal of 10 MHz is sampled with a sample rate of 80 Ms/s with 5 ps_{rms} jitter. What happens with the SNR if the sample speed is increased to 310 Ms/s at the same jitter specification. Compare also the SNR in a bandwidth between 9 and 11 MHz.
- 2.11.** Design a half band filter with 19 non-zero coefficients to get a pass-band stop band ratio of 100 dB. Use a computer program.
- 2.12.** An analog-to-digital converter is sampling at a frequency of just $2.5\times$ the bandwidth. Due to the large spread in passive components, the problem of alias filtering is addressed by placing before the converter a time-discrete filter running at twice the sample rate and before that time-discrete filter a time-continuous filter. Is this a viable approach? There are twice as many samples coming out-of the filter then the converter can handle. Is there a problem?
- 2.13.** Make a proposal for the implementation of the filters in the previous exercise if the bandwidth is 400 kHz and a attenuation of better than 60 dB must be reached starting from 500 kHz.
- 2.14.** In Example 2.14 the second capacitor is twice the size of the first: $C_2 = 2C_1$. Will the signal-to-noise ratio remain the same?
- 2.15.** A sinusoidal signal of 33 MHz is distorted and generates second and third harmonics. It is sampled by a 32 Ms/s or a 132 Ms/s system. Draw the resulting spectra. What sample rate do you favor?
- 2.16.** A signal source delivers a signal that consists of three components: at 3, 4, and 5 MHz. The signal is processed by a sampling system with an unknown sample rate. The output contains in the 0–0.5 MHz band only frequencies at 0.27, 0.36, 0.45, and 0.46 MHz. What sampling frequency was used? Complete the spectrum till 1 MHz.

2.17. An RF oscillator at 2.45 GHz contains harmonic distortion products at 2x and 3x time the oscillation frequency. The available spectrum analyzer can measure up to 10 MHz, but has a 10 GHz input bandwidth sampling circuit with variable sampling rate up to 10 GS/s. Advice how to measure the harmonic distortion.

2.18. A 2 MHz signal is sampled by a 100 Ms/s clock with 10ps_{rms} jitter. In the digital domain the band of interest is limited to DC-5 MHz, Calculate the SNR. The digital circuits repeat a process every 10 ms and this is the cause of the 10 ps jitter. What is the resulting spectrum in DC –5 MHz?

<http://www.springer.com/978-3-319-44970-8>

Analog-to-Digital Conversion

Pelgrom, M.

2017, XXVII, 548 p. 506 illus., 287 illus. in color.,

Hardcover

ISBN: 978-3-319-44970-8