Sampling: Time Domain Discretization

1 A/D Conversion as Two-Dimensional Process

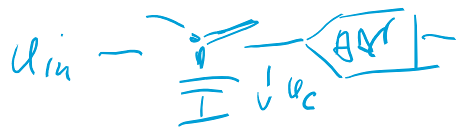
Time-discrete signals exist at particular time points $t_n=t_0+n\cdot T$ only, where $T=1/f_S$ with f_S being the sampling frequency: $x_n = x(n) = x(t_n)$.

Value-discrete signals exist with particular values only, typically $x_{offset}+m\cdot\Delta$ with Δ being the lest significant bit (LSB). For communication purposes there exist also signals with $\Delta=f(amplitude)$, which will not be considered in this chapter.

The **transition from Laplace transformation to z-transformation** occurs with sampling using a constant sampling interval T and is independent of quantization.

Typical we find all 4 forms of signal representations, as illustrated in Fig. 4.1

Situation in time	value	example
a) time-continuous	+ value-continuous:	analog circuitry, e.g. RC lowpass
b) time-discrete	+ value-continuous:	switched capacitors
c) time-continuous	+ value-discrete:	DAC output
d) time-discrete	+ value-discrete:	digital signal processing



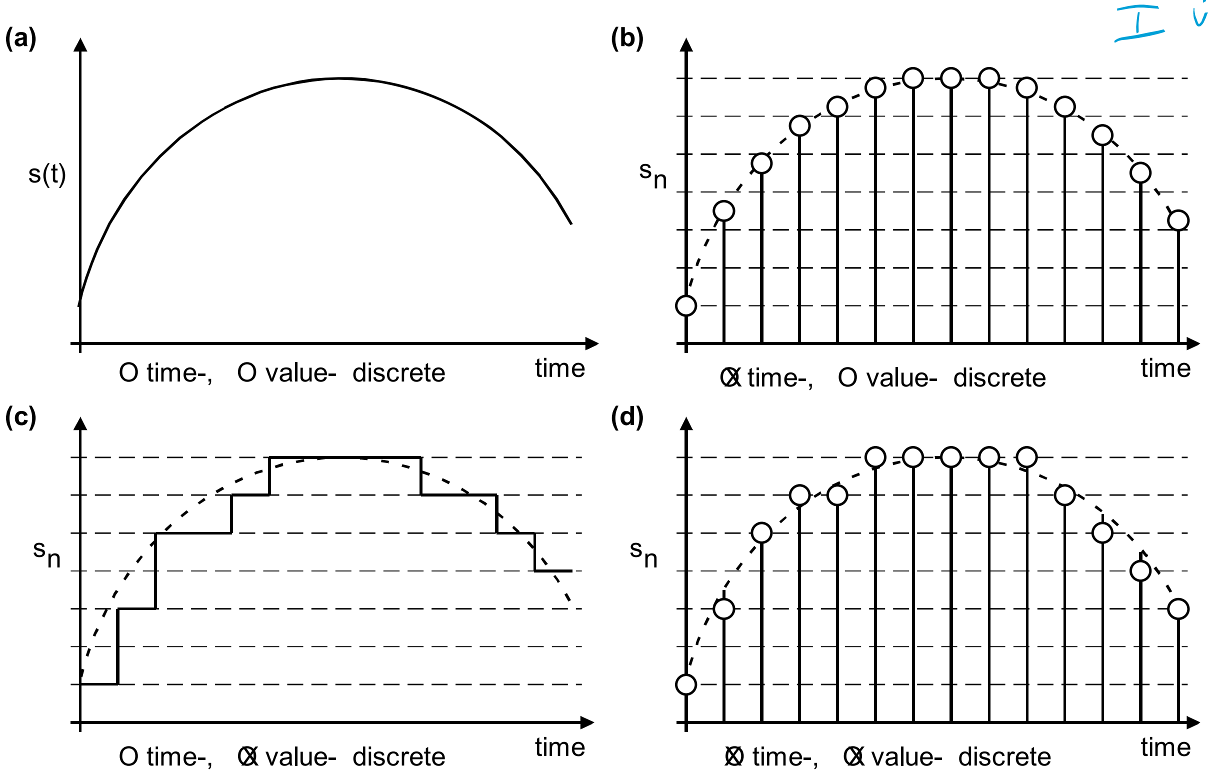


Fig. 1: Signals being: **(a)** analog, **(b)** time-, **(c)** value-, **(d)** time- and value-discrete

2 Sampling Analog Waveforms

2.1 Time-Domain Considerations

2.1.1 The Shannon-Nyquist Criterion

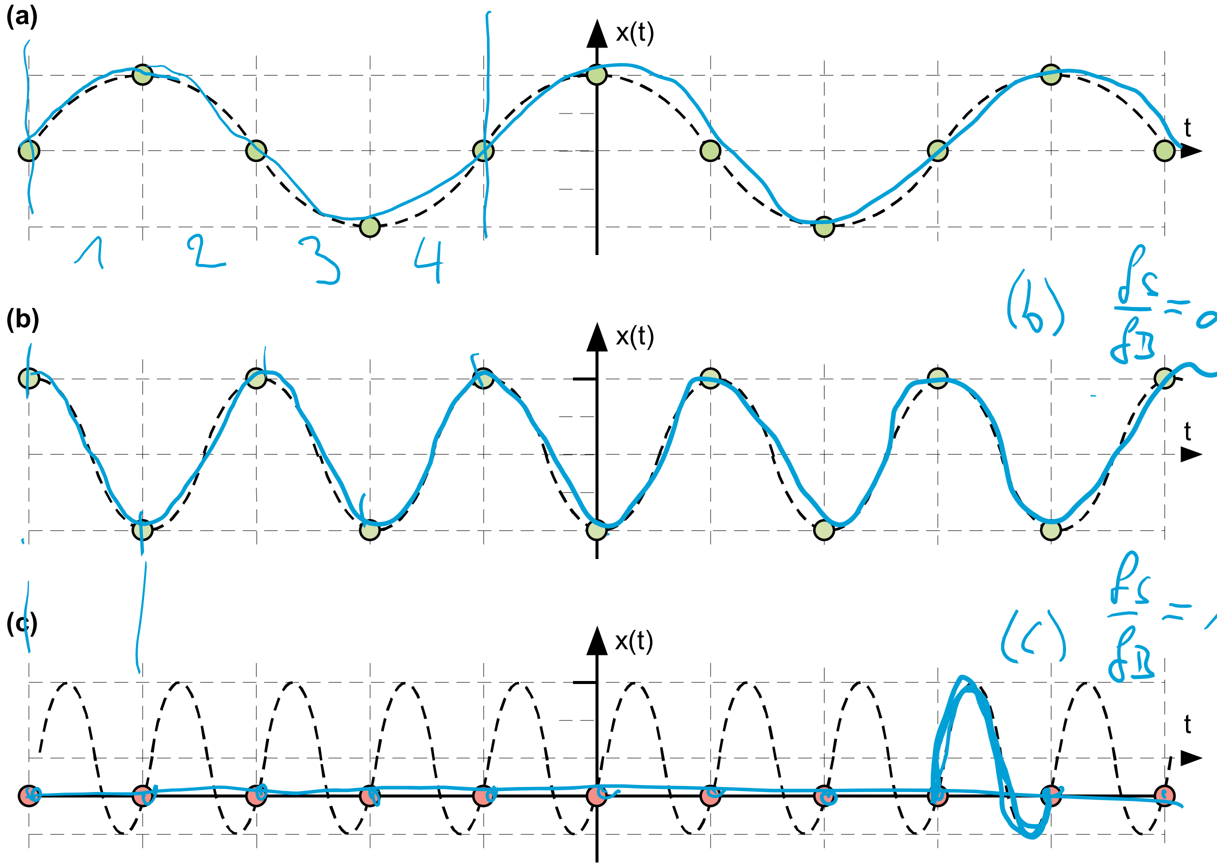


Fig. 2.1.1: Sampling sinusoidal waveforms of with frequency $f_{\#}$, #=a, b, c, at different rates f_s, then trying to interpolate the samples. **(a)** $F_a= f_a/f_s=\frac{1}{4}$, **(b)** $F_b= f_b/f_s/ =\frac{1}{2}$, **(c)** $F_c=f_c/f_s=1$.

Fig. 2.1.1 illustrates sinusoidal waveforms with different frequencies $f_{\#}$, sampled with rate f_S corresponding to sampling interval and relative frequency, respectively:

$T_s = 1 / f_s$ *and* $F = f / f_s$.

Some public sampling rates:

Old telephone:	8 KHz (±3dB band 300...3400Hz)
Audio data on CD:	44.1 KHz = $(1\cdot2\cdot3\cdot5\cdot7)^2$ Hz
Digital Video Broadcasting - Terrestrial (DVB-T):	13.3 MHz

Exercise:

Up to which relative frequency $F=f/f_S$ can curves in Fig. 2.1.1 be reconstructed from samples, when samples are interpolated with lowest possible frequency and amplitude?

According to the Nyquist-Shannon criterion, a sinusoidal wave can be reconstructed from samples, when the sampling frequency is at least twice the sinusoidal waveform's frequency. Consequently, the bandwidth that can be can be transmitted in sampled form, is

$$f_B \leq \tfrac{1}{2} f_S \quad \Leftrightarrow \quad f_S \geq 2 f_B \,. \qquad (1)$$

During digital signal processing (DSP) we are rather interested in frequencies relative to the sampling rate f_s than in physical frequencies. Consequently, we do not compute the true frequencies f and ω but relative frequencies

$$F = f / f_s = f \cdot T_s \quad \textit{and} \quad \Omega = \omega / f_s = \omega\, T_s = 2\pi F. \qquad (2)$$

According to Shannon and Nyquist correct signal reconstruction requires

$$F \leq \tfrac{1}{2} \quad <=> \quad \Omega \leq \pi \,. \qquad (3)$$

Exercise:
The figure below shows three sinusoidal waveforms. Vertical dashed lines are sampling time points. Draw sampled pulses and then reconstruct curves by interpolating with lowest-amplitude, lowest-frequency sinusoidal waves. Also compute relative frequencies F and Ω.

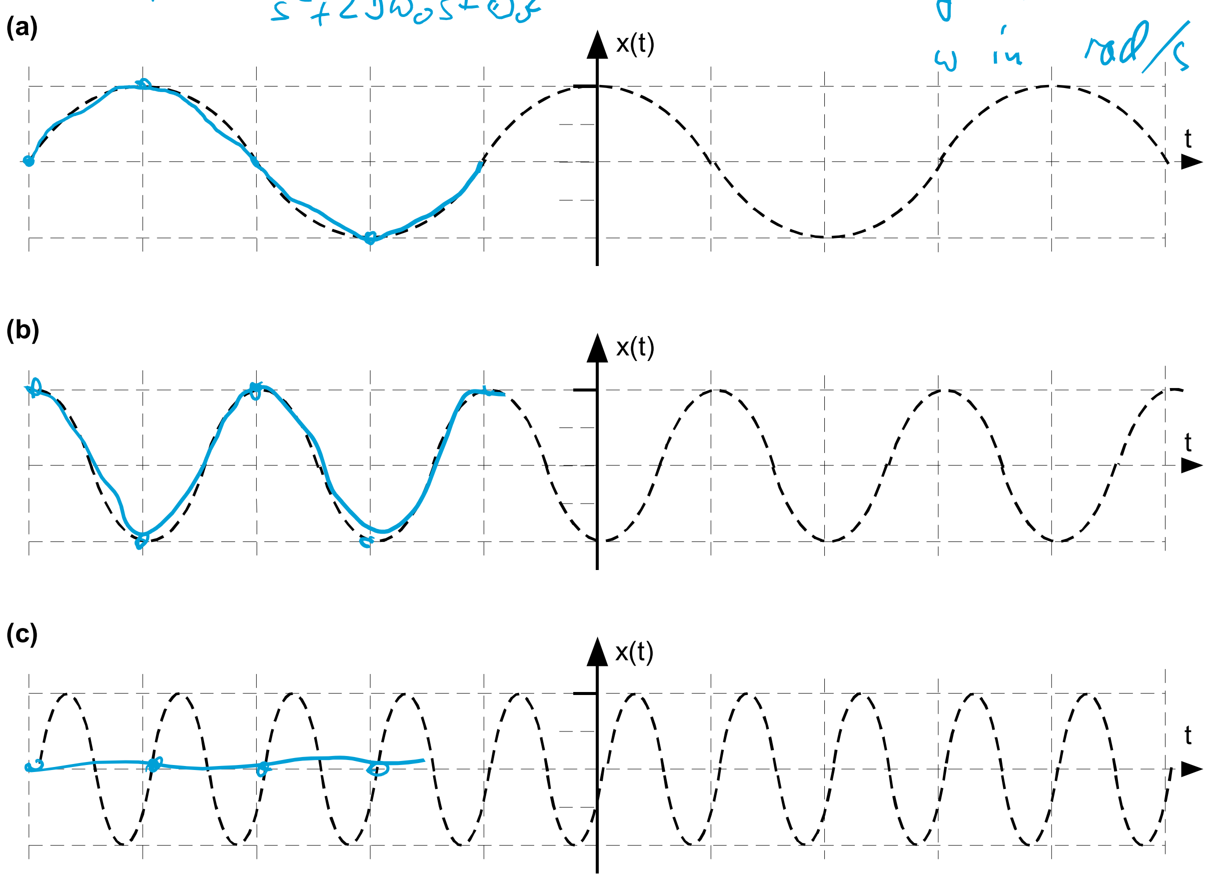


Fig.: Exercise to experience sampling of different sinusoidal waveforms. (Sol. → Fig. 2.1.1)

2.1.2 Problems with the Rule of Shannon and Nyquist

Even when the Nyquist criterion is fulfilled, some problems may occur. In Fig. 2.1.1(b) the sampler picks samples exactly at the extrema of the curves. This cannot be guaranteed as illustrated in Fig. 2.1.2:

- **Fig. part (a)** shows a waveform with frequency $f_{B(a)}=½f_S$, however, the sampler does not get the curve's extrema. Reconstruction delivers a signal with lower amplitude and phase shift.
- **Fig. part (b)** shows a waveform with frequency $f_{B(b)}=½f_S$, however, the sampler gets the curve's zeros. Reconstruction delivers a zero signal.
- **Fig. part (c)** illustrates sampling of a sinusoidal wave with $f_{B(c)}<½f_S$, the *Nyquist* criterion is fulfilled. Reconstruction delivers a signal with beats. This can be considered composed of the original frequency $½f_S$-Δf and a second signal with frequency $½f_S+\Delta f$. An extremely good filter could remove the upper signal, but this is costly.

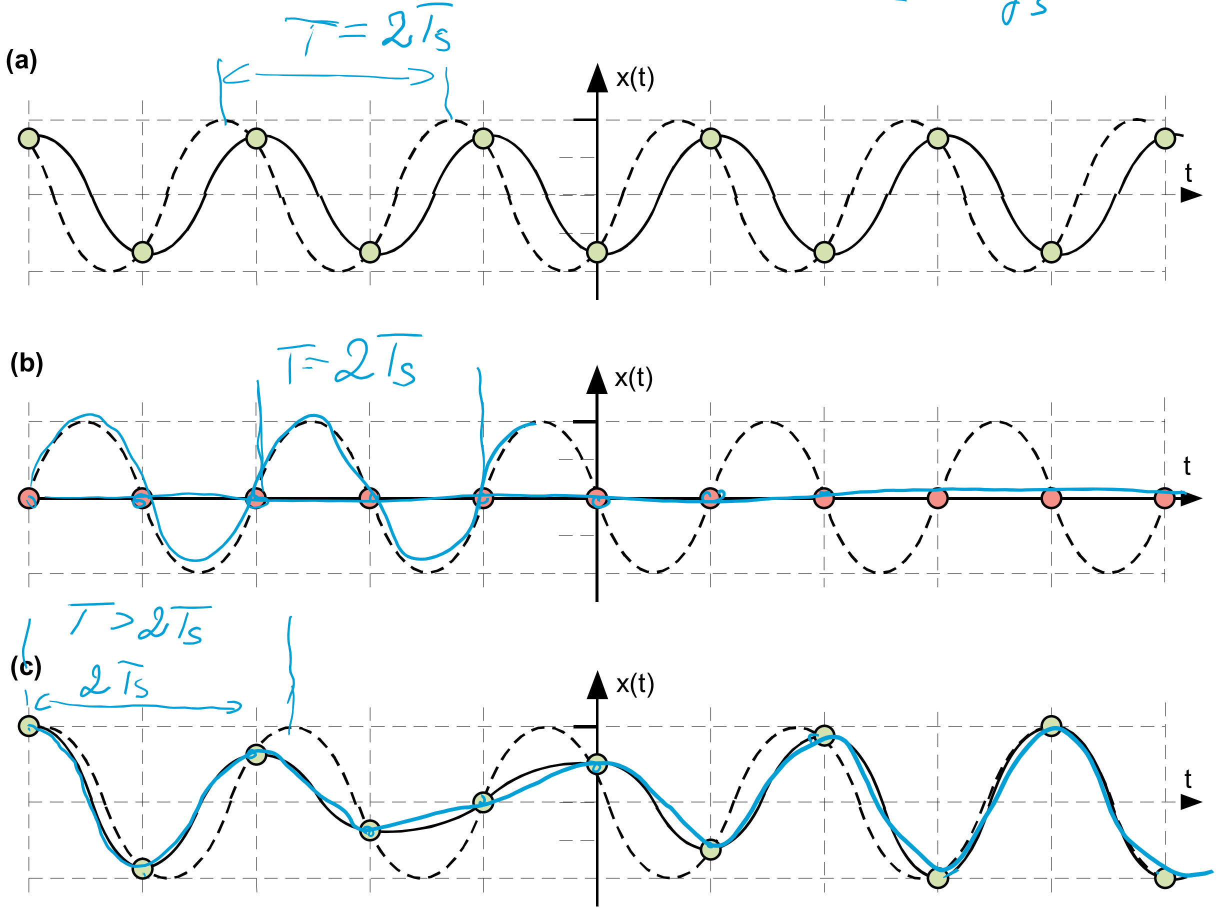


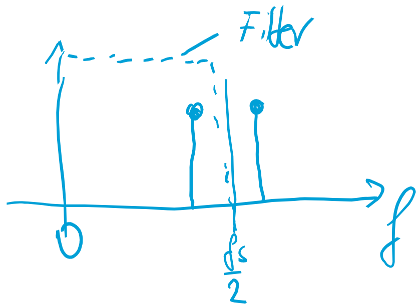
Fig. 2.1.2: Problems close to Nyquist bandwidth $f_B=f_S/2$.

In practical applications, these problems can be overcome with increased sampling rate, e.g.

f_S = **4...10** f_B

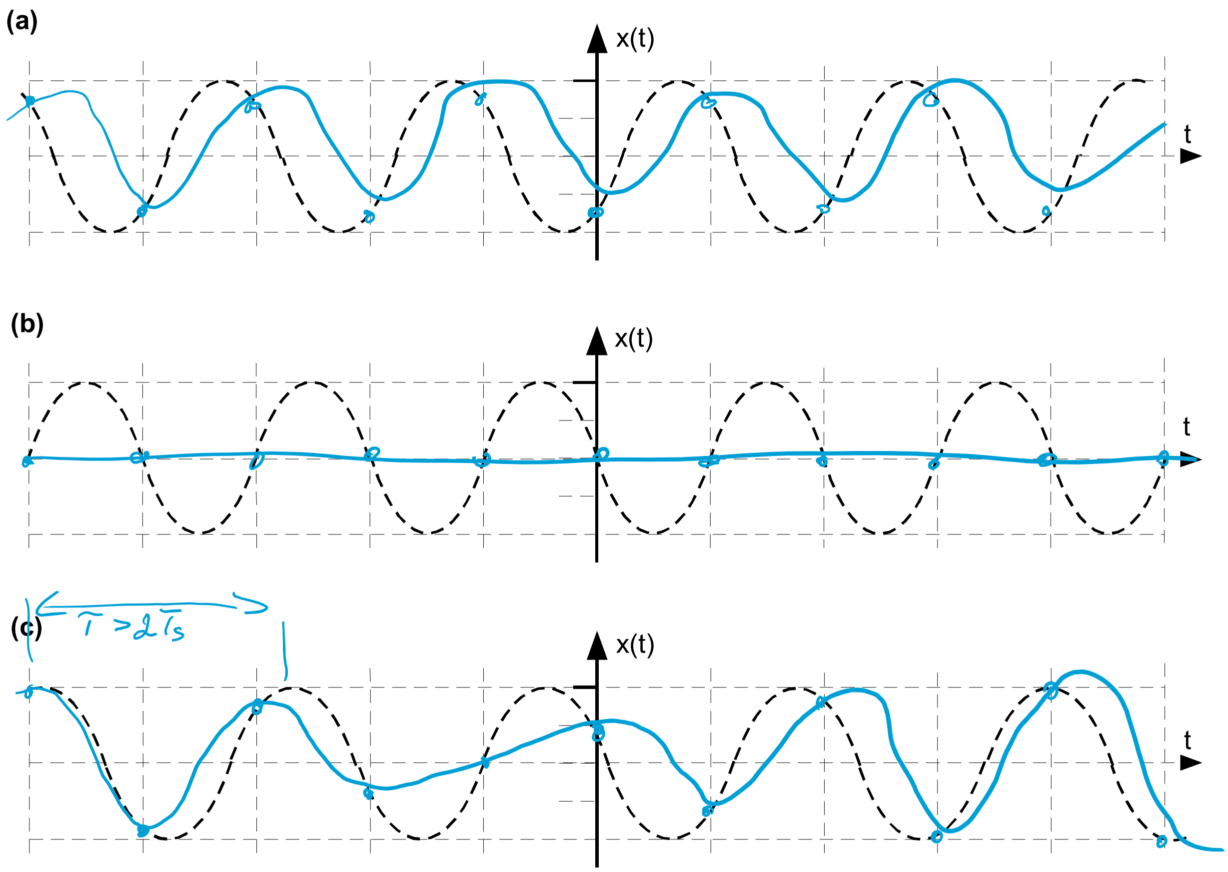
corresponding to an „**O**ver **S**ampling **R**atio“

$OSR = f_S/2f_B$ = **2...5** (4)

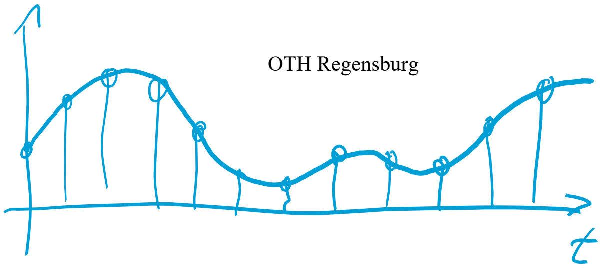


As will be shown later in more detail, sampling is no lowpass filtering! Frequencies >½f_S are not removed but occur at frequencies in the baseband, i.e. in range 0 ... ½f_S.

Exercise 2.1.2: the figure below shows three sinusoidal waveforms and vertical dashed lines as sampling time points. Draw the sampled pulses and then reconstruct curves from them interpolating with lowest-amplitude, lowest-frequency waves. Also compute the relative frequencies F and Ω.



Exercise Fig. 2.1.3: Sampling of frequencies close to $f_S/2$. (Solution → Fig. 2.1.2)



Frequency-Domain Considerations

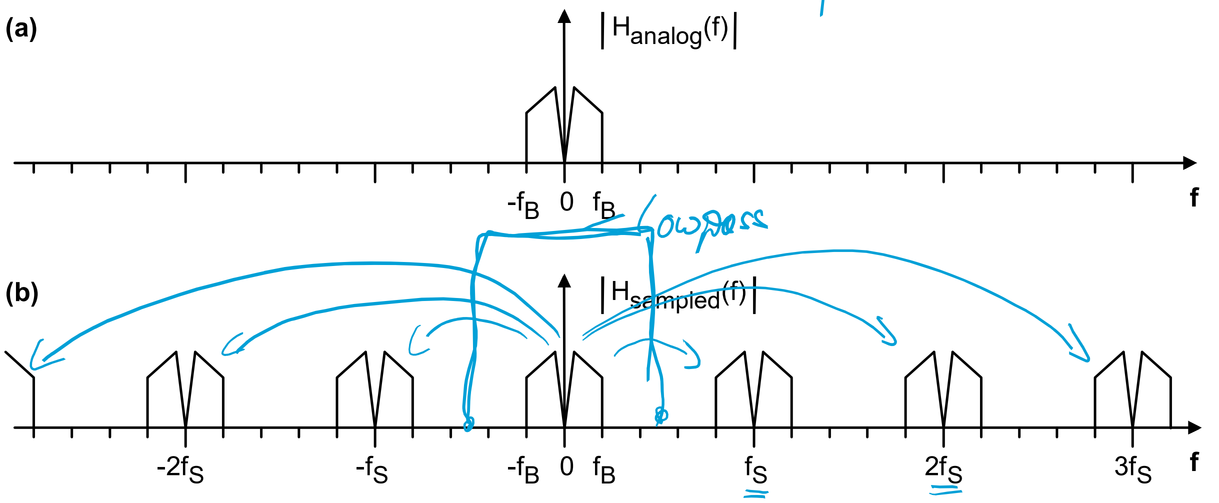


Fig. 2.2-1: Frequency spectrum (**a**) before and **(b)** after sampling with rate f_S.

Fig. 2.2(a) shows the spectrum of an analog signal. Sampling in time-domain with rate f_S translates it to the spectrum illustrated in Fig. part (b). Further spectra are added around integral multiples of the sampling frequency f_S. The baseband around f=0 is essentially the same and can be recovered by lowpass filtering.

Fig. 2.2-2 illustrates an effect called aliasing. The pulse is very short, but the reconstruction has to occur in the band 0...½f_S. This effect is called aliasing, which is completely different from lowpass filtering.

Fig. 2.2-2:
The sampler picks a very short pulse, which is later reconstructed at frequencies in the range 0...½f_S.

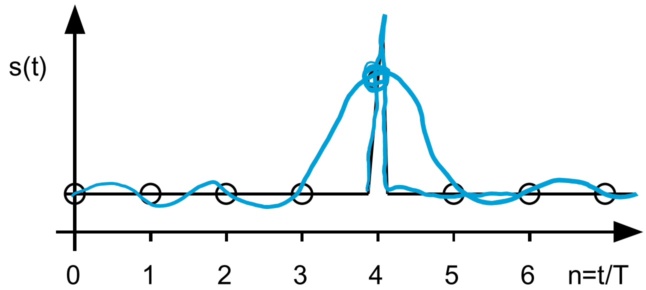


Fig. 2.2-3 illustrates different sampled signal spectra to gain an understanding of aliasing.

- In Fig. part (a) the signal's bandwidth f_{Ba} < ½f_S and the spectra can be separated.
- In Fig. part (b) the signal's bandwidth is at the limit of f_{Bb} = ½f_S and spectra can theoretically be separated with an infinitely good lowpass.
- In Fig. part (c) the signal's bandwidth is beyond the limit: f_{Bc} > ½f_S. Spectra overlap and the original baseband 0...f_{Bc} cannot be recovered any more. This is what we call aliasing.

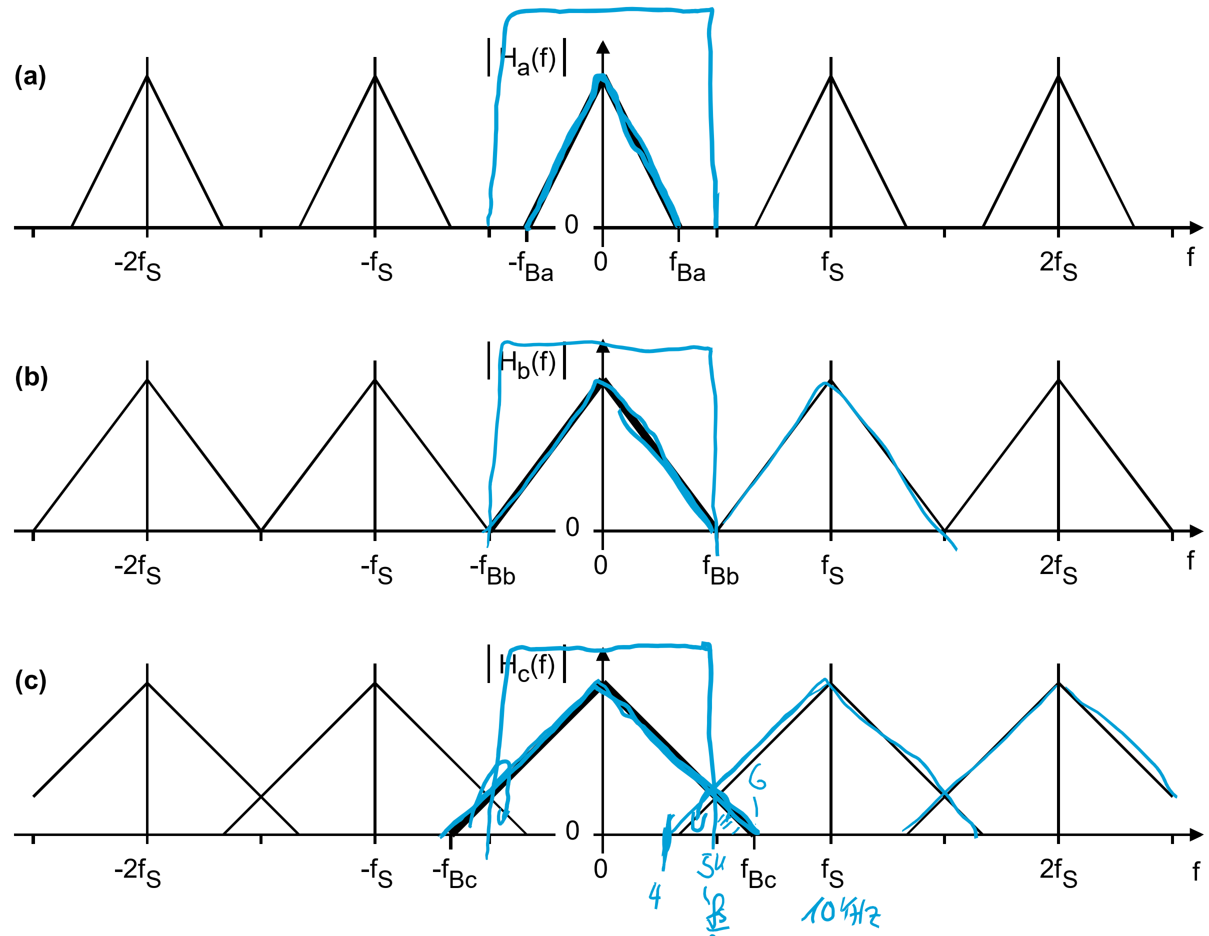


Fig. 2.2-3: (a) Spectrum of oversampled signal, **(b)** sampling at Nyquist rate, **(c)** sampling below Nyquist rate and consequently aliasing: Periodic spectra overlap; they cannot be distinguished by lowpass filtering any more.

This example illustrates that signal spectra of a sampled function $x(nT)$ of $x(t)$ can be recovered if

- $x(t)$ has a band limited Fourier transform $X(f)$ with $X(f)=0$ when $f>f_B$ and
- sampling rate $1/T = f_s > 2f_B$

The energy of all higher signal frequencies is sampled to the alias frequency $f_{alias} \le \frac{1}{2} f_S$ according to

$$f_{alias} = f_{in} - N \cdot f_s \quad \text{with} \quad N = \text{round}(f_{in} / f_S).$$

Do not worry about negative frequencies; they might indicate a phase shift, which does not matter when we have a frequency shift. (Remark: All real sinusoidal waves have their half energy at a positive and the other halt a negative frequency: $\cos(\omega t)=\frac{1}{2}(e^{j\omega t}+e^{-j\omega t})$ and $\sin(\omega t)=(e^{j\omega t}-e^{-j\omega t})/(2j)$. If the positive frequency part becomes negative, then the corresponding negative part becomes positive.)

Exercise 2.2: You record music at a rate of f_S=10KHz. The sampled accord contains 3 KHz, 6 KHz, 12 KHz and 24 KHz. Which frequencies do you hear when you hear your sampled data?

Solutions: 3 → 3, 6 → |6-10|=4, 12 → |12-10|=2, 24 → |24-2x10|=4

3 Signal-Value Hold Circuits

3.1 Why do we need Hold Circuits?

(a) Architecture of a SAR ADC using a DAC in the feedback branch.

(b) Conversion process: approximate *$U_{DAC,out}$* to *$U_{ADC,in}$* by trial and error from most to least significant bit

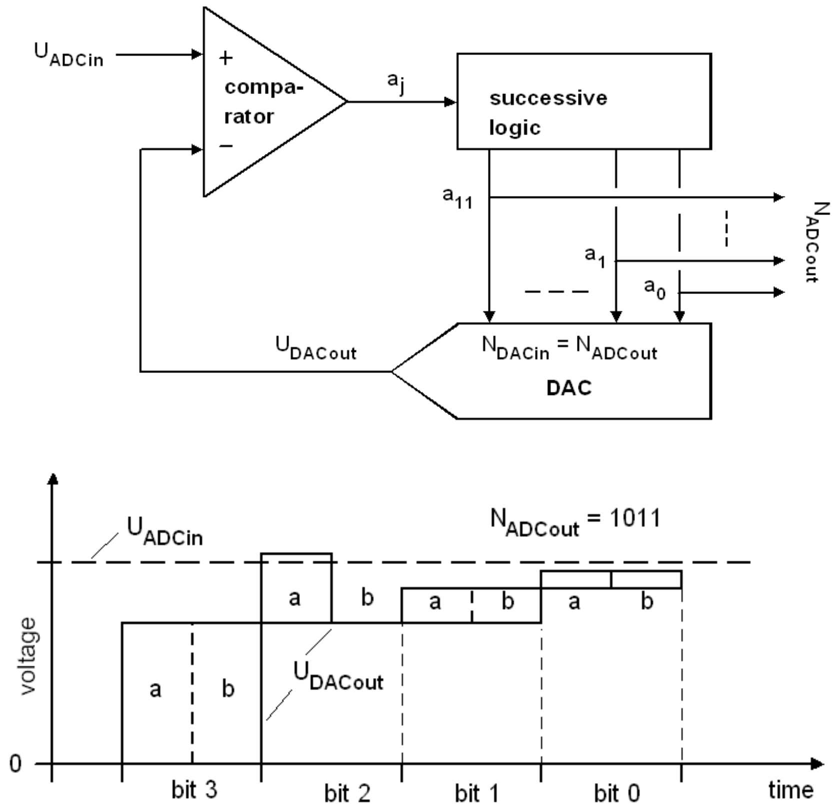


Fig. 3.1: Principle of successive approximation register (SAR) A/D conversion. Phase *a* and *b* in the figure's part (b) can be combined in a single phase.

Fig. 3.1 illustrates the principle of the most common A/D converter (ADC) architecture: The successive approximation register (SAR) type. The successive logic starts from all bits zero to setting the most significant bit in phase *a*, the D/A converter (DAC) converts it to analog and the comparator compares it with the analog input signal. If the DAC output is below input signal U_{DACin}, this bit is confirmed in phase *b*, otherwise it is set back in phase *b*. After a decision about the most significant bit has been taken, this bit is stored in a shift register and the procedure starts with the next significant bit.

It is obvious from Fig. part (b), that input signal U_{DACin} must be held constant during evaluation of one bit after the other. This holding of the input voltage is the task of a sample and hold circuit, which is a mathematical idealization of the physically realizable track and hold circuit.

3.2 Sample & Hold Mathematics, Track & Hold Circuits

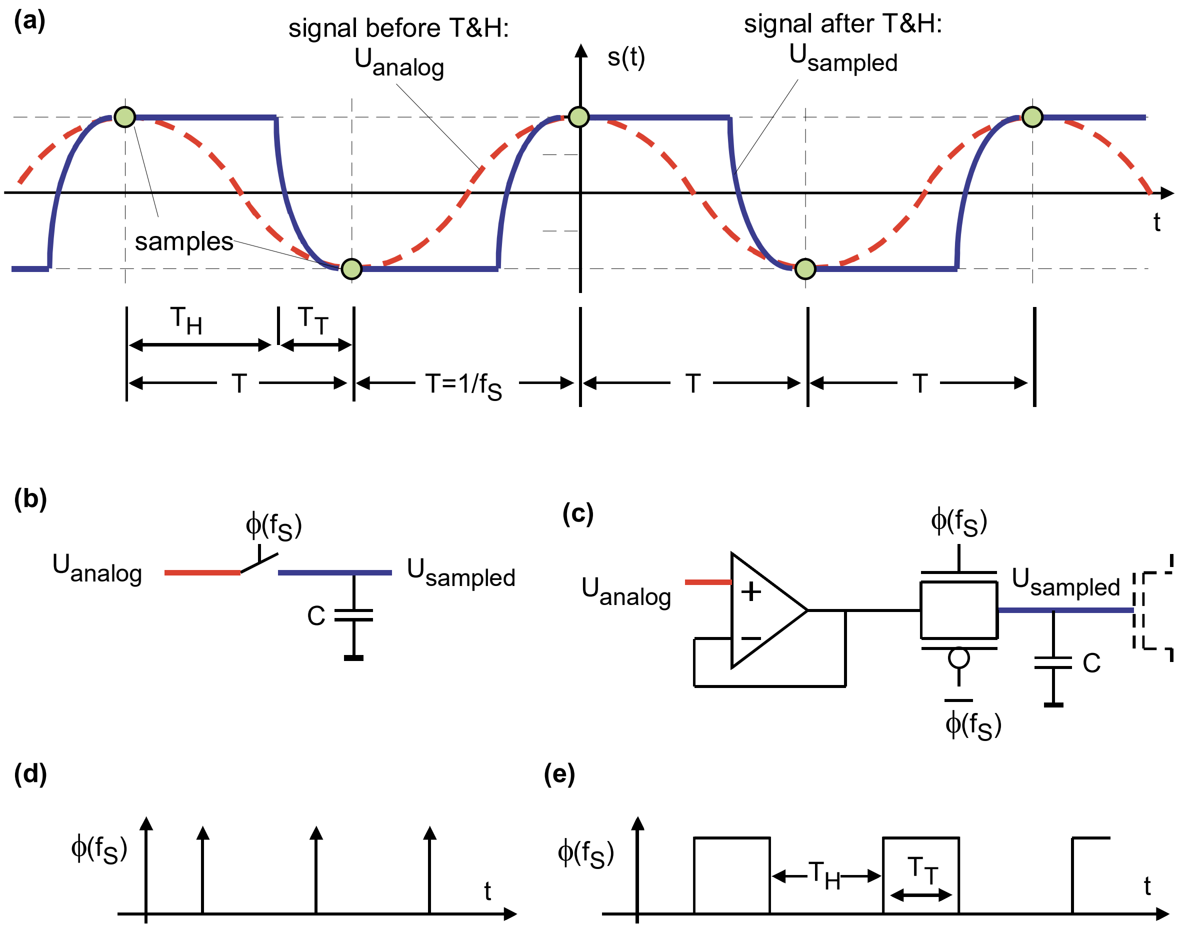


Fig. 3.2: (a) Sampling of a sinusoidal waveform (dashed) at Nyquist rate. The track & hold circuit hold the sample value constant during the conversion time of the ADC. **(b)** T&H circuit idealized **(c)** T&H circuit realized, **(d)** S&H mathematics uses Dirac pulses, **(e)** Track & hold phases T_{hold} (T_H) and T_{track} (T_T).

The **mathematical concept of sampling** a time-continuous function *s(t)* is for single samples $s(a) = \int_{t=-\infty}^{\infty} s(t)\delta(t-a)dt$, and for a sampled waveform $s[n] = \int_{t=-\infty}^{\infty} s(t)\sum_{n=-\infty}^{\infty} \delta(t-nT)dt$

with *δ(t)* being the Dirac function. This is useful for mathematics but difficult to realize.

For **practical sampling applications,** we find track & hold (T&H) circuits. The maximum sampling rate obtainable with track-time T_T and hold-time T_H is

$$f_{S,\max} = \frac{1}{T_{\min}} = \frac{1}{T_T + T_H} .$$

4 Anti-Aliasing Filters

4.1 Nyquist and Over-Sampler Architectures

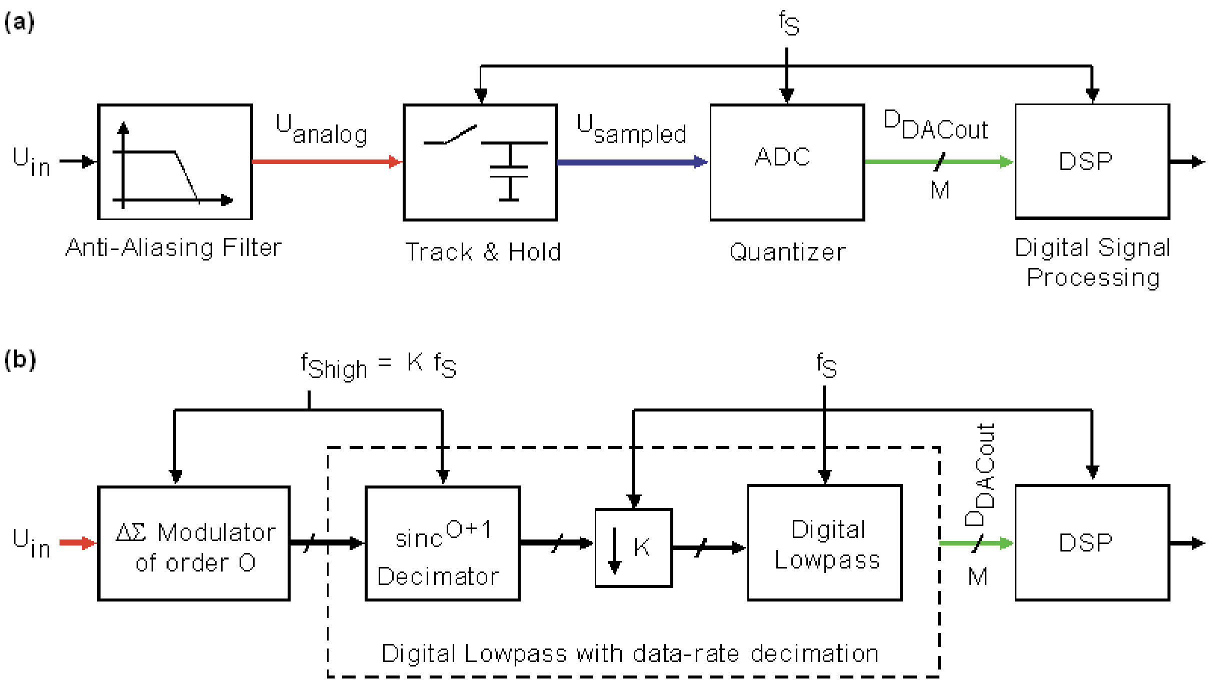


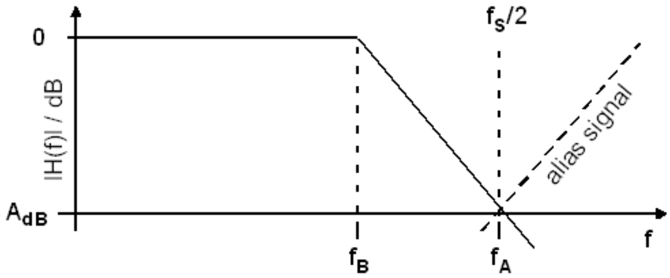
Fig 4.1: (a) Standard system to translate an analog voltage to a digital data stream.
(b) ΔΣ A/D conversion: lowpass filtering shifted to the digital side.

Fig. 4.1(a) illustrates the standard A/D conversion system. The anti-aliasing (AA) filter must be before the sampler, which is followed by the ADC.

Fig. 4.1(b) illustrates the difference for a system based on ΔΣ modulation (DSM). The anti-aliasing filter before the sampler is either strongly relaxed or removed, i.e. lowpass filtering is shifted from the analog to the digital side. This a driving motivation to use DSM.

4.2 Analog Anti-Aliasing Filters for Nyquist Samplers

Fig. 4.2:
Zur Dimensionierung eines Anti-Aliasing-Filters



To avoid aliasing, frequencies >½f_S must be suppressed by anti-aliasing filters. Fig. 4.2 illustrates the asymptotes of such a filter. Formula using lg=log10: To obtain an attenuation A_{dB} within frequency range f_B...f_A the filter needs an order of at least

$$N = \frac{A_{dB}}{20dB \cdot \lg\left(\frac{f_A}{f_B}\right)} \xrightarrow{f_A = f_S/2} \frac{A_{dB}}{20dB \cdot \lg\left(\frac{f_S}{2f_B}\right)} = \frac{60dB}{20dB \cdot \lg\left(\frac{4000Hz}{3400Hz}\right)} = 42.5.$$

Exercise: A telephone signal is sampled at rate f_S=8KHz. Guaranteed bandwidth is f_B=3.4KHZ for an alias suppression of 60dB. Which filter order is required? (Solution below)

A particularly suitable filter type for moderate orders is the Butterworth filter with its characteristics of

$$|H_{BW}(f, f_B, N)| = \frac{1}{\sqrt{1 + (f / f_B)^{2N}}}$$

The Butterworth filter has a
- minimum ripple as the first 2N-1 derivatives are zero in f=0,
- $|H_{BW}(f_B)|$=-3dB for any N, with f_B being the cross point of the asymptotes.

However, it is difficult to build analog lowpass filters with orders > 8, as device tolerances become critical. The standard solution to this problem is a higher sampling rate.

Solution: $N = \dfrac{60dB}{20dB \cdot \lg\left(\frac{8000Hz}{2 \cdot 3400Hz}\right)} = 42.5.$

4.3 Anti-Aliasing Filters for Over Samplers

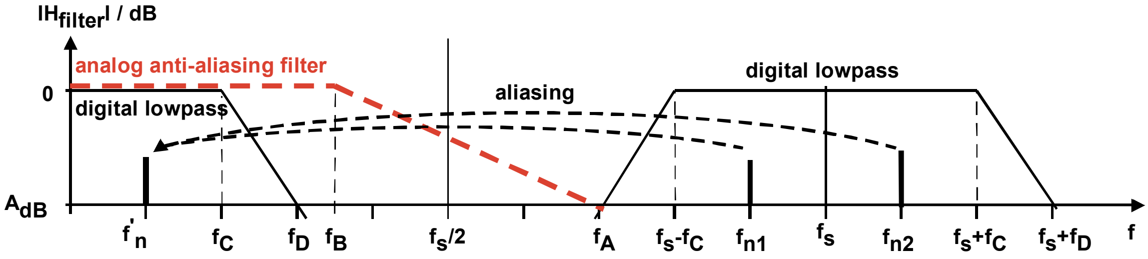


Fig. 4.3: Situation in frequency domain.

Due to oversampling, the anti-aliasing filter was shifted to the digital side, having cut-off frequency f_C and obtaining damping at f_D. On the digital side, any frequency domain characteristics is periodic in sampling frequency f_s.

If an analog anti-aliasing lowpass is required, then it may be strongly relaxed as attenuation needs to be obtained at $f_A = f_S - f_D$.

4.4 Spatial Aliasing

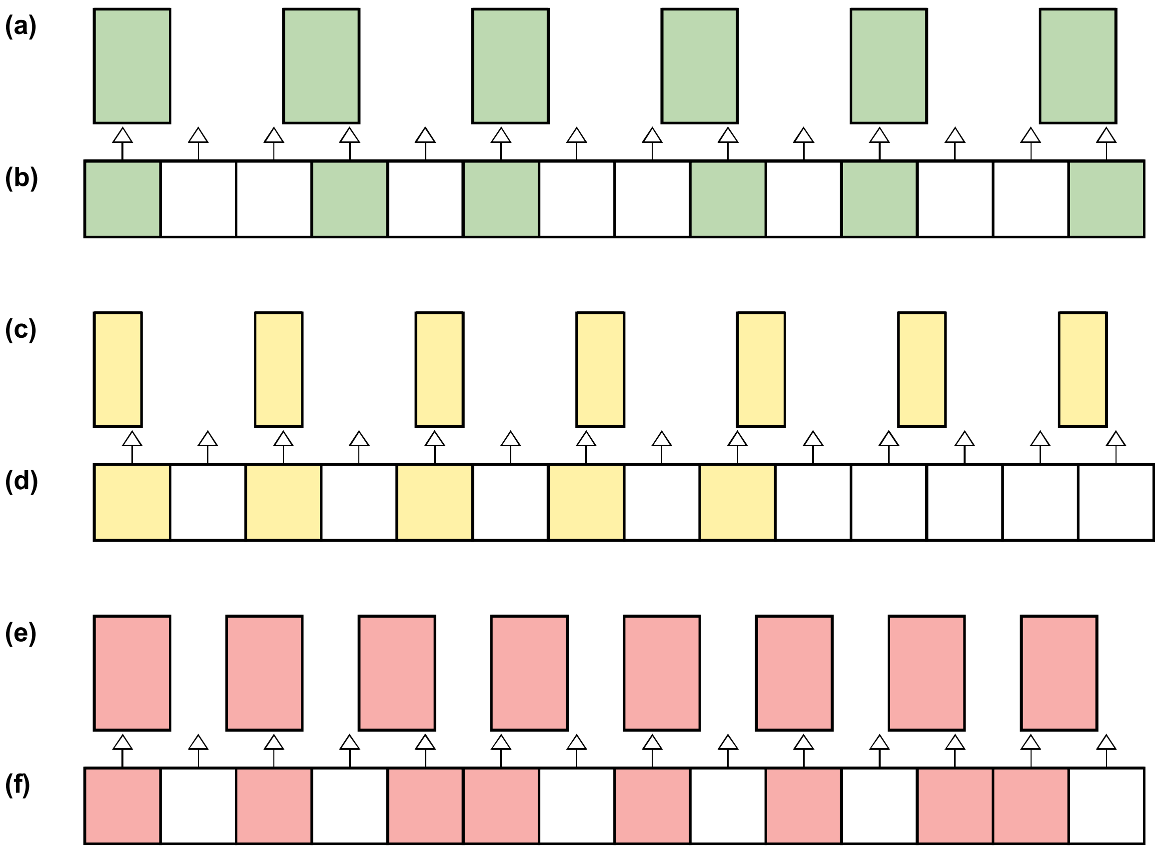


Fig. 4.4: Alle Strukturelemente in **Fig. part** (a) können in **Fig. part** (b) aufgelöst werden. **Fig. part** (c) hat eine Grundfrequenz, die aufgelöst werden könnte, aber in (d) nicht wird. **Fig. part** (e) hat eine nicht mehr auflösbare, räumliche Frequenz, (f) ist grob fehlerhaft.

Aliasing does not only occur over time axis. Fig. 4.4 illustrates spatial aliasing.

Fig. part (a) shows a periodic row of green squares. **Fig. part (b)** shows pixels to represent them. The arrows make the decision if a pixel is green or white. All green squares can be represented by the pixel density.

Fig. part (c) shows yellow squares on an image. Fig. part **(d)** illustrates that some of them are completely omitted, although Nyquist criterion for sinusoidal waves would predict that they could be mapped.

Fig. part (e) contains sufficiently wide red squares so that they cannot be omitted, but the frequency is subject to aliasing: 3rd and 4th square form one big red square in **part (f)**.

5 Changing Data Rates

5.1 Retrieve Time-Continuous Baseband

5.1.1 The Dirac Pulse Concept

A Dirac pulse is assumed to be infinitely narrow having a pulse area (and consequently integral) of 1. The integral over $a \cdot \delta(t-t_1)$ is a.

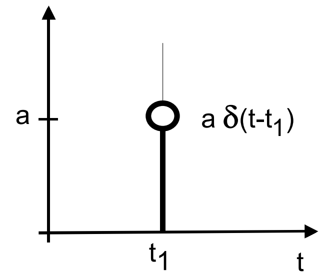
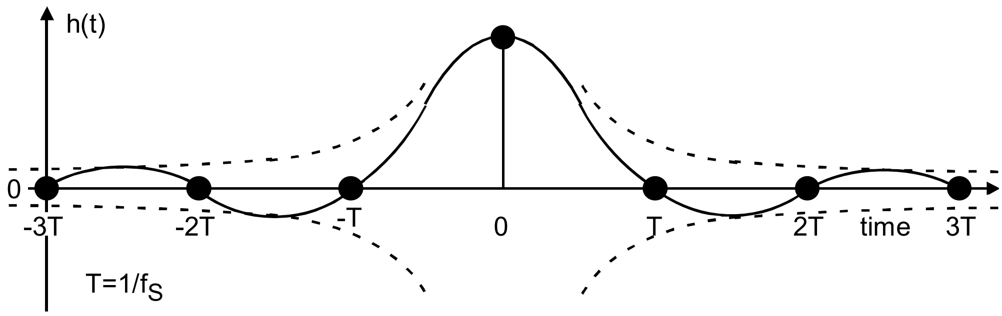


Fig. 5.1.1: Dirac function

5.1.2 Impulse Response of the Ideal Lowpass

Fig. 5.1.2: Cutout of impulse response of the ideal lowpass with cut-off frequency ½*fS*.



The impulse response of the ideal lowpass with cut-off frequency $f_C = f_S/2$ is a sinc function having the same zeros like a sinusoidal function with frequency f_C. It is defined as

$$\text{sinc(t)} = \frac{1}{\pi t} \cdot \sin \pi t \quad \text{with} \quad \frac{\sin x}{x} = \frac{1}{x} \cdot \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} ... \right) = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \frac{x^8}{9!} ...$$

We see that

$$\text{sinc(t)} \cong 1 - \frac{(\pi t)^2}{3!} \xrightarrow{t \to 0} 1$$

The impulse response of the ideal lowpass with cut-off frequency $f_C = f_S/2$ has only one tap different from zero.

Remark: An ideal lowpass does exist in real world, because it is not causal, i.e. it would have to begin with its impulse response before the impulse arrived.

5.1.3 Interpolating Sampled Values Using an Ideal Lowpass

Fig. 5.1.3: Interpolation of samples with ideal lowpass.

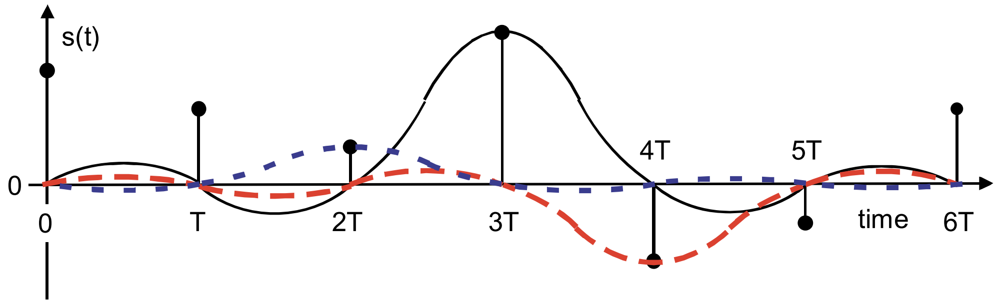


Fig. 5.1.3 shows arbitrary samples $s_n(t_n)$ at time points $t_n=nT$. For better overview only three impulse responses of the ideal lowpass with cut-off frequency $f_C=f_S/2$ were printed: $h(2T)$ with blue points, $h(3T)$ black solid and $h(4T)$ red dashed. Any impulse response crosses "its" impulse and is zero at any other sampling time point. As the ideal lowpass is an LTI (linear and time invariant) system, it sums all impulse responses. This sum is its output signal, which is crossing all sampled values.

As any of the impulse responses is infinitely smooth (i.e. all derivatives are smooth), it is obvious that also their sum has to be infinitely smooth. This way using a sum of "shape functions" for interpolation is a method widely used in mathematics (e.g. Lagrange interpolation).

5.2 Up-sampling: Redundant Increase of Sampling Rate

A CD delivers a data stream of 44.1 KHz, so that sound can be replayed with 22.05 KHz. Fig. 5.2 illustrates, that it is difficult for an analog smoothing filter to separate the spectra around f=0 and $f=f_S$. To make things easier we want to up-sample the CD's data stream by a factor 4.

With digital filters we can modify the data stream to the result shown in Fig. part (e), so that the demands to the analog smoothing filter are strongly relaxed, which is indicated by lowpass $|H_{TP,ana}(f)|$ in Fig. part(f).

Up-sampling causes redundancy, because we can remove the added samples without loss of information. This sample removal is called down-sampling or decimation.

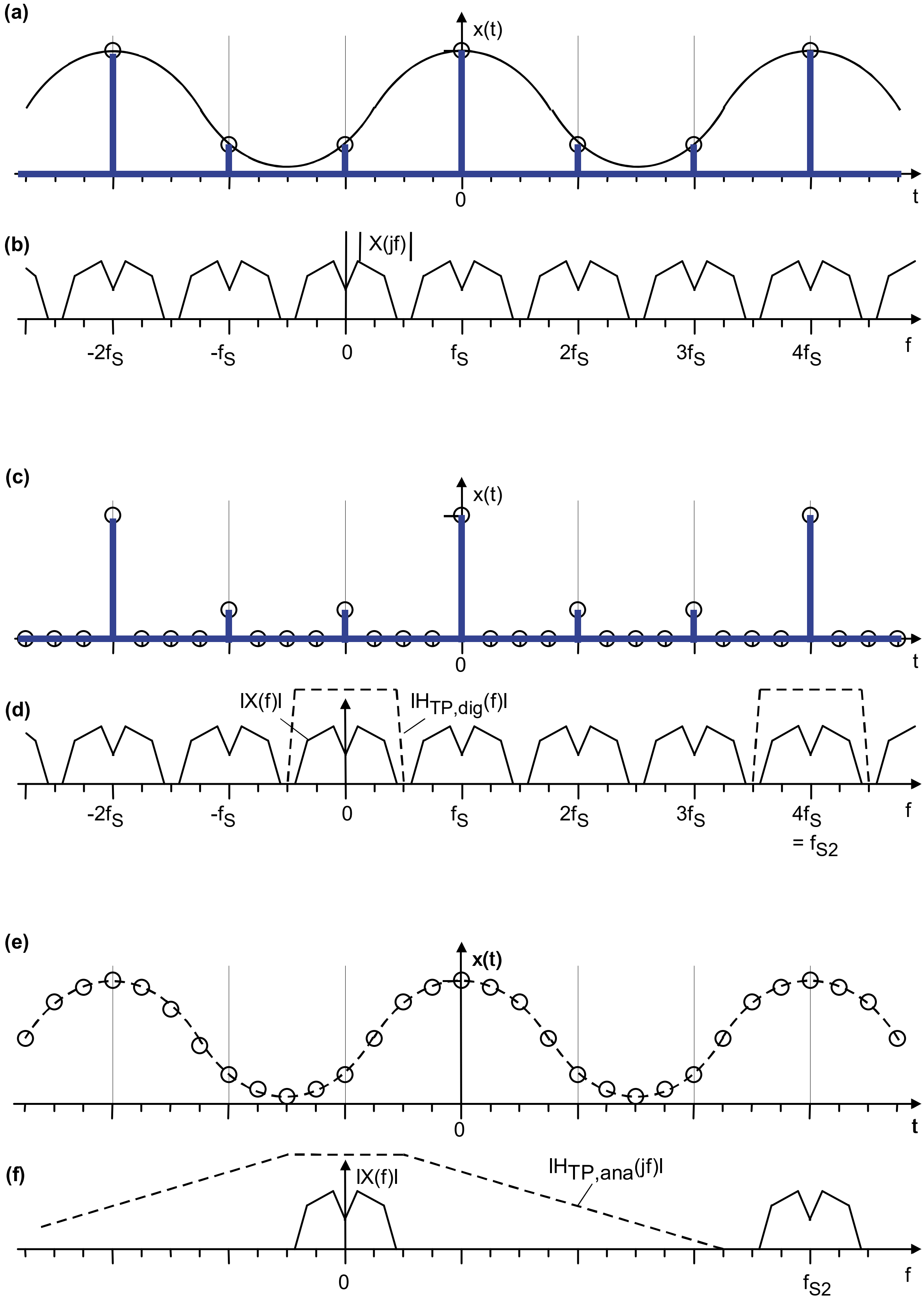


Fig. 5.2: Up-sampling a signal by adding additional samples without changing the baseband information.

The Approach in Fig. 5.2:

- **Fig. part (a)** shows an analog signal as shape and origin of sampled pulses. Sampling frequency f_S is slightly higher than $2f_B$.
- **Fig. part (b)** shows the frequency spectrum of the sampled pulses in Fig. part (a).
- **Fig. part (c)** we added zero samples into any gap of samples in Fig. part (b). To do so we had to increase the clock frequency by a factor four: f_2=4s.
- **Fig. part (d)** shows the frequency spectrum of Fig. part(c). As adding zeros did not change the time-domain function in (c) compared to (a), there cannot be a change in the frequency domain function of (d) compared to (b).
- **Fig. part (e)** is the situation of applying a digital filter $H_{TP,dig}(f)$ indicated by dashed lines in part (d).
- **Fig. part (f)** illustrates that the spectra around f_S, $2f_S$ and $3f_S$ were removed by the digital lowpass. $H_{TP,ana}(f)$ indicates the analog smoothing lowpass - which will most probably not required as nobody can hear the frequencies around f_{S2}.

Remark: High-order lowpasses are nowadays easier build in digital than in analog technology. Furthermore, digital transfer functions can be exactly reconstructed, while analog filters depend on device tolerances.

5.3 Decimation (=Down-Sampling)

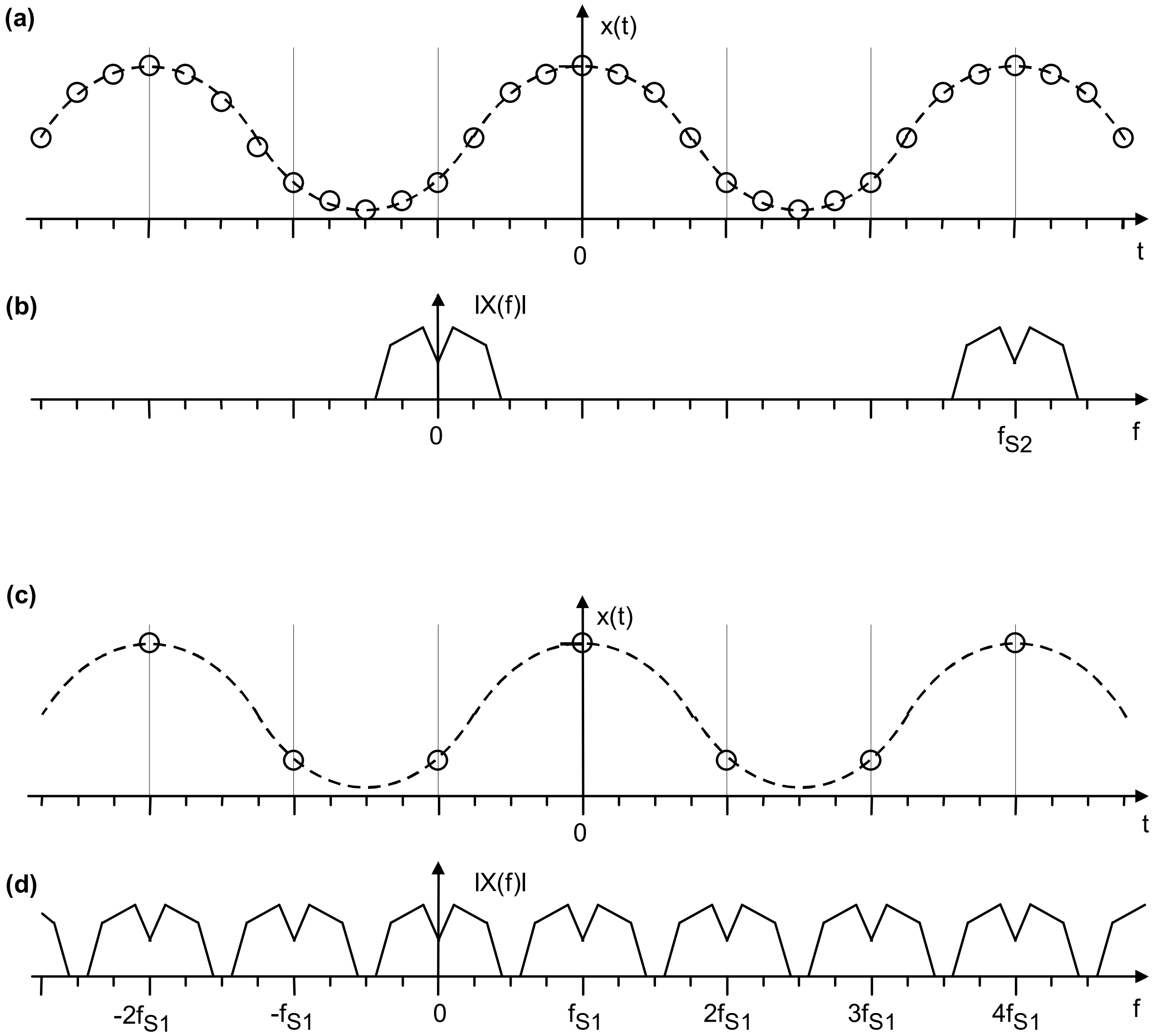


Fig. 5.3: Down-sampling a data stream by a factor four.

Decimation or down-sampling is sampling rate reduction. It is the inversion of up-sampling.

The sequence in Fig. 5.3 illustrates:

- **Fig. part (a)** shows a digital signal limited to $f<f_s/8$, i.e. with oversampling ratio *OSR* > 4.
- **Fig. part (b)** shows the frequency spectrum of (a).
- **Fig. part (c)** shows the decimated signal of (a): 3 of 4 samples were removed.
- **Fig. part (d)** illustrates the frequency spectrum of (c) after lowering the sampling rate by a factor 4.

Note: Decimation or down-sampling is a reduction of a time-discrete sampling rate. It is typically preceded by a lowpass to avoid aliasing.

5.4 Sub-Sampling

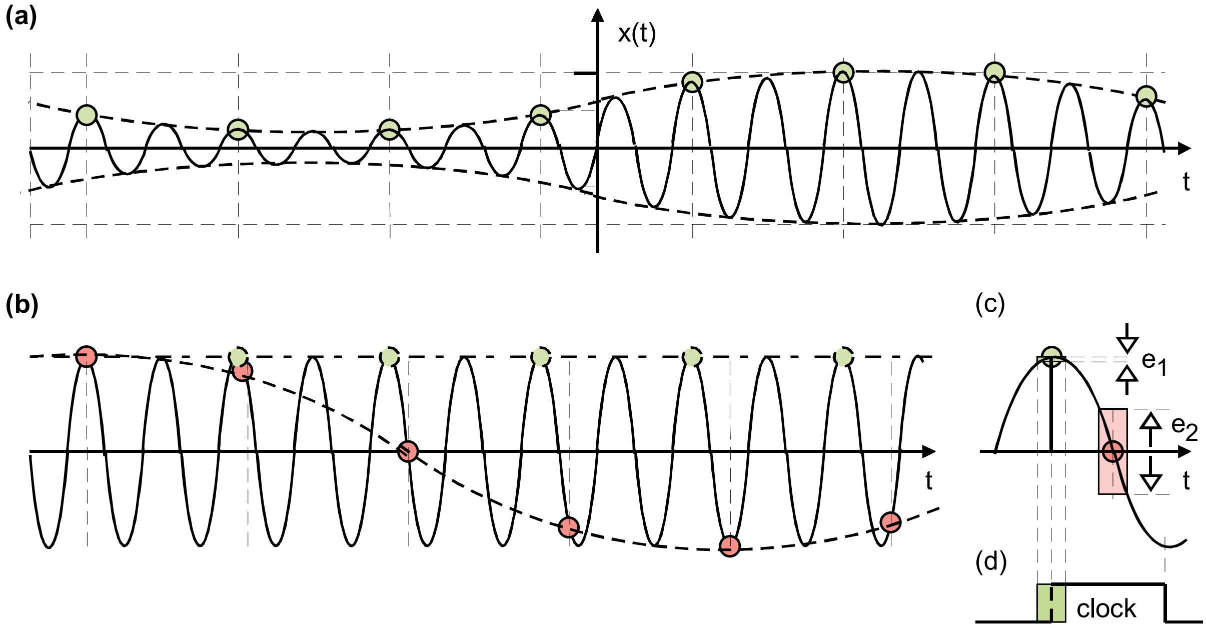


Fig. 5.4: (a) Demodulation of an amplitude modulated Signal by sub-sampling; (b) the carrier frequency is not exactly an integral multiple of the sampling frequency, we get an erroneous difference signal; (c) optimal (green) and worst (red) sampling time point; (d) sampling signal in time domain.

Fig. 5.4 (a) illustrates, that aliasing can be used to demodulate amplitude modulated (AM) signals.

Fig. part (b) illustrates, that this kind of AM demodulation has to sample the correct phase (green) to pick the tops of the carrier signal (green). When the frequency relation $f_{carrier}/f_S$ is not exactly integral, we generate additionally erroneous difference signals.

Fig. part (c) illustrates, that the optimum carrier phase to sample is its maximum (or minimum). This yields maximum sample amplitude a_1, minimum error e_1 and consequently minimum e_1/a_1, caused by phase noise of the sampling signal. The error e_2 occurs by same phase noise near the carries zero crossing (a_2=0), maximum slope and consequently $e_2/a_2 \rightarrow \infty$.

Fig. part (d) shows the positive edge of the sampling clock and its phase noise.

Note: Sub-sampling takes advantage of aliasing. It typically comes with a PLL (Phase Locked Loop) to guarantee sampling at exact phase and frequency.

Change of Signal Sampling Rate:

Up-sampling (interpolation) is increasing sampling frequency by an integral factor by introduction of additional sampling time point with sampling value zero and subsequent interpolation (lowpass, anti-aliasing) filtering.

Down-sampling (decimation, re-sampling) is lowering the sampling frequency by an integral factor avoiding aliasing. Therefore, anti-aliasing lowpass filters are required before sampling.

Sub-sampling is taking advantage of aliasing for demodulation and comes typically with a phase-locked loop (PLL)

6 References

[1] W. Kellermann, Vorlesung „Digitale Nachrichtensysteme“, FH Regensburg, 1998
[2] U. Tietze, Ch. Schenk, Halbleiterschaltungstechnik, Kap. 24: Digitale Filter, 10. Auflage, Springer Verlag