Chair of Distributed Systems and Security
School of Computation, Information and Technology
Technical University of Munich

Computer Networking and IT Security (CNS)

INHN0012 – WiSe 2024/25

Prof. Dr.-Ing. Stephan Günther

Chair of Distributed Systems and Security
School of Computation, Information and Technology
Technical University of Munich

Friday 1st November, 2024 08:03

Chapter 1: Physical layer

Signals, information, and their meaning

Definition – Signals and symbols
Signals are time-dependent and measurable physical quantities. Defined measurable signal changes can be assigned a symbol. These symbols are the physical representation of information.

Examples for signals

- light, e. g. transmission of Morse code in navigation
- voltage, e. g. telegraphy
- sound, e. g. language and music

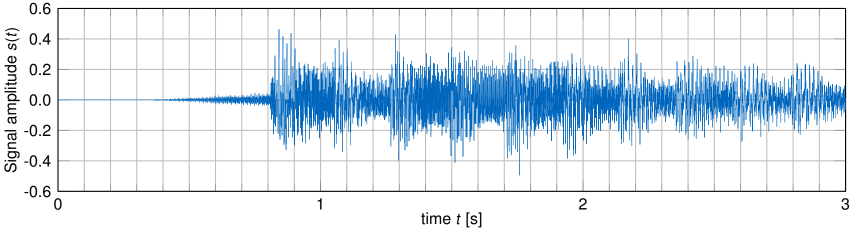


Figure 1: The first 3 s of „Sunrise Avenue – Hollywood Hills“

Definition – Information content
The information content of a symbol (sign or character) expresses how much information is transmitted by the sign.

The information content has the following properties:

- The less frequently a character occurs, the higher its information content.
- The information content of a string is the sum of the information content of the individual characters provided that characters occur independently from each other.
- The information content of predictable characters is 0

The logarithm is the simplest function to define the information content with these properties.

Claude Elwood Shannon

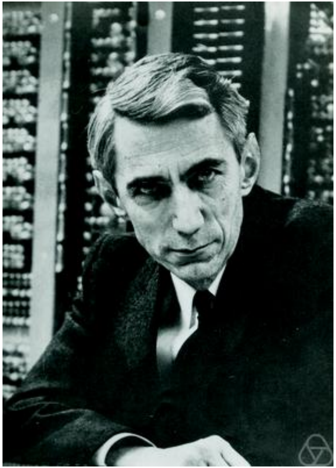
∗ April 30, 1916 † February 24, 2001
AT&T Bell Labs: 1941–1958, then professor at the MIT

A Mathematical Theory of Communication
by Claude E. Shannon
In: The Bell System Technical Journal,
Vol. 27, No 3, 1948, pp. 379–423 and
Vol. 27, No 4, 1948, pp. 623–656
https://dl.acm.org/doi/pdf/10.1145/584091.584093

Communication in the Presence of Noise
by Claude E. Shannon
Proc. Inst. Radio Eng. (IRE) Vol. 37, 1949, pp.10-21
Online retyped copy of the paper:
https://www.noisebridge.net/images/e/e5/Shannon_noise.pdf

Communication Theory of Secrecy Systems
by Claude E. Shannon
In: The Bell System Technical Journal,
Vol. 28, No. 4, 1949, pp. 656–715
http://netlab.cs.ucla.edu/wiki/files/shannon1949.pdf

Prediction and Entropy of Printed English
by Claude E. Shannon
In: The Bell System Technical Journal,
Vol. 30, No. 1, 1951, pp. 50–64,
https://archive.org/details/bstj30-1-50



Definition – Information

Information consists in the uncertainty of being able to predict changes in a signal. The information content of a character $x \in \mathcal{X}$ from an alphabet $\mathcal{X}$ depends on the probability $p(x)$ that the information-carrying signal takes on the value or range of values assigned to this character at the time of observation. The information content I of the character x with probability of occurrence $p(x)$ is defined as

$$I(x) = -\log_2 p(x) \quad \text{mit} \quad [I] = \text{bit}.$$

[1] Will be covered in Theory of Computation and Information Theory (INHN0013).

Definition – Information

Information consists in the uncertainty of being able to predict changes in a signal. The information content of a character $x \in \mathcal{X}$ from an alphabet $\mathcal{X}$ depends on the probability $p(x)$ that the information-carrying signal takes on the value or range of values assigned to this character at the time of observation. The information content I of the character x with probability of occurrence $p(x)$ is defined as

$$I(x) = -\log_2 p(x) \quad \text{mit} \quad [\,I\,] = \text{bit}.$$

Definition – Entropy

The average information content of a source is called entropy

$$H(X) = \sum_{x \in \mathcal{X}} p(x) I(x) = -\sum_{x \in \mathcal{X}} p(x) \log_2 (p(x))\,.$$

[1] Will be covered in Theory of Computation and Information Theory (INHN0013).

Definition – Information
Information consists in the uncertainty of being able to predict changes in a signal. The information content of a character $x \in \mathcal{X}$ from an alphabet $\mathcal{X}$ depends on the probability $p(x)$ that the information-carrying signal takes on the value or range of values assigned to this character at the time of observation. The information content I of the character x with probability of occurrence $p(x)$ is defined as

$$I(x) = -\log_2 p(x) \quad \text{mit} \quad [\,I\,] = \text{bit}.$$

Definition – Entropy
The average information content of a source is called entropy

$$H(X) = \sum_{x \in \mathcal{X}} p(x) I(x) = -\sum_{x \in \mathcal{X}} p(x) \log_2 (p(x)) \,.$$

Note: We sometimes use the notations $p(x)$ or p_x as a short form of $\Pr[X = x]$ (read as "the probability that the random variable X takes the value x")[1].

[1] Will be covered in Theory of Computation and Information Theory (INHN0013).

Examples:

1. Deterministic, discrete source that always emits the character 'A':

$\boxed{X} \longrightarrow$ A A A A A . . .

$$I(\mathrm{A}) = -\log_2\left(\Pr[X = \mathrm{A}]\right) = -\log_2(1) = 0\,\mathrm{bit}$$

Information and entropy

Examples:

1. Deterministic, discrete source that always emits the character 'A':

$$\boxed{X} \longrightarrow \text{A A A A A} \ldots \qquad I(\text{A}) = -\log_2\left(\Pr[X = \text{A}]\right) = -\log_2(1) = 0 \text{ bit}$$

2. Binary, discrete source which emits the characters '0' or '1' in a completely unpredictable way:

$$\boxed{X} \longrightarrow 0\,1\,1\,0\,1 \ldots$$

$$I(0) = -\log_2\left(\Pr[X = 0]\right) = -\log_2(0.5) = 1 \text{ bit}$$

$$I(1) = -\log_2\left(\Pr[X = 1]\right) = -\log_2(0.5) = 1 \text{ bit}$$

The entropy $H(X) = \sum_i p_i I(x_i)$ of thus source is

$$H(X) = -(p_0 \log_2(p_0) + p_1 \log_2(p_1)) = -(-0.5 - 0.5) = 1 \text{ bit/symbol}.$$

Information and entropy

Examples:

1. Deterministic, discrete source that always emits the character 'A':

   $\boxed{X} \longrightarrow$ A A A A A ...

   $$I(\mathrm{A}) = -\log_2\left(\Pr[X = \mathrm{A}]\right) = -\log_2(1) = 0\,\text{bit}$$

2. Binary, discrete source which emits the characters '0' or '1' in a completely unpredictable way:

   $\boxed{X} \longrightarrow$ 0 1 1 0 1 ...

   $$I(0) = -\log_2\left(\Pr[X = 0]\right) = -\log_2(0.5) = 1\,\text{bit}$$

   $$I(1) = -\log_2\left(\Pr[X = 1]\right) = -\log_2(0.5) = 1\,\text{bit}$$

   The entropy $H(X) = \sum_i p_i I(x_i)$ of thus source is

   $$H(X) = -\left(p_0 \log_2(p_0) + p_1 \log_2(p_1)\right) = -(-0.5 - 0.5) = 1\,\text{bit}/\text{symbol}.$$

3. Unordered characters of a long German text, i. e., $X \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \dots, \mathrm{Z}\}$:

   $\boxed{X} \longrightarrow$ E W T I L E M H C A B ...

   $$I(\mathrm{E}) = -\log_2\left(\Pr[X = \mathrm{E}]\right) = -\log_2(0.1740) \approx 2.52\,\text{bit}$$

   The entropy $H(X)$ of this source is

   $$H(X) = -\sum_{i=1}^{N} p_i \log_2(p_i) \approx 4.0629\,\text{bit}/\text{symbol},$$

   i. e., German text can be encoded with slightly more than 4 bit per character on average.

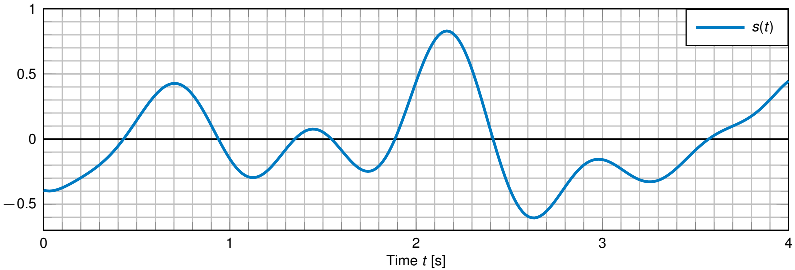
Note: This applies only to memoryless sources or sufficiently long texts, respectively. Otherwise, conditional probabilities must be considered.

What is the meaning of specific signal?
A signal transports information. Only by an interpretation rule this information gets a meaning, i. e., there must be a mapping between symbols (physical signal values or value ranges) and data.

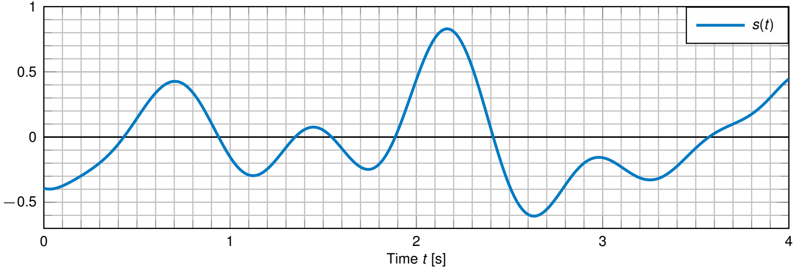
Example: Given a binary alphabet with the characters $X \in \{0,1\}$. The interpretation rule is

$$x = \begin{cases} 0 & s(t) \leq 0, \\ 1 & \text{otherwise.} \end{cases}$$

What is the meaning of the signal shown below?



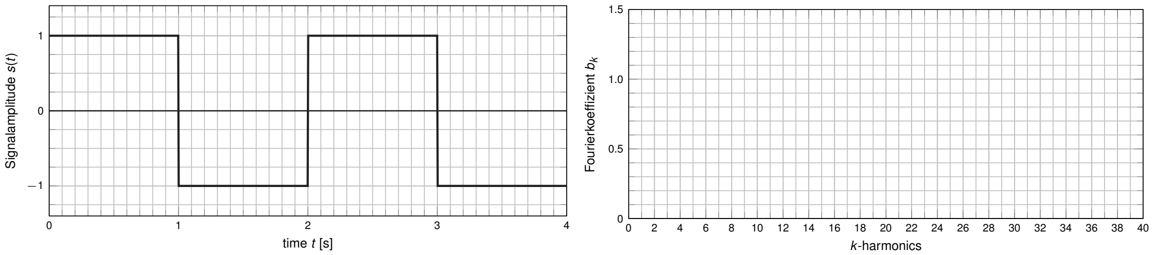
Open questions



- At what time intervals are samples taken? (Time discretization)
- Does more frequent sampling also automatically mean more information? (Sampling theorem)
- How to round continuous signal values? (Quantization)
- What is the interpretation rule of sampled data? (Line coding)
- Which interfering factors play a role? (noise, attenuation, distortion, . . .) (Channel coding)
- And how is such a signal generated in the first place? (Impulse shaping, modulation)

Chapter 1: Physical layer

Periodic time signals can be understood as a superposition of sine and cosine oscillations of different frequencies:

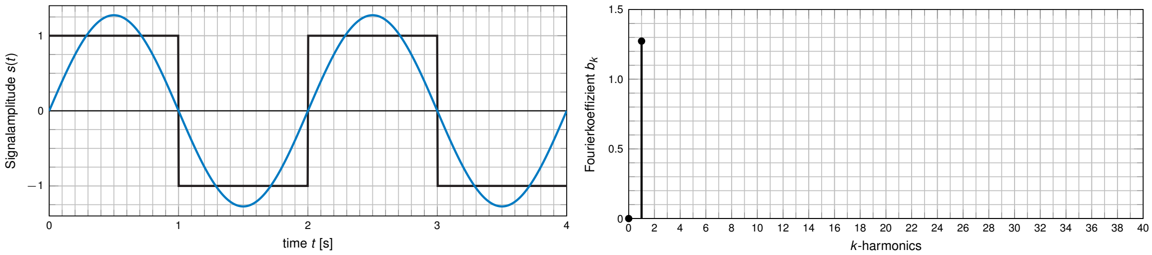


(a) time signal $s(t)$ (b) spectrum $S(f)$

Fourier Series: (with $\omega = 2\pi/T$, period $T = 2\,\text{s}$)

$$s(t) \approx \frac{a_0}{2}$$

Periodic time signals can be understood as a superposition of sine and cosine oscillations of different frequencies:

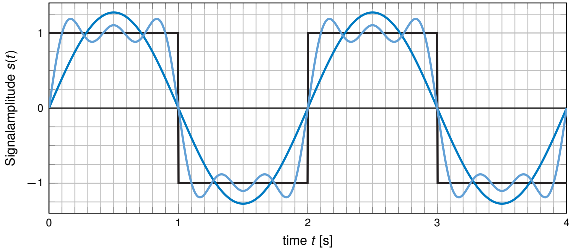


(a) time signal $s(t)$ (b) spectrum $S(f)$

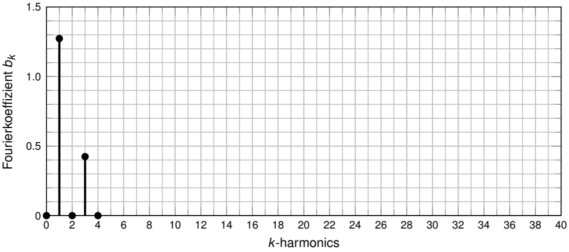
Fourier Series: (with $\omega = 2\pi/T$, period $T = 2\,\mathrm{s}$)

$$s(t) \approx \frac{a_0}{2} + a_1 \cos(\omega t) + b_1 \sin(\omega t)$$

Periodic time signals can be understood as a superposition of sine and cosine oscillations of different frequencies:



(a) time signal $s(t)$

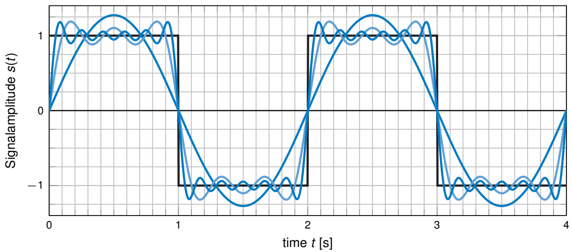


(b) spectrum $S(f)$

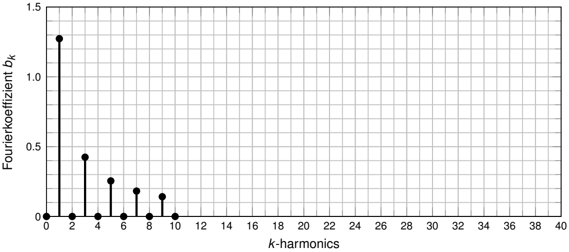
Fourier Series: (with $\omega = 2\pi / T$, period $T = 2\,\mathrm{s}$)

$$s(t) \approx \frac{a_0}{2} + \sum_{k=1}^{4} \left(a_k \cos(k\omega t) + b_k \sin(k\omega t)\right)$$

Periodic time signals can be understood as a superposition of sine and cosine oscillations of different frequencies:



(a) time signal $s(t)$

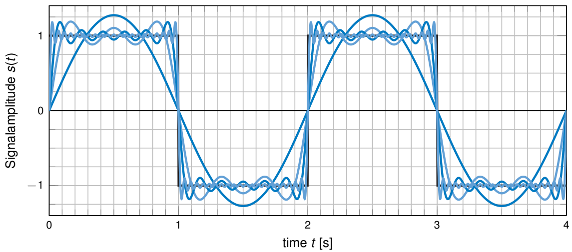


(b) spectrum $S(f)$

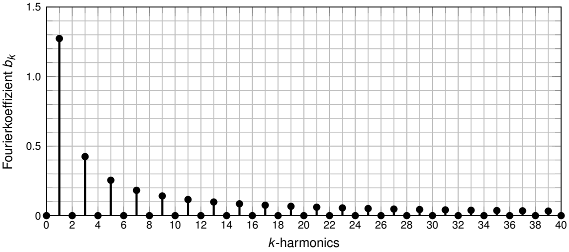
Fourier Series: (with $\omega = 2\pi/T$, period $T = 2\,\mathrm{s}$)

$$s(t) \approx \frac{a_0}{2} + \sum_{k=1}^{10} \left(a_k \cos(k\omega t) + b_k \sin(k\omega t)\right)$$

Periodic time signals can be understood as a superposition of sine and cosine oscillations of different frequencies:



(a) time signal $s(t)$

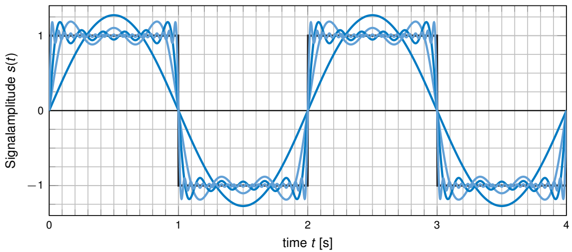


(b) spectrum $S(f)$

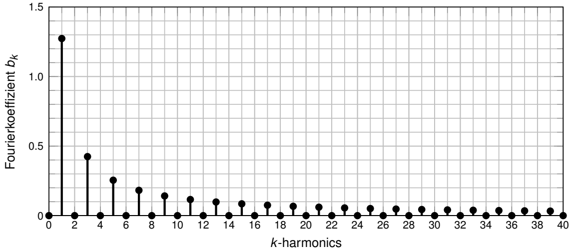
Fourier Series: (with $\omega = 2\pi/T$, period $T = 2\,\mathrm{s}$)

$$s(t) \approx \frac{a_0}{2} + \sum_{k=1}^{40} \left(a_k \cos(k\omega t) + b_k \sin(k\omega t)\right)$$

Periodic time signals can be understood as a superposition of sine and cosine oscillations of different frequencies:



(a) time signal $s(t)$



(b) spectrum $S(f)$

Fourier Series: (with $\omega = 2\pi/T$, period $T = 2\,\mathrm{s}$)

For the limit $N \to \infty$ holds:

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos(k\omega t) + b_k \sin(k\omega t) \right)$$

Fourier series

A periodic signal $s(t)$ can be reconstructed as the sum of weighted sinus and cosinus functions. The resulting series $s(t)$ is called Fourier series:

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos(k\omega t) + b_k \sin(k\omega t)\right).$$

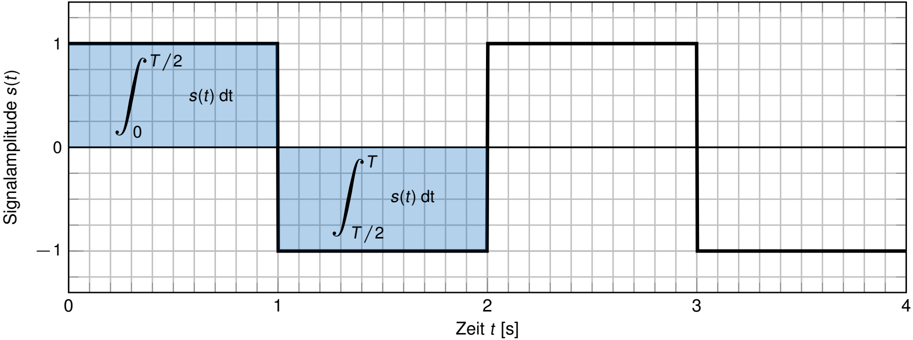
The sum component with index k is called the k-harmonic. The constant component $a_0/2$ represents a shift of the signals amplitude regarding the y-axis and therefore is a constant factor of the function. The angular frequency $\omega = 2\pi/T$ defines the periodicity T of the signal

The weights a_k und b_k can be calculated as follows:

$$a_k = \frac{2}{T}\int_0^T s(t)\cos(k\omega t)\,\mathrm{d}t \quad \text{and} \quad b_k = \frac{2}{T}\int_0^T s(t)\sin(k\omega t)\,\mathrm{d}t.$$

Signal properties

- Calculating the coefficients a_k and b_k can be done through simple calculations
- Some signal properties can be seen directly:

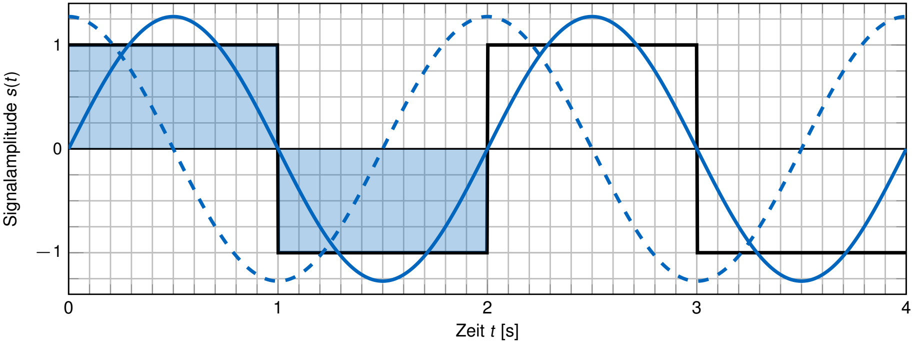


- Point symmetry regarding $\left(T/2, 0\right) \quad \Rightarrow \quad a_0 = \frac{2}{T} \int_0^T s(t)\,dt = 0$
- No constant component

Question: What holds for the signal $s'(t) = s(t) + c$ with $c > 0$?

Signal properties

- Calculating the coefficients a_k and b_k can be done through simple calculations
- Some signal properties can be seen directly:



- All weights of the cosinus components are zero, that is $a_k = 0 \; \forall k \in \mathbb{N}$
- Reason: $s(t)$ is in-phase with the sinus

Question: What happens if we shift $s(t)$ by 90°?

Fourier Transform

So far, we have only looked at periodic signals. So what happens if we introduce non-periodic signals to the equation?

- We cannot use the fourier series anymore
- We result with a continuous spectrum, rather than a discrete one

Fourier Transformation

The fourier transform of a steady rising, integrable function $s(t)$ is defined as

$$s(t) \quad \circ\!\!-\!\!\bullet \quad S(\omega) = \int_{-\infty}^{\infty} s(t)e^{-j\omega t}\,\mathrm{d}t = \int_{-\infty}^{\infty} s(t)\,(\cos(\omega t) - j\sin(\omega t))\,\mathrm{d}t,$$

where j denotes the imaginary unit and $\omega = 2\pi f$ the angular frequency. The equivalency $e^{jx} = \cos(x) + j\sin(x)$ is known as Euler's formula.

So far, we have only looked at periodic signals. So what happens if we introduce non-periodic signals to the equation?

- We cannot use the fourier series anymore
- We result with a continuous spectrum, rather than a discrete one

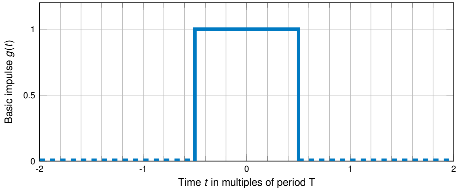
Fourier Transformation

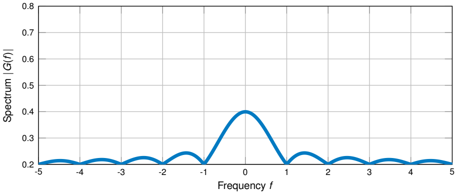
The fourier transform of a steady rising, integrable function $s(t)$ is defined as

$$s(t) \circ\!\!-\!\!\bullet\; S(\omega) = \int_{-\infty}^{\infty} s(t)e^{-j\omega t}\,\mathrm{d}t = \int_{-\infty}^{\infty} s(t)\,(\cos(\omega t) - j\sin(\omega t))\,\mathrm{d}t\,,$$

where j denotes the imaginary unit and $\omega = 2\pi f$ the angular frequency. The equivalency $e^{jx} = \cos(x) + j\sin(x)$ is known as Euler's formula.

Example: Square impulse and associated spectrum





Chapter 1: Physical layer

Sampling, reconstruction, and quantization

Naturally occurring signals are continuous in time and continuous in value, i.e., they take on arbitrary real values at infinitely many points in time.

Problem for computers:

- limited memory
- limited precision

Sampling, reconstruction, and quantization

Naturally occurring signals are continuous in time and continuous in value, i.e., they take on arbitrary real values at infinitely many points in time.

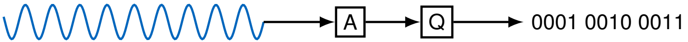
Problem for computers:

- limited memory
- limited precision

Solution approach: Discretization of signals in the

- time domain (sampling) and
- value domain (quantization).

A discrete-time and discrete-value signal is digital and is stored in fixed-length words.



Comparison: Usage of fixed or floating point numbers instead of real numbers corresponds to rounding (quantization) to a finite number of discrete steps.

The signal $s(t)$ is sampled using the unit pulse (Dirac pulse) $\delta[t]$ at equidistant intervals T_a (sampling interval) for $n \in \mathbb{Z}$:

$$\hat{s}(t) = s(t) \sum_{n=-\infty}^{\infty} \delta[t - nT_a], \text{ with } \delta[t - nT_a] = \begin{cases} 1 & t = nT_a, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\hat{s}(t)$ is nonzero only at times nT_a for integer n, we agree on the notation $\hat{s}[n]$ for discrete-time but continuous-value signals.

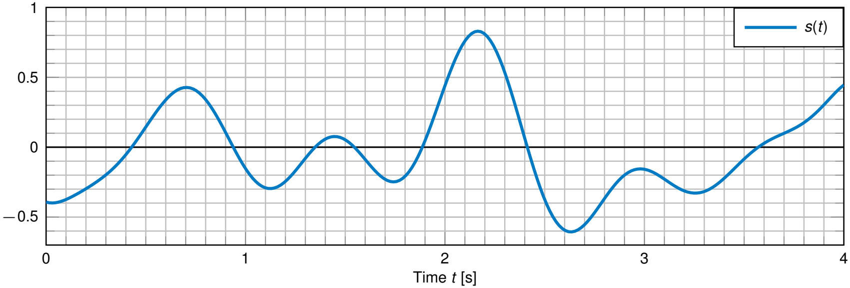


Figure 2: Time continuous signal $s(t)$

The signal $s(t)$ is sampled using the unit pulse (Dirac pulse) $\delta[t]$ at equidistant intervals T_a (sampling interval) for $n \in \mathbb{Z}$:

$$\hat{s}(t) = s(t) \sum_{n=-\infty}^{\infty} \delta[t - nT_a], \text{ with } \delta[t - nT_a] = \begin{cases} 1 & t = nT_a, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\hat{s}(t)$ is nonzero only at times nT_a for integer n, we agree on the notation $\hat{s}[n]$ for discrete-time but continuous-value signals.

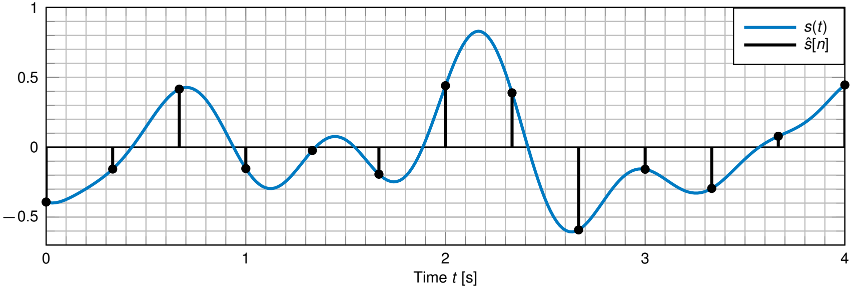


Figure 2: Time continuous signal $s(t)$ and corresponding discrete samples $\hat{s}[n]$

Reconstruction

By means of the samples $\hat{s}[n]$ it is possible to approximate or even reconstruct the original signal $s(t)$:

$$s(t) \approx \sum_{n=-\infty}^{\infty} \hat{s}[n] \cdot \operatorname{sinc}\left(\frac{t - nT_a}{T_a}\right).$$

- Samples are support points and
- serve as weights for a suitable approximate function (trigonometric interpolation, cf. polynomial interpolation).

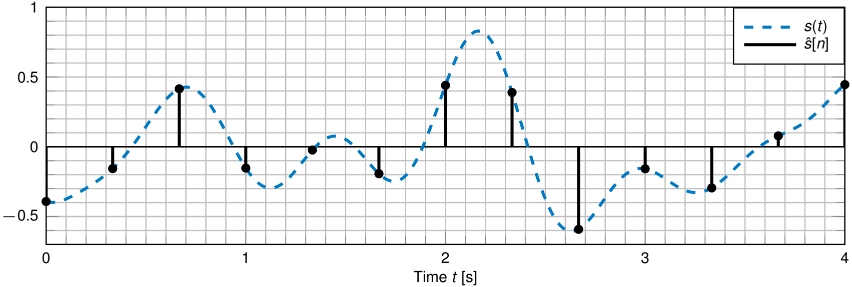


Figure 3: Samples $\hat{s}[n]$

Reconstruction

By means of the samples $\hat{s}[n]$ it is possible to approximate or even reconstruct the original signal $s(t)$:

$$s(t) \approx \sum_{n=-\infty}^{\infty} \hat{s}[n] \cdot \operatorname{sinc}\left(\frac{t - nT_a}{T_a}\right).$$

- Samples are support points and
- serve as weights for a suitable approximate function (trigonometric interpolation, cf. polynomial interpolation).

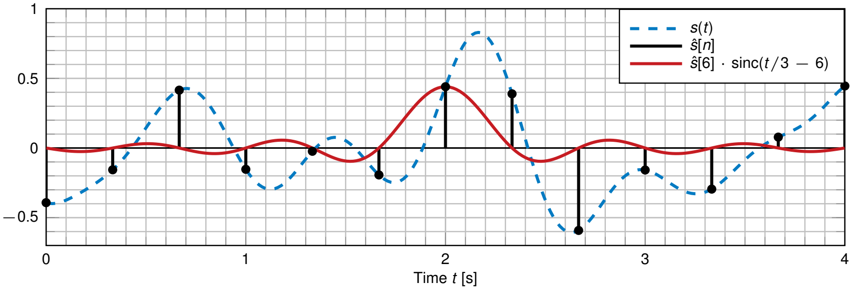


Figure 3: Each sample $\hat{s}[n]$ serves as a weight for the reconstruction function.

Reconstruction

By means of the samples $\hat{s}[n]$ it is possible to approximate or even reconstruct the original signal $s(t)$:

$$s(t) \approx \sum_{n=-\infty}^{\infty} \hat{s}[n] \cdot \operatorname{sinc}\left(\frac{t - nT_a}{T_a}\right).$$

- Samples are support points and
- serve as weights for a suitable approximate function (trigonometric interpolation, cf. polynomial interpolation).

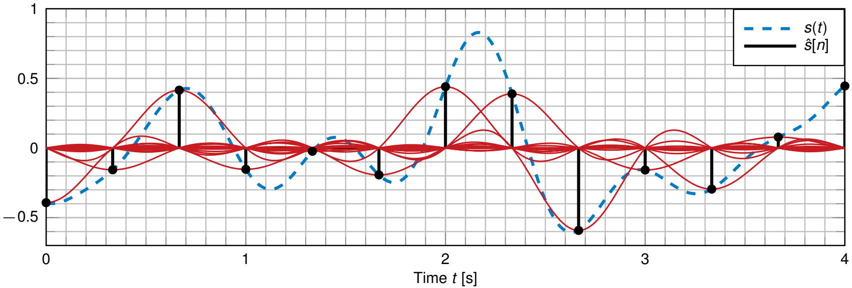


Figure 3: Each sample $\hat{s}[n]$ serves as a weight for the reconstruction function.

Reconstruction

By means of the samples $\hat{s}[n]$ it is possible to approximate or even reconstruct the original signal $s(t)$:

$$s(t) \approx \sum_{n=-\infty}^{\infty} \hat{s}[n] \cdot \operatorname{sinc}\left(\frac{t - nT_a}{T_a}\right).$$

- Samples are support points and
- serve as weights for a suitable approximate function (trigonometric interpolation, cf. polynomial interpolation).

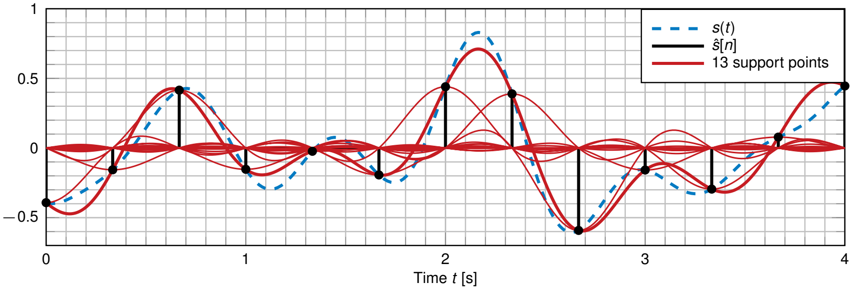


Figure 3: The sum of the weighted functions approaches the original signal depending on the number of summing elements.

Reconstruction

By means of the samples $\hat{s}[n]$ it is possible to approximate or even reconstruct the original signal $s(t)$:

$$s(t) \approx \sum_{n=-\infty}^{\infty} \hat{s}[n] \cdot \operatorname{sinc}\left(\frac{t - nT_a}{T_a}\right).$$

- Samples are support points and
- serve as weights for a suitable approximate function (trigonometric interpolation, cf. polynomial interpolation).

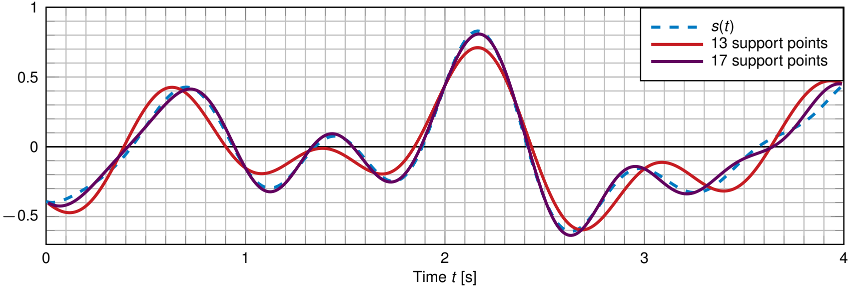


Figure 3: The sum of the weighted functions approaches the original signal depending on the number of summing elements.

Reconstruction

Under which conditions is a lossless reconstruction possible?

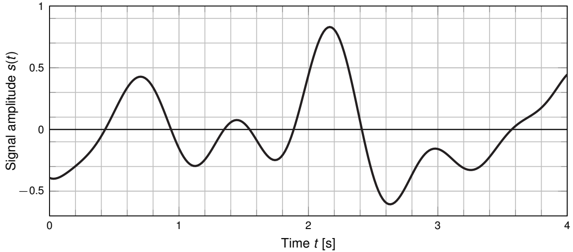
- Multiplication in the time domain corresponds to convolution in the frequency domain:

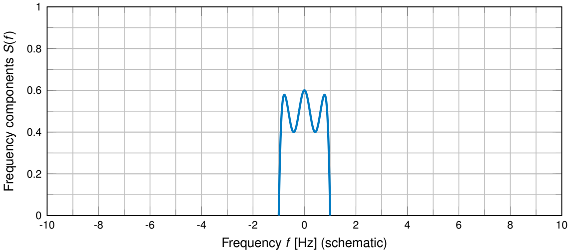
$$s(t) \cdot \delta[t - nT] \quad \circ\!\!-\!\!\bullet \quad \frac{1}{T} S(f) * \delta[f - n/T].$$

- This convolution with unit pulses corresponds to a shift of $S(f)$ along the abscissa.

Consequently, the sampling of the signal $s(t)$ at intervals T_a corresponds to the periodic repetition of its spectrum $S(f)$ at intervals $f_a = {}^1/_{T_a}$.

Example: Sampling of a signal $s(t)$ band limited on some maximum frequency B with sampling frequency $f_a = 4B$:





Under which conditions is a lossless reconstruction possible?

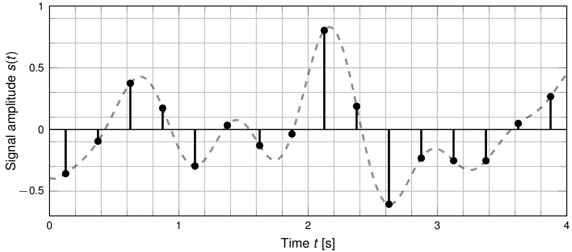
- Multiplication in the time domain corresponds to convolution in the frequency domain:

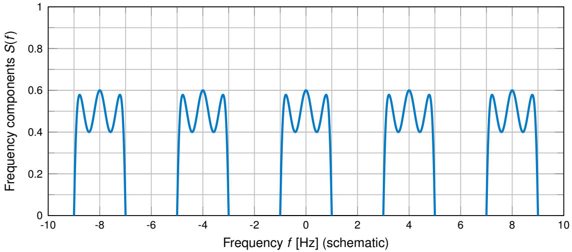
$$s(t) \cdot \delta[t - nT] \quad \circ\!\!-\!\!\bullet \quad \frac{1}{T} S(f) * \delta[f - n/T].$$

- This convolution with unit pulses corresponds to a shift of $S(f)$ along the abscissa.

Consequently, the sampling of the signal $s(t)$ at intervals T_a corresponds to the periodic repetition of its spectrum $S(f)$ at intervals $f_a = {}^1/_{T_a}$.

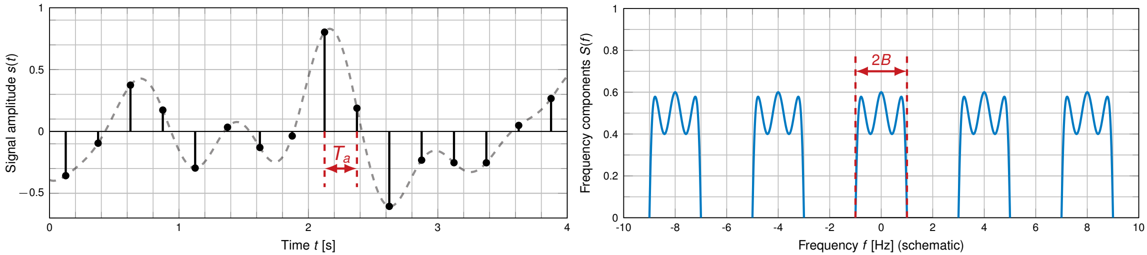
Example: Sampling of a signal $s(t)$ band limited on some maximum frequency B with sampling frequency $f_a = 4B$:





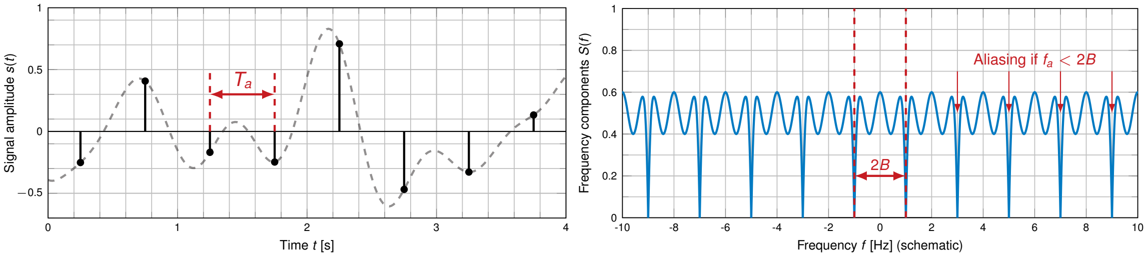
Shannon and Nyquist sampling theorem

A signal $s(t)$ band-limited to $|f| \leq B$ is fully described by equidistant samples $\hat{s}[n]$, provided they are no farther apart than $T_a \leq {}^{1}/{}_{2B}$. The sampling frequency, which allows a complete signal reconstruction, is consequently bounded below by $f_a \geq 2B$.



Shannon and Nyquist sampling theorem

A signal $s(t)$ band-limited to $|f| \leq B$ is fully described by equidistant samples $\hat{s}[n]$, provided they are no farther apart than $T_a \leq 1/2B$. The sampling frequency, which allows a complete signal reconstruction, is consequently bounded below by $f_a \geq 2B$.



- If one chooses $f_a < 2B$, the periodic repetitions of the spectrum overlap
- This effect is known as aliasing
- In that case, a lossless reconstruction is no longer possible.

Quantization

The samples $\hat{s}[n] \in \mathbb{R}$ are still continuous in the range of values and cannot be stored exactly.

Solution: Quantization

- In order to differentiate between $M = 2^N$ signal stage, we need code words of N bit
- A specific code word is assigned to each signal stage in the process
- The signal stages are distributed in the quantization interval in a sensible way
- What is “sensible”?

Quantization

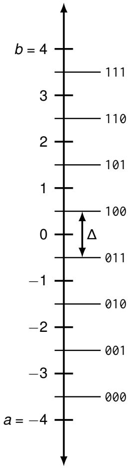
The samples $\hat{s}[n] \in \mathbb{R}$ are still continuous in the range of values and cannot be stored exactly.

Solution: Quantization

- In order to differentiate between $M = 2^N$ signal stage, we need code words of N bit
- A specific code word is assigned to each signal stage in the process
- The signal stages are distributed in the quantization interval in a sensible way
- What is “sensible”?

Example: Linear quantization with mathematical rounding

- This scheme is optimal if all values within I_Q occur with equal probability
- Step width $\Delta = \dfrac{b - a}{M}$
- Within I_Q the maximum quantization error is $q_{\max} = \Delta/2$
- Signal values outside I_Q are mapped to the largest or smallest signal stage, respectively $\Rightarrow$ the quantization error is unbounded outside of I_Q



Quantization

The samples $\hat{s}[n] \in \mathbb{R}$ are still continuous in the range of values and cannot be stored exactly.

Solution: Quantization

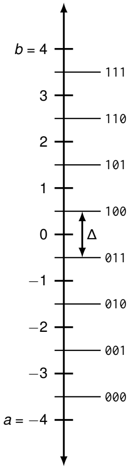
- In order to differentiate between $M = 2^N$ signal stage, we need code words of N bit
- A specific code word is assigned to each signal stage in the process
- The signal stages are distributed in the quantization interval in a sensible way
- What is “sensible”?

Example: Linear quantization with mathematical rounding

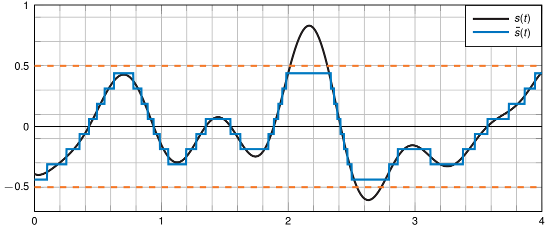
- This scheme is optimal if all values within I_Q occur with equal probability
- Step width $\Delta = \dfrac{b-a}{M}$
- Within I_Q the maximum quantization error is $q_{\max} = \Delta/2$
- Signal values outside I_Q are mapped to the largest or smallest signal stage, respectively $\Rightarrow$ the quantization error is unbounded outside of I_Q

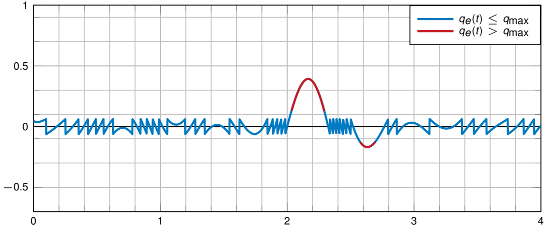
What if the signal values are not uniformly distributed?

- Linear quantization is typically suboptimal
- Non-linear quantization is used, for example, in the digitization of speech or music



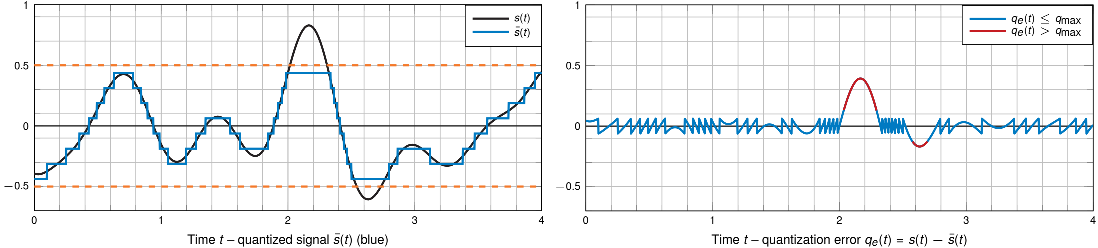
Example: Linear quantization within the interval $I = [-0.5; 0.5]$ mit $N = 3$ bit:





Code word	000	001	010	011	100	101	110	111
Signal stage	$-7/16$	$-5/16$	$-3/16$	$-1/16$	$1/16$	$3/16$	$5/16$	$7/16$

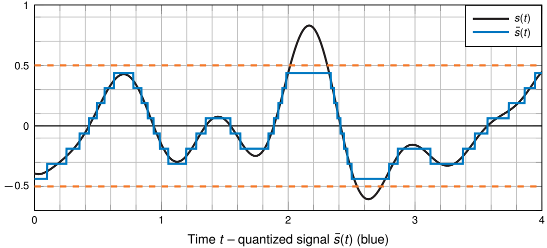
Example: Linear quantization within the interval $I = [-0.5; 0.5]$ mit $N = 3$ bit:

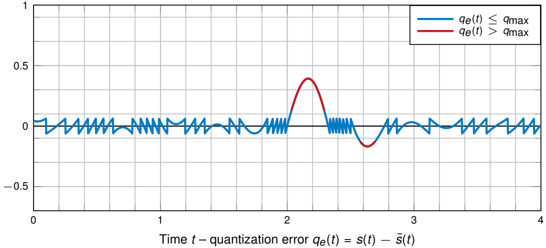


Code word	000	001	010	011	100	101	110	111
Signal stage	$-7/16$	$-5/16$	$-3/16$	$-1/16$	$1/16$	$3/16$	$5/16$	$7/16$

Question: Why is the highest signal stage at $7/16$ and not at $1/2$?

Example: Linear quantization within the interval $I = [-0.5; 0.5]$ mit $N = 3$ bit:





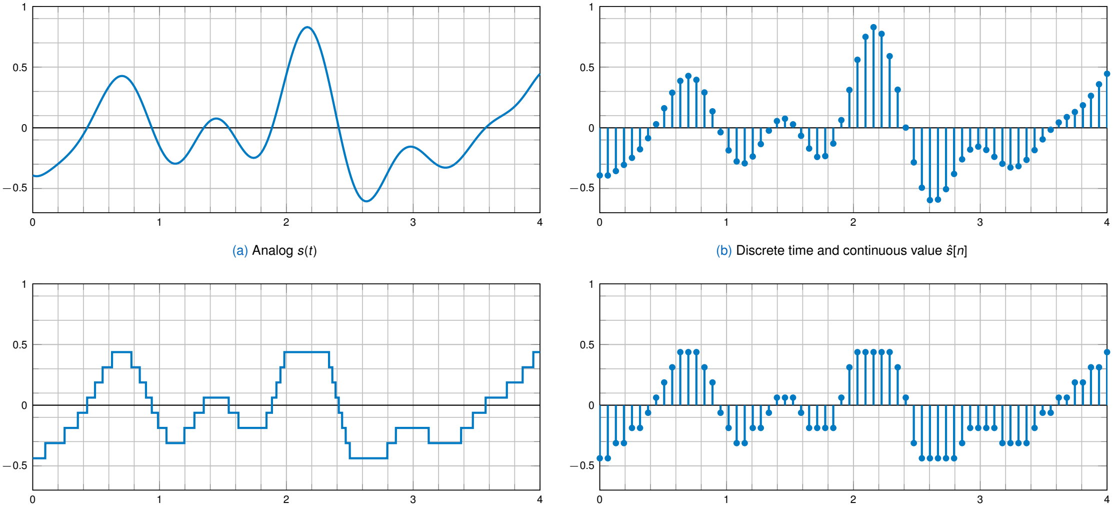
Code word	`000`	`001`	`010`	`011`	`100`	`101`	`110`	`111`
Signal stage	$-7/16$	$-5/16$	$-3/16$	$-1/16$	$1/16$	$3/16$	$5/16$	$7/16$

Question: Why is the highest signal stage at $7/16$ and not at $1/2$?

Remarks:

- The assignment of code words to signal levels is in principle arbitrary
- However, one often chooses a code which reduces the effect of single bit errors
  (e. g. Gray code: Adjacent codewords differ only in one binary digit each, i. e., the Hamming distance is 1).

Signal types (overview)



(a) Analog $s(t)$

(b) Discrete time and continuous value $\hat{s}[n]$

(c) Time continuous and discrete value $\bar{s}(t)$

(d) Digital $s[n]$

Chapter 1: Physical layer

Transmission channel

From the last subchapter we should know:

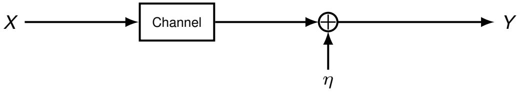
- What are the differences between analog, discrete-time, discrete-value and digital signals?
- How must a signal be sampled so that no information is lost?
- Under what conditions can a naturally occurring signal be reconstructed from sampled and quantized values without loss?
- How should the samples be quantized if within the quantization interval each signal level is equally likely?

In this subchapter we clarify the following questions:

- What influence does the transmission channel have on a signal?
- What is the theoretically maximum achievable transmission rate?

Channel effects

Model of a (linear, time-invariant) channel with one input and one output:



Our model considers:

- Attenuation (Signal amplitude of the useful signal at the output is lower than at the input)
- Low pass filter (higher frequencies are attenuated more than low ones)
- Delay (the transfer takes some time)
- Noise in shape of additive white Gaussian noise (AWGN)[1]

Among others, we do not consider the following effects:

- Interference by other transmissions
- Reflections of our own signal
- Distortions due to non linear filtering effects, among others in dependency of the signal amplitude
- Time variant effects, e. g. objects or people may have an influence on wireless transmissions

[1] AWGN is a simplifying model conception of noise processes. In reality, there is no AWGN.

Example:

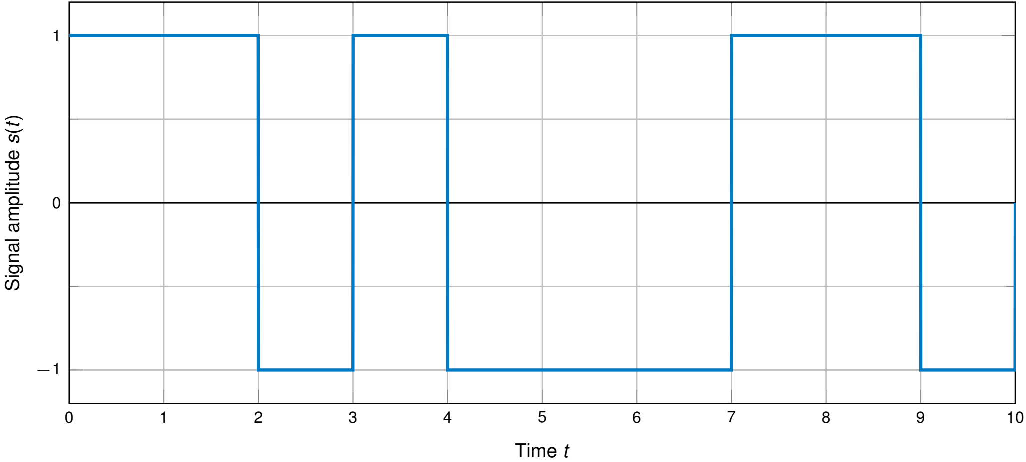


Figure 4: Idealized transmitted signal

Example:

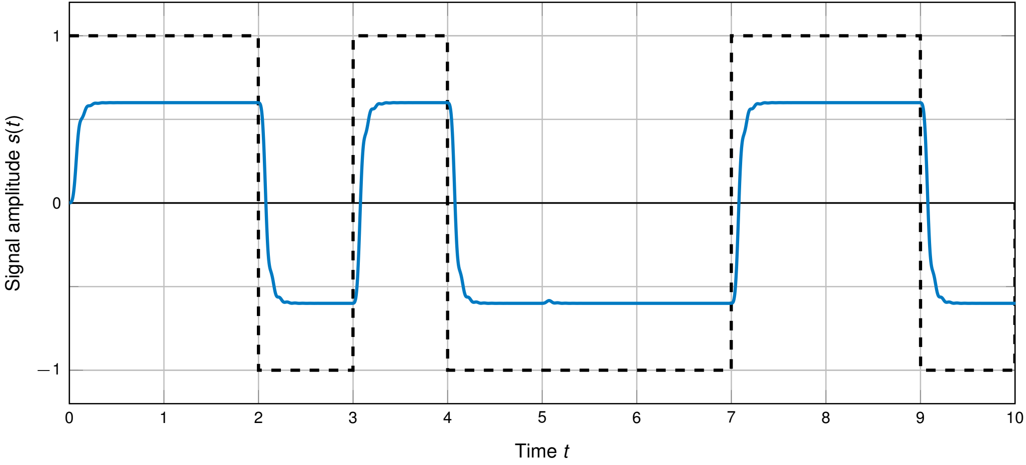


Figure 4: Transmit signal after attenuation and low-pass influences by the channel

Example:

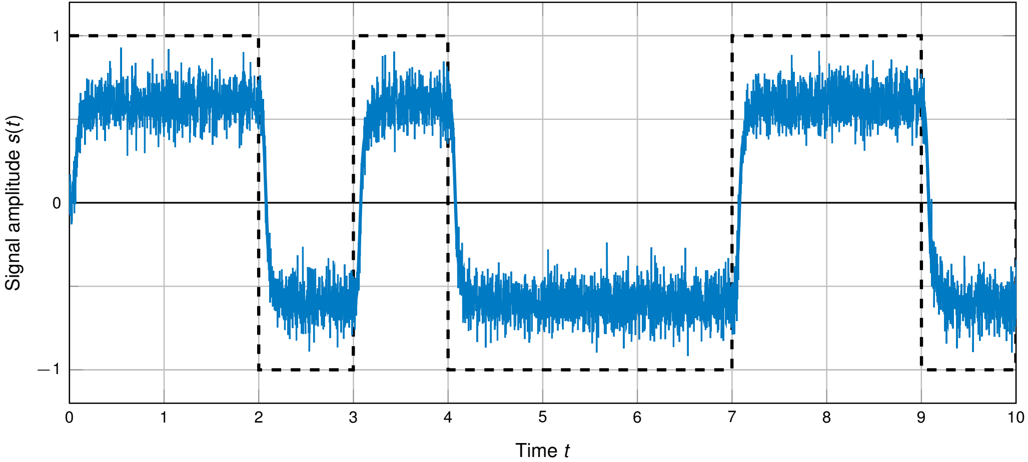


Figure 4: Transmit signal after attenuation and low-pass influences through the channel as well as with AWGN

Channel capacity

We have already seen that

- a channel has similar effects like a low pass filter and
- additional noise distorts the transmission.

Because of the low-pass characteristic of channels, one can speak of a channel bandwidth B:

- Low frequencies pass unhindered (low pass)
- High frequencies are attenuated
- Above a certain frequency, the attenuation is so strong that the relevant signal components can be neglected

Simplified we assume a sharp upper bound for B:

- Frequencies $|f| < B$ pass
- Frequencies $|f| \geq B$ are filtered

What is the achievable data rate on a channel with bandwidth B?

For this we need a connection between

- the channel bandwidth B,
- the number M of distinguishable signal stages, and
- the relation between the power of the useful signal and the noise.

Noise-free, binary channel

We remember the sampling theorem:

A signal bandlimited to B must be sampled at least at the frequency $2B$ in order to reconstruct the signal without loss, i.e. so that no information is lost.

Viewed the other way around:

- We obtain up to $2B$ distinguishable[2] symbols from a signal limited to bandwidth B.
- If you scan more frequently, you do not gain any new information.
- This leads to a new interpretation of the frequency $f = 2B$, which is also called Nyquist rate.

Definition: Nyquist rate

Let B be the cutoff frequency of a bandlimited channel. Then the Nyquist rate $f_N = 2B$ is

- a lower bound for the sampling frequency that allows a complete reconstruction of the signal and
- an upper bound for the number of symbols per time interval that are distinguishable after transmission.

[2] Sufficiently sensitive measuring systems provided

Channel capacity

Noise-free, M-ary channel

Assuming that not only two but $M = 2^N$ distinguishable symbols can be transferred. How does the achievable data rate change?

We remember quantization and entropy:

- With a word width of N bit, $M = 2^N$ discrete signal stages can be represented.
- If a source emits all characters (symbols) with the same probability, the entropy (and thus the average information) of the source is maximal.

Consequently, for the transmission rate over a channel of bandwidth B, we obtain the maximum rate $2B\log_2(M)$ bit.

Hartleys theoreom

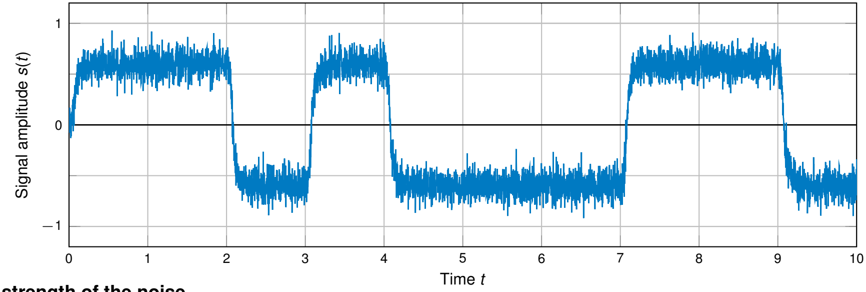
On a channel of bandwidth B with M distinguishable signal stages, the channel capacity bounded above by

$$C_H = 2B\log_2(M)\ \text{bit}.$$

Interesting: If we could distinguish any number of signal stages from each other, the achievable data rate would be unlimited! Where is the problem?

Noise

- Noise makes it hard to tell signal levels apart
- The finer the signal levels are selected, the more difficult this becomes



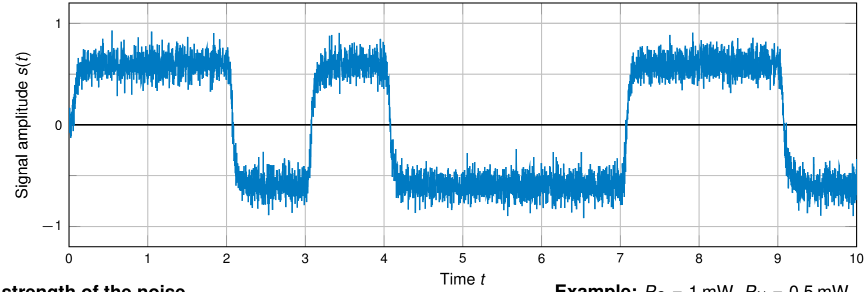
Measure of the strength of the noise

$$\text{SNR} = \frac{\text{Signal power}}{\text{Noise power}} = \frac{P_S}{P_N}$$

The Signal to Noise Ratio (SNR) is given as dB: $\text{SNR}\,\text{dB} = 10 \cdot \log_{10}(\text{SNR})$

Noise

- Noise makes it hard to tell signal levels apart
- The finer the signal levels are selected, the more difficult this becomes



Measure of the strength of the noise

$$\mathrm{SNR} = \frac{\text{Signal power}}{\text{Noise power}} = \frac{P_S}{P_N}$$

Example: $P_S = 1\,\mathrm{mW}$, $P_N = 0{,}5\,\mathrm{mW}$

$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{1}{0{,}5}\right)\,\mathrm{dB} \approx 3{,}0\,\mathrm{dB}$$

The Signal to Noise Ratio (SNR) is given as dB: $\mathrm{SNR\,dB} = 10 \cdot \log_{10}(\mathrm{SNR})$

Theorem of Shannon and Hartley

On a channel of bandwidth B with additive white noise with noise power P_N and signal power P_S, the upper bound for the achievable data rate is

$$C_S = B \log_2 \left(1 + \frac{P_S}{P_N}\right) \text{ bit.}$$

Derivation of the theorem: see Shannon's 1949 publication Communication in the Presence of Noise [1].

Theorem of Shannon and Hartley

On a channel of bandwidth B with additive white noise with noise power P_N and signal power P_S, the upper bound for the achievable data rate is

$$C_S = B \log_2\left(1 + \frac{P_S}{P_N}\right) \text{ bit.}$$

Derivation of the theorem: see Shannon's 1949 publication Communication in the Presence of Noise [1].

Comparison with Hartley's law:

$$C_H = 2B\log_2(M) = 2B\log_2\left(\frac{b-a}{\Delta}\right) \text{ bit.}$$

Theorem of Shannon and Hartley

On a channel of bandwidth B with additive white noise with noise power P_N and signal power P_S, the upper bound for the achievable data rate is

$$C_S = B \log_2 \left(1 + \frac{P_S}{P_N}\right) \text{ bit.}$$

Derivation of the theorem: see Shannon's 1949 publication Communication in the Presence of Noise [1].

Comparison with Hartley's law:

$$C_H = 2B \log_2(M) = 2B \log_2 \left(\frac{b-a}{\Delta}\right) \text{ bit.}$$

- The interval limits a,b here refer to the unquantized signal
- With $\alpha = a + \Delta/2$ and $\beta = b - \Delta/2$ as minimum and maximum quantized signal amplitude, respectively, we obtain

$$C_H = 2B \log_2 \left(\frac{\beta - \alpha + \Delta}{\Delta}\right) = B \log_2 \left(\left(1 + \frac{\beta - \alpha}{\Delta}\right)^2\right) = B \log_2 \left(1 + \frac{(\beta - \alpha)^2}{\Delta^2} + 2\frac{\beta - \alpha}{\Delta}\right). \tag{1}$$

Theorem of Shannon and Hartley

On a channel of bandwidth B with additive white noise with noise power P_N and signal power P_S, the upper bound for the achievable data rate is

$$C_S = B \log_2 \left(1 + \frac{P_S}{P_N} \right) \text{ bit.}$$

Derivation of the theorem: see Shannon's 1949 publication Communication in the Presence of Noise [1].

Comparison with Hartley's law:

$$C_H = 2B \log_2(M) = 2B \log_2 \left(\frac{b - a}{\Delta} \right) \text{ bit.}$$

- The interval limits a,b here refer to the unquantized signal
- With $\alpha = a + \Delta/2$ and $\beta = b - \Delta/2$ as minimum and maximum quantized signal amplitude, respectively, we obtain

$$C_H = 2B \log_2 \left(\frac{\beta - \alpha + \Delta}{\Delta} \right) = B \log_2 \left(\left(1 + \frac{\beta - \alpha}{\Delta} \right)^2 \right) = B \log_2 \left(1 + \frac{(\beta - \alpha)^2}{\Delta^2} + 2 \frac{\beta - \alpha}{\Delta} \right). \tag{1}$$

Just as with C_S, we get a logarithm of 1 + SNR, where this time the SNR is a quantization noise:

- **C_S considers only additive noise of the channel but no quantization erros.**
- **C_H considers only signal stages and thus noise due to quantization but no channel effects.**
- The missing mixed term in (1) compared to C_S is related to the assumption of independence between signal and noise ($\mathrm{E}[x\eta] = \mathrm{E}[x]\mathrm{E}[\eta]$). The quantization error is of course not independent of the input signal – for this reason (1) cannot be put into the same form as C_S without approximation.

Summary

The channel capacity C is limited by two factors:

- **The number *M* of distinguishable symbols**
  Even a noise-free channel is of no use if we can only use two symbols.
- **Signal-to-Noise Ratio (SNR)**
  If the signal-to-noise ratio SNR is too low, the distance Δ between the signal stages may have to be increased and thus the number of distinguishable symbols reduced to ensure reliable discrimination.

The channel capacity C is thus limited by the following upper bound:

$$C < \min\{C_H, C_S\} = \min\left\{2B\log_2(M), B\log_2\left(1+\mathrm{SNR}\right)\right\} \text{ bit.}$$

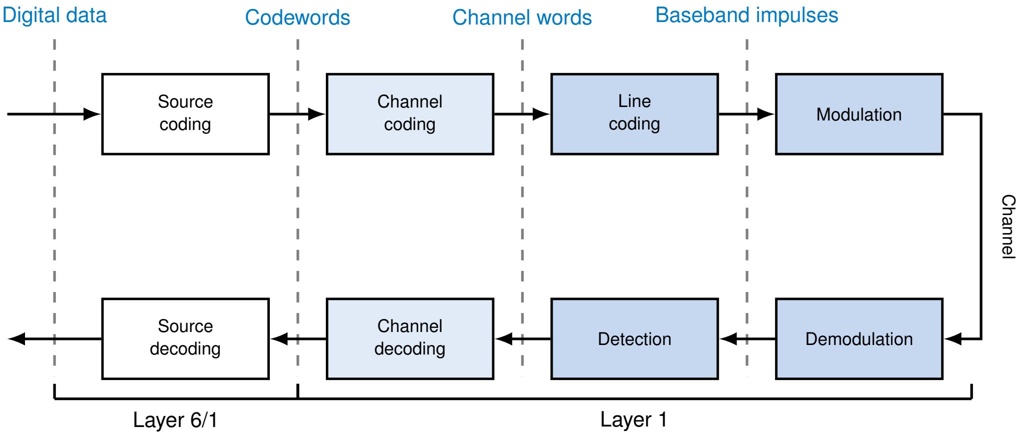
Remarks:

- This is just a model – with highly simplifying assumptions.
- How to construct a channel code with just the right amount of redundancy so that C is maximized is an open problem in information theory. ($\leftarrow$ challenge!)
- We are talking here about data rates in the information-theoretical sense, i. e., the data to be transmitted is available without redundancy. This is never guaranteed in real systems
  - Payloads are not necessarily (and never optimally) compressed before transmission
  - In addition to the payloads, control information (headers) is required ($\rightarrow$ more on that later).

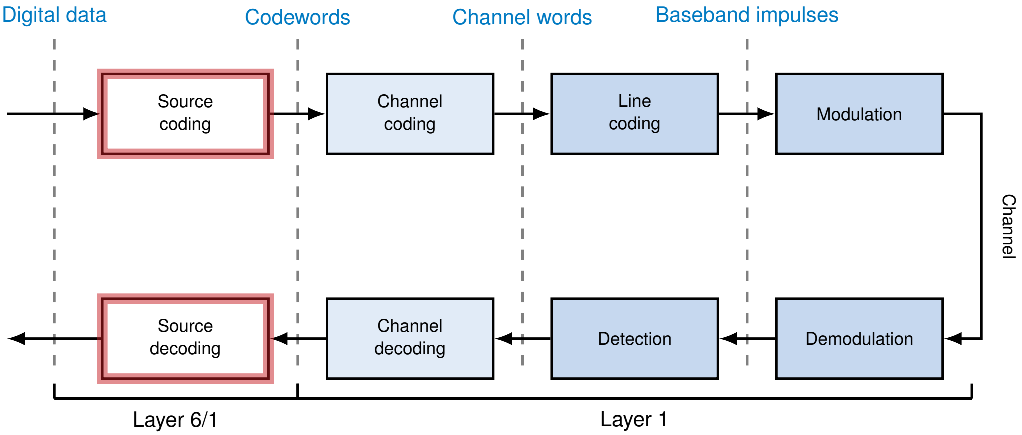
  $\Rightarrow$ The net data rate that can actually be achieved is below the information-theoretic barrier.

Chapter 1: Physical layer

Message transmission



Source coding [4]



Source coding
The goal of source coding is to remove (unstructured) redundancy from the data to be transmitted by mapping bit sequences to code words. This corresponds to lossless data compression.

Source encoding can occur in different layers of the ISO/OSI model:

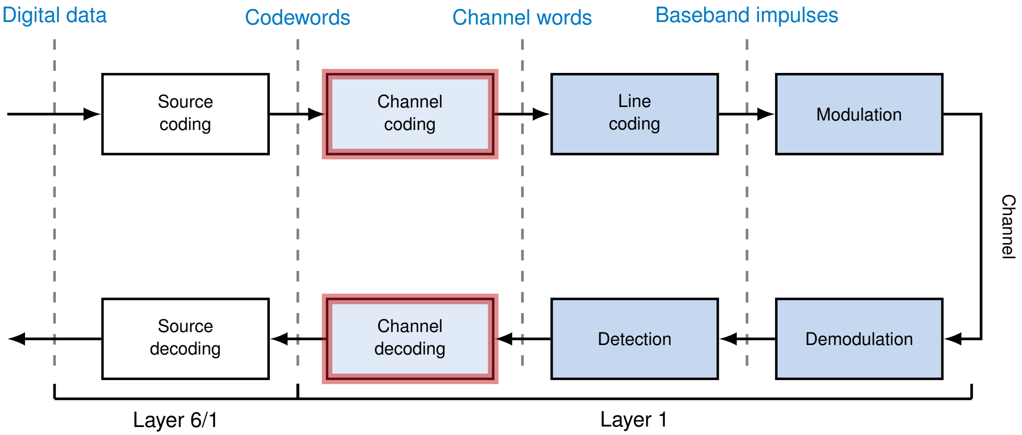
- Data compression can take place on the presentation layer (layer 6)
- Data may already be in compressed form (lossless compressed file formats, e. g. ZIP, PNG).
- In mobile communications (digital voice transmission), the source coding may happen even at layer 1.
- In local area networks such as Ethernet and WLAN there is commonly no explicit source coding

Examples:

- Huffman code
- Lempel-Ziv / Lempel-Ziv-Welch (LZW)
- Run-Length Enconding (RLE)

→ In Chapter 5 we will go into the Huffman code, which is also covered by the lecture *Theory of Computation and Information Theory*.

Channel coding [4]



No feasible transmission channel is perfect. One measure of this is the bit error probability p_e:

- Characteristic for Ethernet over copper cable: $p_e \approx 10^{-8}$
- Characteristic for WLAN: $p_e \approx 10^{-6}$ or more
- Characteristic for unsecured long range radio transmission: $p_e \approx 10^{-4}$ or more

Mind game:

- Assume an unsecured radio transmission with bit error probability $p_e = 10^{-4}$, and let bit errors be independently and uniformly distributed
- Assume a packet length of $L = 1500\,\mathrm{B} = 12\,000\,\mathrm{bit}$
- $\Pr[$„no bit error in the packet“$] = (1 - 10^{-4})^{12000} \approx 30\,\%$

$\Rightarrow$ 70 % of the transmitted packets would contain at least one bit error.

[3] Note that error detection and error correction are different goals. Not every error that is detected can also be corrected.

Channel coding [4]

No feasible transmission channel is perfect. One measure of this is the bit error probability p_e:

- Characteristic for Ethernet over copper cable: $p_e \approx 10^{-8}$
- Characteristic for WLAN: $p_e \approx 10^{-6}$ or more
- Characteristic for unsecured long range radio transmission: $p_e \approx 10^{-4}$ or more

Mind game:

- Assume an unsecured radio transmission with bit error probability $p_e = 10^{-4}$, and let bit errors be independently and uniformly distributed
- Assume a packet length of $L = 1500\,\text{B} = 12\,000\,\text{bit}$
- Pr [„no bit error in the packet“] $= (1 - 10^{-4})^{12000} \approx 30\,\%$

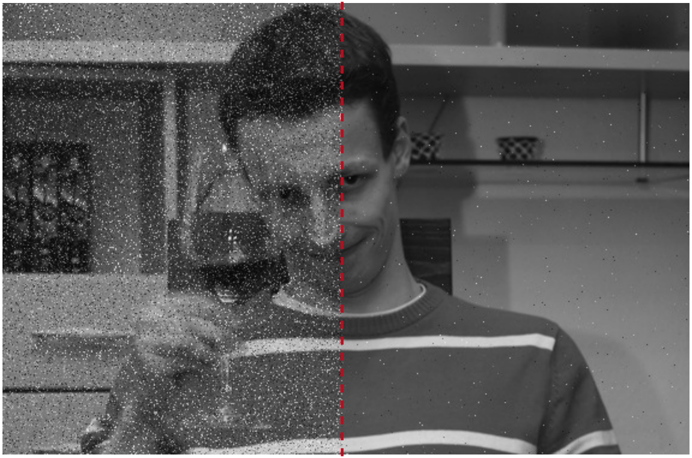
$\Rightarrow$ 70 % of the transmitted packets would contain at least one bit error.

Channel coding
The aim of channel coding is to add structured redundancy to the data to be transmitted so that the largest possible number of bit errors can be detected and corrected.[3]

[3] Note that error detection and error correction are different goals. Not every error that is detected can also be corrected.

Channel coding [4]

Example: Uncompressed image (bitmap) transmitted over an imperfect channel



without channel coding | with channel coding

Channel coding [4]

Minor transmission errors are tolerable in analog systems:

- Noise or crackling on a telephone connection
- Snow (noise) in analog TV
- FM radio

In digital systems, transmission errors can have serious consequences:

- Transmission of compressed or encrypted data (possible error propagation during decoding)
- Error-free transmission may be required, e. g. a downloaded application may be unusable even with single bit error

Additional protocols and mechanisms are therefore needed

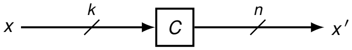
- to at least detect transmission errors that occur despite channel coding and
- to repeat a transmission if necessary.

$\Rightarrow$ Interaction of checksums and acknowledgment procedures, typically at the layers 2, 4, and 7.

Block codes divide a data stream

- in blocks of length k and
- map these blocks to code words of length $n > k$ while
- adding $n - k$ bit for error detection and correction.

The ratio $R = \frac{k}{n}$ is called **code rate**.

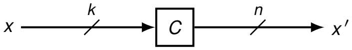


Channel coding [4]

Block codes divide a data stream

- in blocks of length k and
- map these blocks to code words of length $n > k$ while
- adding $n - k$ bit for error detection and correction.

The ratio $R = \frac{k}{n}$ is called code rate.



Example: Repetition code

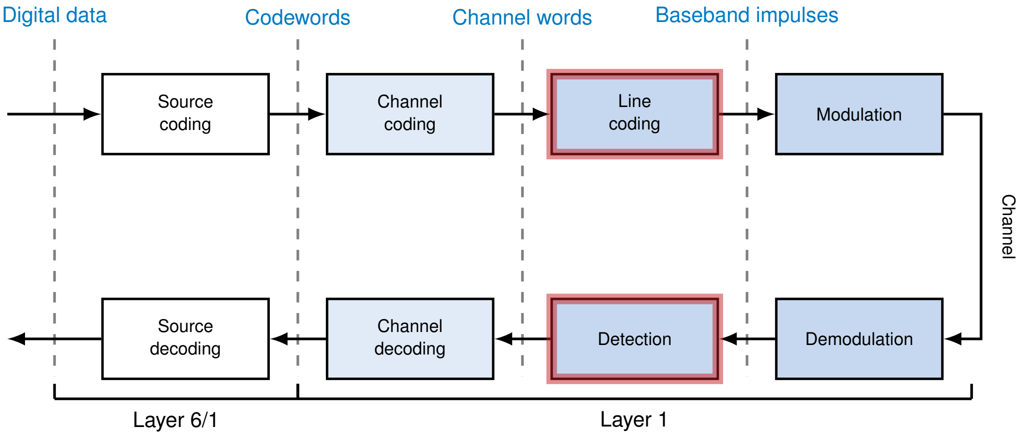
- $k = 1$, $n = 3$, mapping: $0 \mapsto 000$, $1 \mapsto 111$
- Decoding fails if 2 bit or more per block are flipped:

$$\Pr[\text{„decoding failure“}] = \binom{3}{2} p_e^2 (1 - p_e) + \binom{3}{3} p_e^3 \approx \Big|_{p_e = 10^{-4}} 3 \cdot 10^{-8}$$

- New problem:
  - The number of bits to be sent is tripled
  - In the error-free case, the achievable data rate would thus decrease to $1/3$

$\Rightarrow$ Cost / benefit ratio between error probability and redundancy depends on the current bit error rate

Line coding


Digital data
Codewords
Channel words
Baseband impulses
Source coding
Channel coding
Line coding
Modulation
Channel
Source decoding
Channel decoding
Detection
Demodulation
Layer 6/1
Layer 1

Definition – Line codes

Line codes (not to be confused with channel codes) define the sequence of a certain kind of basic pulses representing bits or groups of bits. Such a sequence of basic impulses is called transmitting impulse.

In the context of line codes, we understand a symbol to be a physically. measurable change in the time signal.

Important properties of line codes:

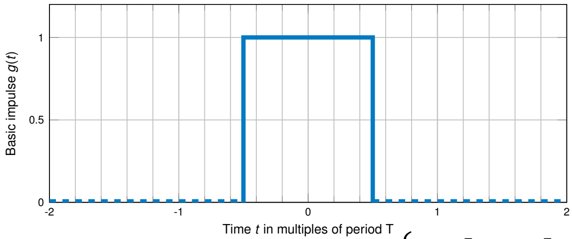
- Number of signal levels (binary, ternary, . . .)
- Number of bits encoded per symbol
- Symbol rate (called Baud rate), unit bd

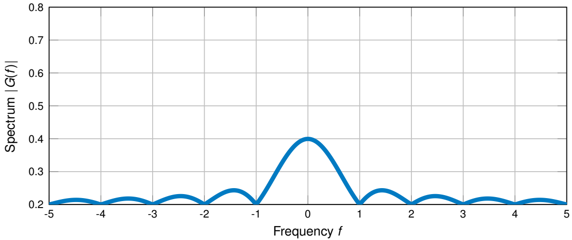
Optional properties of line codes:

- Clock recovery
- DC freedom
- Additional control characters (e. g. 4B5B code $\rightarrow$ more on that later)

Depending on the type of basic pulses used and their sequence, line codes have an influence on the required channel bandwidth. As a rule of thumb: the more abrupt signal changes there are, the wider the spectrum required. (see examples)

Basic impulses: rectangular impulse





$$g(t) = \begin{cases} 1 & -\frac{T}{2} \leq t < \frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \quad \circ\!\!-\!\!\bullet \quad G(f)^{1} = \frac{1}{\sqrt{2\pi}} \frac{\sin(\pi f)}{\pi f}$$

Advantages

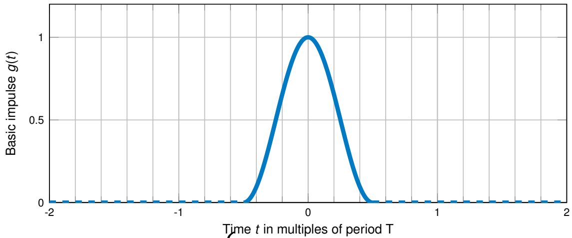
- Most simple representation of data in the time domain
- Basis for various transmitting impulses ($\rightarrow$ more on that later)

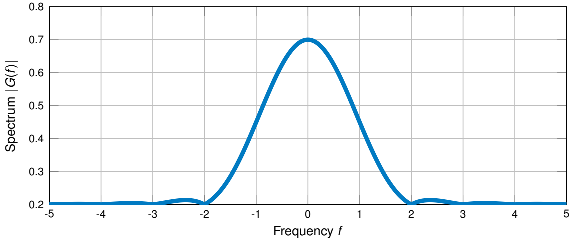
Disadvantages

- Abrupt signal changes practically difficult to implement at high frequencies
- Slowly decaying spectrum $\Rightarrow$ high frequency components

[1] The spectrum $G(f)$ is determined using the Fourier transformation: $G(f) = \int_{-\infty}^{\infty} g(t)\,(\cos(2\pi f t) - j\sin(2\pi f t))\,\mathrm{d}t$ where j denotes the imaginary unit.

Basic impulses: $\cos^2$ impulse





$$g(t) = \begin{cases} (\cos(2\pi t))^2 & -\frac{T}{2} \leq t < \frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \quad ○\!\!-\!\!\!-\!\!\bullet \quad G(f) = \frac{1}{4\pi}\left(\frac{\sin(\pi(f-1))}{f-1} + \frac{2\sin(\pi f)}{f} + \frac{\sin(\pi(f+1))}{f+1}\right)$$

Advantages

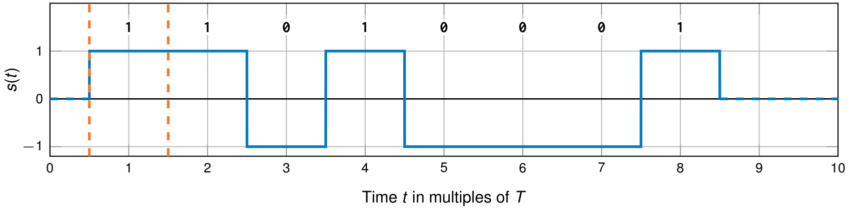
- Fast decaying spectrum since few high frequency components
- Therefore lower influence of low passes

Disadvantages

- The maximum signal amplitude $g(t) = 1$ is reached only in the middle of the impulse
- This makes sampling more difficult if the transmitter and receiver are not synchronized

Line coding

Line codes: Non-Return-To-Zero (NRZ)



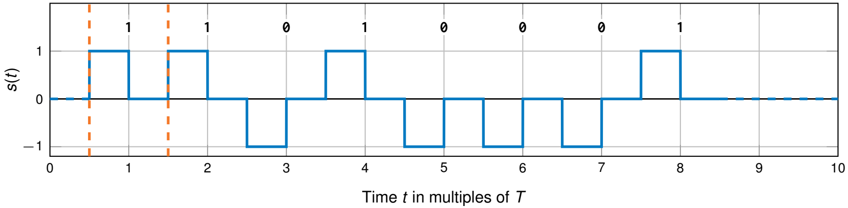
Coding rule:

- Transmit pulse $g(t) = \mathrm{rect}\,(t)$ with period T
- Possible assignment of weights $d_n = \begin{cases} 1 & b_n = 1 \\ -1 & b_n = 0 \end{cases}$
- Transmitted signal is defined as $s(t) = \sum_{n=0}^{\infty} d_n g(t - nT)$

Properties:

- Binary code (only two signal levels)
- Efficiency 1 Symbol/bit
- No clock recovery (long sequences of same bit)
- Not free of DC
- Slowly decaying frequency components

Line codes: Return-To-Zero (RZ)



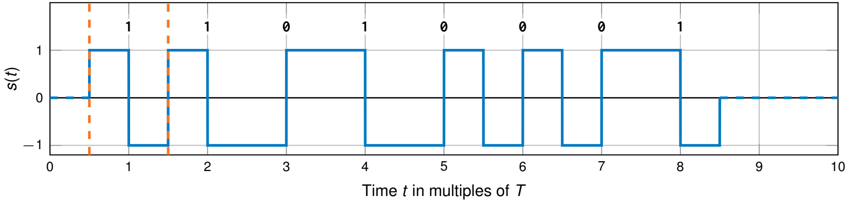
Kodiervorschrift:

- Transmit pulse $g(t) = \text{rect}\left(2t + \frac{T}{2}\right)$ with period T
- Possible assignment of weights $d_n = \begin{cases} 1 & b_n = 1 \\ -1 & b_n = 0 \end{cases}$
- Transmitted signal is defined as $s(t) = \sum_{n=1}^{\infty} d_n g(t - nT)$

Properties:

- Binary code (only two signal levels)
- Efficiency 2 Symbols/bit
- Clock recovery through forced level changes simple
- Not free of DC
- Slower decay of high frequency components than NRZ

Line code: Manchester-Code



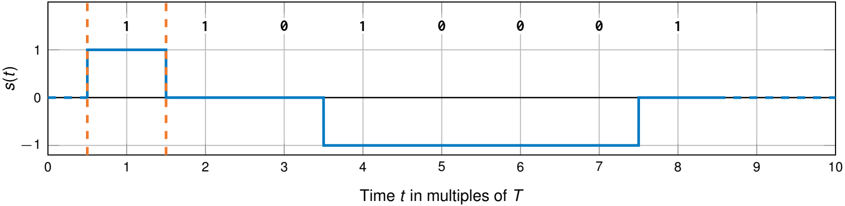
Coding rule:

- Transmit pulse $g(t) = \text{rect}\left(2t + \frac{T}{2}\right) - \text{rect}\left(2t - \frac{T}{2}\right)$ with period T
- Possible assignment of weights $d_n = \begin{cases} 1 & b_n = 1 \\ -1 & b_n = 1 \end{cases}$
- Transmitted signal is defined as $s(t) = \sum_{n=1}^{\infty} d_n \cdot g(t - nT)$

Properties:

- Binary code (only two signal levels)
- Efficiency 2 Symbols/bit
- Clock recovery through forced level changes simple
- DC free since each transmit pulse is already DC free
- Even slower decay of high frequency parts than RZ

Line code: Multi-Level-Transmit 3 (MLT3)



Coding rule:

- Transmit pulse $g(t) = \text{rect}(t)$ (rectangular pulse) with period T
- Weights $d_n = \sin\left(\frac{\pi}{2}\sum_{k=1}^{n} b_k\right)$
  (→ dependent on the number of previously occured 1-bits)
- Transmit signal defined as $s(t) = \sum_{n=1}^{\infty} d_n g(t - nT)$

Properties:

- Ternary code (three signal levels)
- Efficiency 1 bit/Symbol
- No clock recovery (long sequences of same 0-bit leads to no change of the signal level)
- Not DC free
- Fast decay of high frequency components since the fundamental period is reduced by the periodic signal waveform

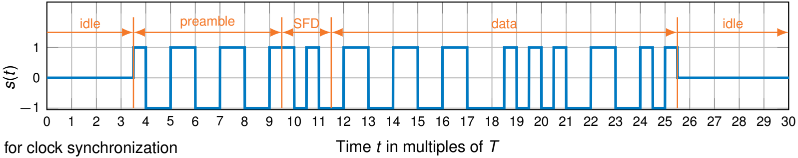
Open questions: How can a receiver detect if

- symbols represent data at all (medium could be “idle”) and
- how can the beginning or end of a message be detected?

Option 1: Violation of coding rule

- If the medium is idle, invalid baseband pulses can be transmitted
- A fixed number of alternating bits can be sent before the start of a message (preamble)
- Start of the message is indicated by a second sequence (Start Frame Delimiter (SFD)).
- This works with NRZ, RZ and Manchester Code (e. g. zero level), but not with MLT3 (zero level here means a sequence of 0-bits).

Example: Manchester-Code with preamble



- Preamble allows for clock synchronization
- Start Frame Delimiter (SFD) at the end of the preamble signals the beginning of the message
- Coding rule violation (zero signal level) indicates and idle medium
- Used by IEEE 802.3a/i (10 Mbit/s Ethernet over coaxial and twisted pair cables → more on that later)

Option 2: Control characters

- Define a block code that divides channel words into groups of k bit and maps to $n > k$ bit.
- This block code is not for error correction (task of channel coding), but only for providing control characters.
- The mapping can be selected in such a way that, when transmitting valid channel words,
  - clock recovery and
  - DC freeomd

  become possible even with line codes such ase NRZ, RZ, and MLT3.
- Invalid code words that are neither data words nor control characters can be used for error detection

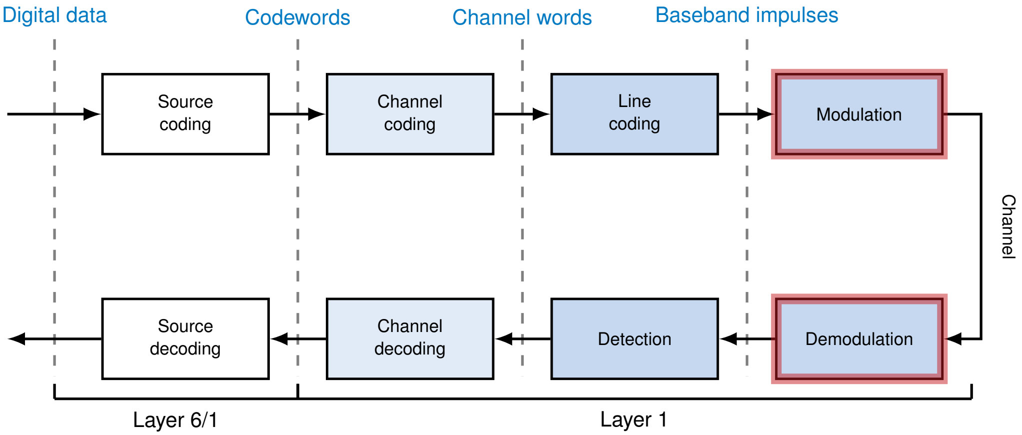
Example 1: 4B5B code

- $k = 4$ bit channel words are mapped to $n = 5$ bit code words
- The assignment between channel words and code words is chosen so that at least one signal change occurs in each block of 5 bit (clock recovery for NRZ and MLT3).
- The additional code words are used as control characters (start/stop, idle, . . .)
- Used in combination with MLT3 by IEEE 802.3u (100 Mbit/s FastEthernet over twisted pair cables)

Beispiel 2: 8B10B code

- $k = 8$ bit channel words are mapped to $n = 10$ bit code words
- Assignment similar to 4B5B, but here DC freedom ican also be guaranteed over time
- Used by PCIe, Serial-ATA, USB . . .

Modulation [6]



So far we have considered only baseband signals:

- Time-shifted transmission pulses are weighted.
- Temporally limited transmission pulses (we have got to know only such) possess an infinitely extended spectrum.
- Provided that the transmission channel is exclusively available for baseband transmission, this is not a problem at first.

What if the channel is used by several transmissions simultaneously

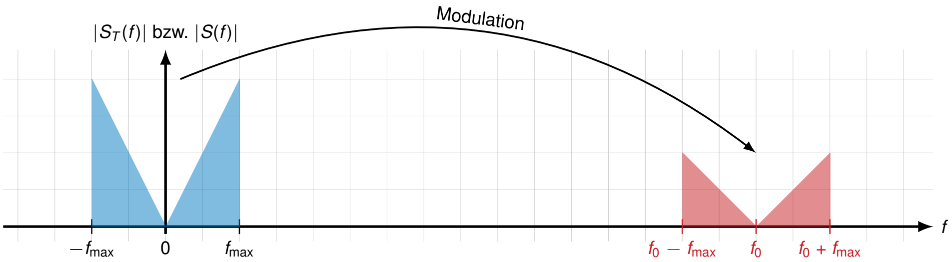
- The baseband signal (or its basic pulses) is lowpass filtered, which corresponds to a limitation of the spectrum (and thus a slight distortion of the time signal).
- Subsequently, the filtered baseband signal can be modulated to a carrier signal.
- This corresponds to a shift of the spectrum (multiplication in the time domain corresponds to a shift in the frequency domain).
- If several transmissions share one channel in this way, we speak of Frequency Division Multiplex (FDM).

Principle sequence of digital modulation processes

- The transmit pulses $g(t)$ are limited to a maximum frequency $f_{\max}$ by means of low-pass filtering. We refer to the pulses filtered in this way as $g_T(t)$.
- The also band-limited transmit signal $s_T(t)$ is modulated on a carrier signal of frequency f_0:

$$s(t) = s_T(t) \cdot \cos(2\pi f_0 t) = \left(\sum_{n=1}^{\infty} d_n \cdot g_T(t - nT) \right) \cos(2\pi f_0 t).$$

Schematic sequence in the frequency domain:



Spectrum of the transmit signal $s_T(t)$ in the baseband

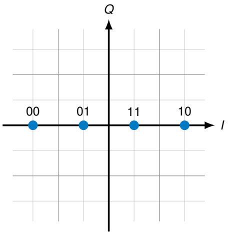
Spectrum of the bandpass signal after modulation $s(t)$

4-ASK (Amplitude Shift Keying)

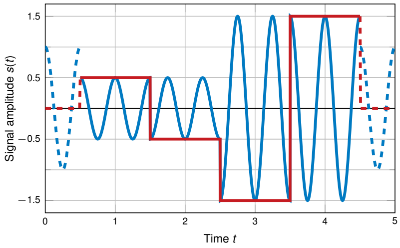
- A distinction is made between 4 signal levels $\Rightarrow$ 2 bit/Symbol
- Only the amplitude of the carrier signal is modulated

Example: Possible weights $S = \left\{-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}\right\}$

- Two bits of the data stream are mapped to a symbol $d \in S$ each, e.g. $00 \mapsto -\frac{3}{2}$, $01 \mapsto -\frac{1}{2}, \dots$
- The symbol sequence d_n changes the amplitude of a basic pulse (e.g. a low pass filtered square pulse)
- The resulting baseband signal is multiplied by a carrier signal (modulation)



(a) Constellation diagram

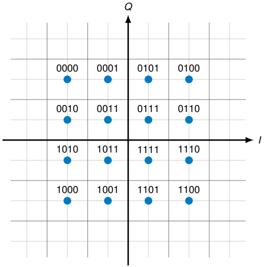


(b) Transmit signal $s(t)$ (blue), baseband signal $s_T(t)$ (red).
Simplification: s_T was not low-pass filtered here!

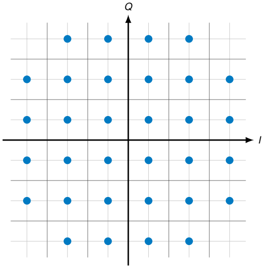
Quadrature-Amplitude-Modulation (QAM)

- You can mix cosine and sinusoidal carrier signals
- Separation possible by orthogonality of sine and cosine
- The cosine is called inphase part, the sine is called quadrature part
- The data rate can be doubled in this way

$$s(t) = \left(\sum_{n=1}^{\infty} d_{In} \cdot g_T(t - nT) \right) \cos(2\pi f_0 t) - \left(\sum_{n=1}^{\infty} d_{Qn} \cdot g_T(t - nT) \right) \sin(2\pi f_0 t)$$



(c) 16-QAM



(d) 32-QAM

Modulation [6]

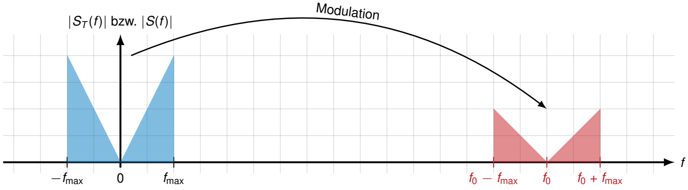
- QAM simply doubles the data rate?
- Have we disproved Shannon?

Modulation [6]

- QAM simply doubles the data rate?
- Have we disproved Shannon?

Of course not: [7] Due to the frequency shift, the bandpass signal occupies double the bandwidth compared to the baseband signal. This shifts the negative frequency components from the baseband into the positive range, forming an

- upper side band, which represents the non-negative frequency components, and a
- lower side band, which represents the non-positive frequency components of the baseband signal.



Spectrum of the transmit signal $s_T(t)$ in the baseband

Spectrum of the bandpass signal after modulation $s(t)$

- Modulation thus doubled the required bandwidth.
- This "lost degree of freedom" can be compensated again by mixing sine and cosine carriers.

The upper bound for the achievable data rate is therefore still valid.

Summary

What we should know:

- What are the differences and goals between source coding, channel coding and line coding?
- How do simple block codes work, e. g. repetition code?
- Why are additional procedures needed for error detection despite all coding procedures?
- How do the line codes introduced in this chapter work?
- What are the respective advantages and disadvantages of the line codes introduced here?
- How could these line codes be extended to more than two or three signal levels?
- What is the principle of modulation?
- How does frequency division multiplex
- How are signal space allocation, modulation method and the achievable data rate related?
- How does Phase Shift Keying (PSK) work and what is a valid signal space mapping for PSK?

Transmission media

We differentiate between

- wireline and
- wireless transmissions

as well as between

- accustic and
- electromagnetic waves.

In the field of digital data transmission, electromagnetic waves are predominantly used. Few exceptions here are

- tone dialing (e. g. „dial-up“ used by internet connections in POTS[4]) and
- submarine wireless communication.

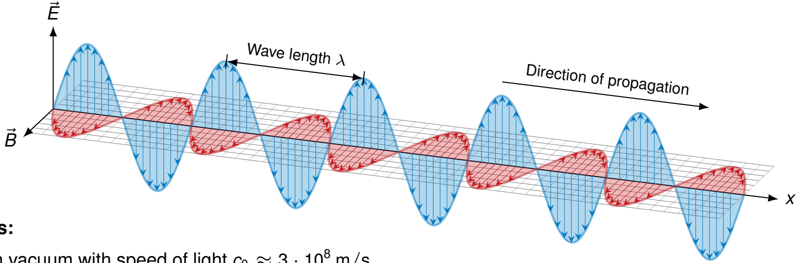
In the following, we provide an overview of

- what EM waves actually are,
- frequencies in the EM spectrum, and
- which types of transmission media are frequently used in wireline networks.

[4] Plain old telephone system

Electromagnetic waves

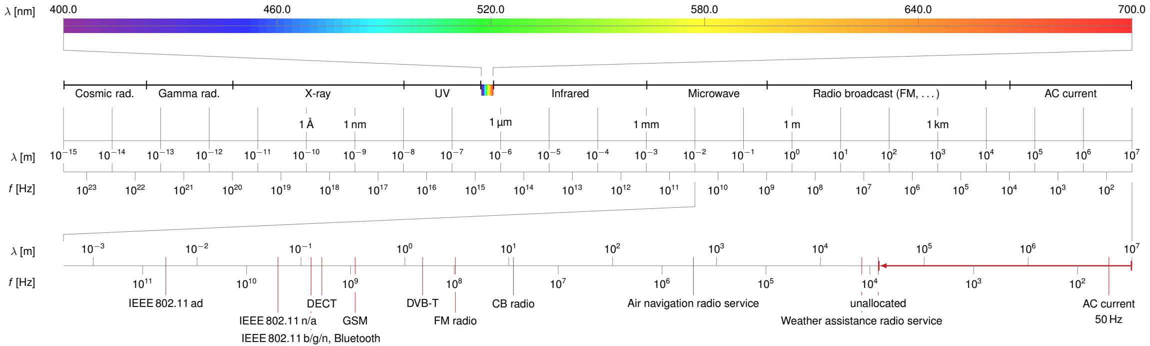
Electromagnetic waves consist of an electric ($\vec{E}$) and magnetic ($\vec{B}$) component, each orthogonal to the other and to the direction of propagation:



Important properties:

- Propagation in vacuum with speed of light $c_0 \approx 3 \cdot 10^8$ m/s
- Unlike sound waves, no medium is required for propagation
- Within a medium (conductor, air), the propagation velocity is $c = \nu c_0$, where $\nu < 1$ is called relative propagation velocity, e. g. $\nu \approx 0.7$ in optical fibers or $\nu \approx 2/3$ in coaxial conductors.
- The wave length λ describes the spatial extension of a wave period in the medium
- The frequency f results from the speed of light and the wave length in the medium to $f = c/\lambda = c_0/\lambda_0$
- At the transition from vacuum into a medium the frequency f remains constant, wave length and propagation velocity change proportionally to each other

The figure below shows a schematic representation of the EM spectrum:

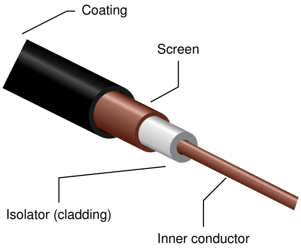


The following are predominantly used for digital data transmission:

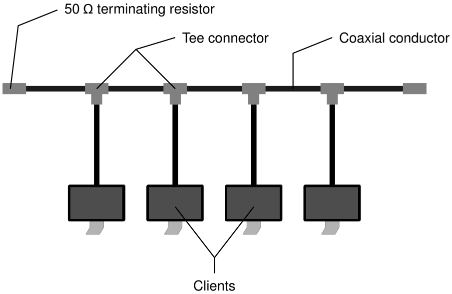
- the frequency band between MHz and $\sim$60 GHz (WLAN / IEEE 802.11 ad),
- the optical spectrum up to $\lambda \approx$ 1 nm, and
- frequencies in the baseband up to a couple of GHz.

Coaxial conductors

- Among others, used for IEEE 802.3a („10Base2 Ethernet“, 10 Mbit/s)
  - Forms a common bus to which all participants are connected
  - Only one participant can send at any time
- Other areas of application:
  - TV cable network
  - High frequency technology (connection to antennas in wireless networks)
  - Twinax cables for 40 und 100 Gbit Ethernet over short distances ($\sim$ 7 m)



(a) Schematic structure [5]



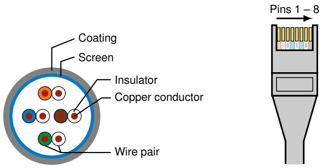
(b) 10Base2 Bus (IEEE 802.3a)

Structure

- 2 or 4 wire pairs consisting of copper strands
- Each wire pair is twisted (thus the name twisted pair)
- Second wire of a pair carries inverse signal level (differential coding)
- Twisting and inverse signal levels reduce crosstalk
- RJ-45 or smaller RJ-11 connectors

Usage

- Local networks (most Ethernet standards for client connections) with RJ-45 connector
- Telephone connection (analog and ISDN) with RJ-11 connector

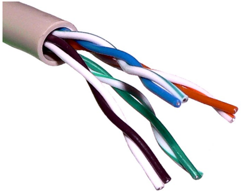


(a) Cable cross section (b) Stecker

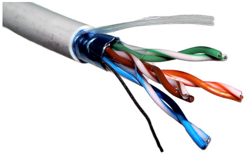
Twisted-Pair-Kabel

Dependent on the type of shielding and screening we differentiate between

- UTP (unshielded twisted pair)
- STP (shielded twisted pair)
- S/UTP (screened / unshielded twisted pair)
- S/STP (screened / shielded twisted pair)



(c) UTP [3]

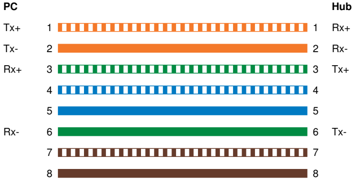


(d) Screened UTP [2]

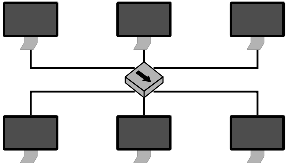
Shielding and screening influences the

- signal quality (e. g. crosstalk between wire pairs) and
- flexibility of the cable (well shielded cables are thicker and stiffer).

Connecting multiple computers via hub (or switch) using straight-through cable at 100BASE-TX

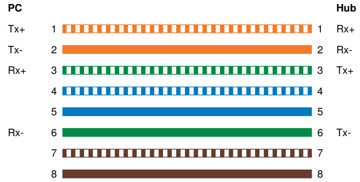


(a) Straight-through

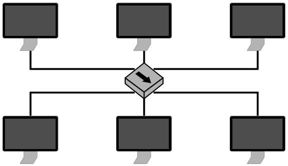


(b) Hub creates a physical bus, half-duplex

Connecting multiple computers via hub (or switch) using straight-through cable at 100BASE-TX

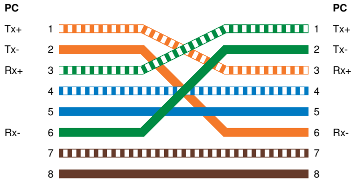


(a) Straight-through



(b) Hub creates a physical bus, half-duplex

- Direct connection of two computers via cross-over cable

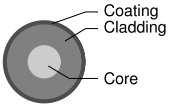


(a) Cross-over

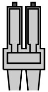
(b) Point-to-point, full-duplex

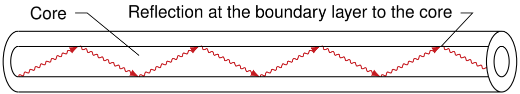
Optical fibers

- Light is transmitted within the fiber core
- Core and cladding each have different optical densities $\rightarrow$ Refractive index ensures approximate total internal reflection
- Single-mode fibers avoid scattering due to very small core diameter $\rightarrow$ low losses, but very sensitive (cable break)
- Multi-mode fibers have a larger core diameter and therefore tend to scatter $\rightarrow$ higher losses, but less sensitive



(a) Cable cross section


(b) LC connector



(c) Side view of a fiber

Advantages over electrical conductors:

- Very high data rates possible
- Long range connections
- No crosstalk
- Galvanic decoupling of transmitter and receiver

Summary

Transmission media

- For digital communication one uses electromagnetic waves
  - in the frequency range up to a couple of GHz and
  - in the optical spectrum.
- As medium one uses either
  - electrical conductors (copper) or
  - optical fibers.
- radio transmissions do not require a medium, since electromagnetic waves (unlike sound waves) propagate in vacuum
- The medium used has an influence on the speed of propagation.

In the next chapter find answers to the questions

- how nodes can access a shared medium (medium access control) and
- messages can be sent to a specific neighboring node (addressing in local networks).

Summary

We should know,

- what the information content of characters as well as the entropy of a message source mean,
- what effects fast level changes in the time domain have on the frequency domain,
- how signals can be sampled, quantized and reconstructed,
- how to determine the maximum achievable data rate depending on bandwidth, SNR and number of distinguishable symbols,
- what is the difference between channel and source coding,
- how line codes such as RZ, NRZ, Manchester, and MLT-3 work,
- what is the difference between baseband transmissions and modulated signals,
- which frequency ranges are used for digital transmission, and
- what fundamentally different types of transmission media are used.

Literature and references

[1] *Bell Syst. Tech. J.*. C. E. Shannon.
Communication in the presence of noise, 1949.
https://www.noisebridge.net/images/e/e5/Shannon_noise.pdf.

[2] B. Ivo.
A stripped foiled twisted-pair f/utp) cable, 2007.
https://commons.wikimedia.org/wiki/File:FTP_cable3.jpg.

[3] B. Ivo.
Unshielded twisted-pair cable with different twist rates, 2007.
https://commons.wikimedia.org/wiki/File:UTP_cable.jpg.

[4] E. Stein.
Taschenbuch Rechnernetze und Internet, chapter Codierung und Modulation, pages 59–66.
Fachbuchverlag Leipzig, 2. edition, 2004.

[5] Tkgd2007 and Fleshgrinder.
Koaxialkabel Schnittmodell, 2009.
https://commons.wikimedia.org/wiki/File:Coaxial_cable_cutaway_new.svg.

[6] M. Werner.
Nachrichtentechnik – eine Einführung für alle Studiengänge, chapter Digitale Übertragung im Basisband, pages 127–136.
Vieweg + Teubner, 6. edition, 2007.

[7] M. Werner.
Nachrichtentechnik – eine Einführung für alle Studiengänge.
Vieweg + Teubner, 6. edition, 2007.