Communication Networks

VITMAB06

Analog & Digital Voice Transmission
Voice codecs

Gusztáv Adamis

BME TMIT

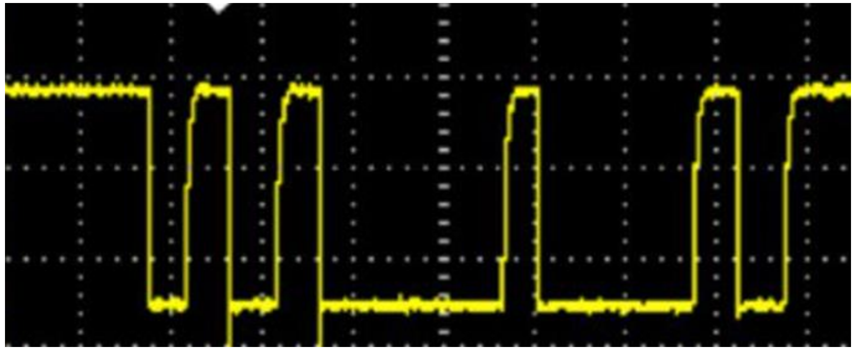
2024

TMiT

©

Voice transmission

1. Analog voice transmission
2. Analog & digital signals
3. Voice digitalization
4. Codecs



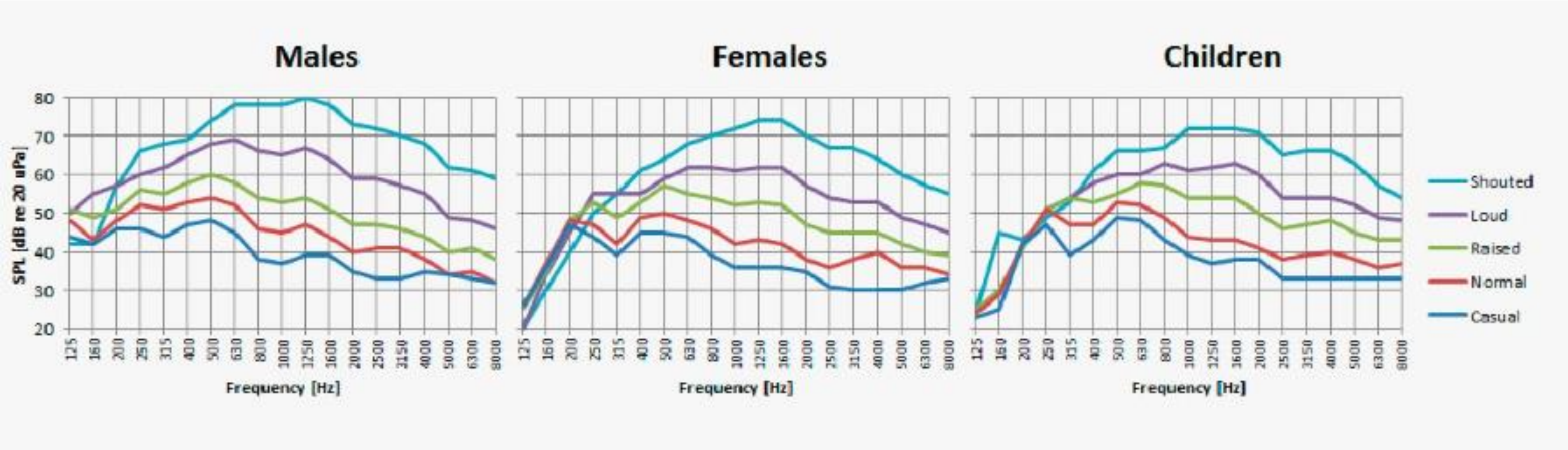
Analog voice transmission

- Terminal: sound wave ↔ analog electrical signal
- What do we know about the voice?
  - The human ear hears about 20 Hz - 20 kHz
    - Depends on age (you probably ~12 kHz)

  - https://www.rapidtables.com/tools/tone-generator.html

Analog voice transmission

- Terminal: sound wave ↔ analog electrical signal
- What do we know about the voice?
  - The human ear hears about 20 Hz - 20 kHz
    - Depends on age (you probably ~12 kHz)
  - The upper limit of the speech is 6-7 kHz



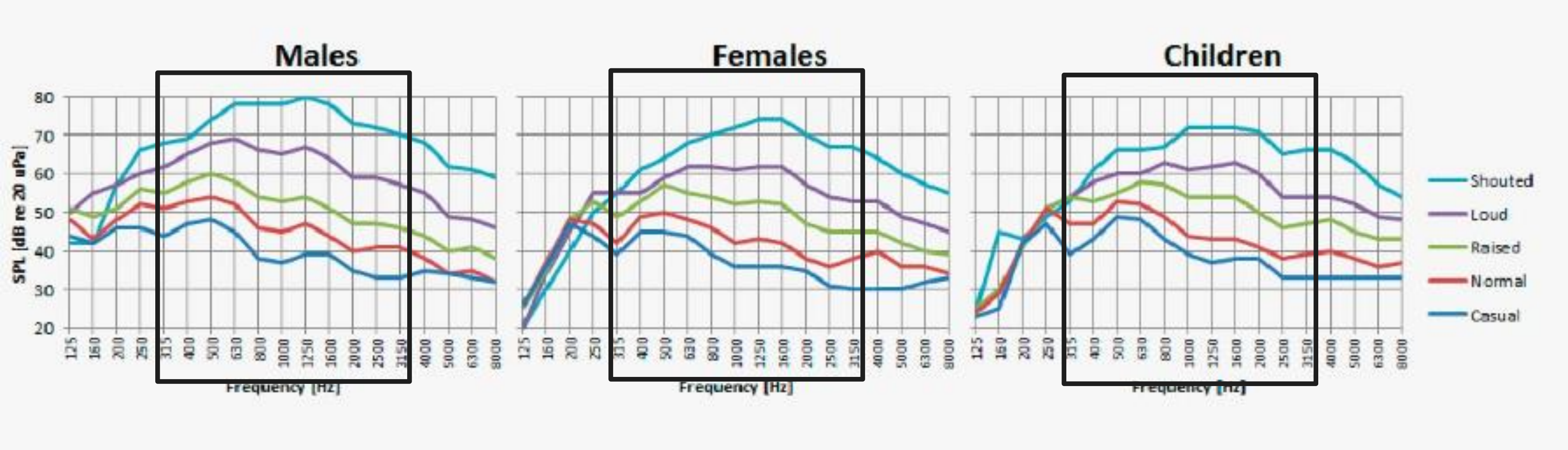
In telecommunication, voice means human's voice (speech)

Analog voice transmission

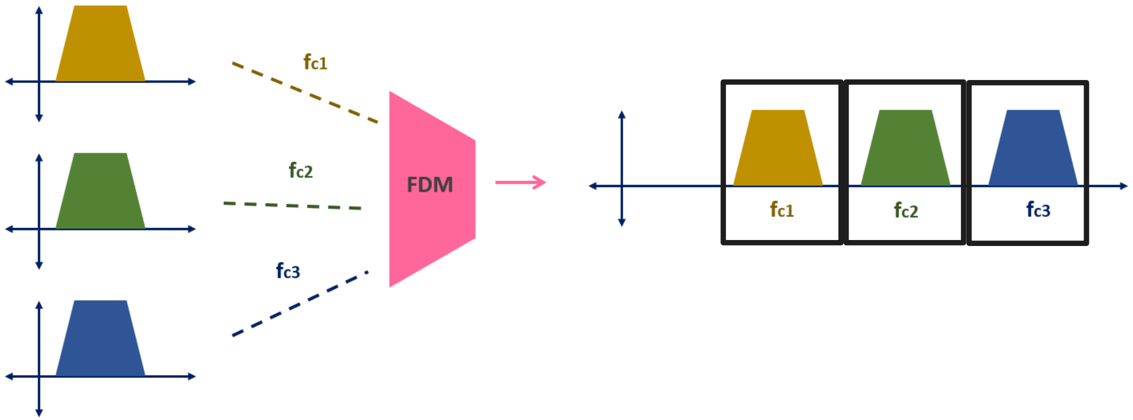
- How many Hz wide should a voice channel be?
  - the goal is simply to transmit intelligible, understandable speech
  - \+ economical aspects!
- Early multiplexed trunks used FDM (Frequency Division Multiplexing)
  - the narrower a voice channel is, the more can fit on a trunk line
- Nowadays: digital transmission – TDM (Time Division Multiplexing), IP
  - but bitrate matters here, too …
  - and it is proportional to the transmitted bandwidth (see later: PCM)
- Understandability and voice quality as a function of transmitted frequency :
  - 500…1000 Hz: poor
  - 500…1500 Hz: tolerable
  - 400…2000 Hz: satisfactory
  - 300…2500 Hz: sufficient
  - <u>300…3400 Hz: good</u>
  - 200…3500 Hz: very good
- Decision: **0,3 – 3,4 kHz band**
  - 3,1 kHz + guarding gaps = **A voice channel will be 4 kHz wide**

Analog voice transmission

- Voice 300-3400 Hz (0.3-3.4 kHz)

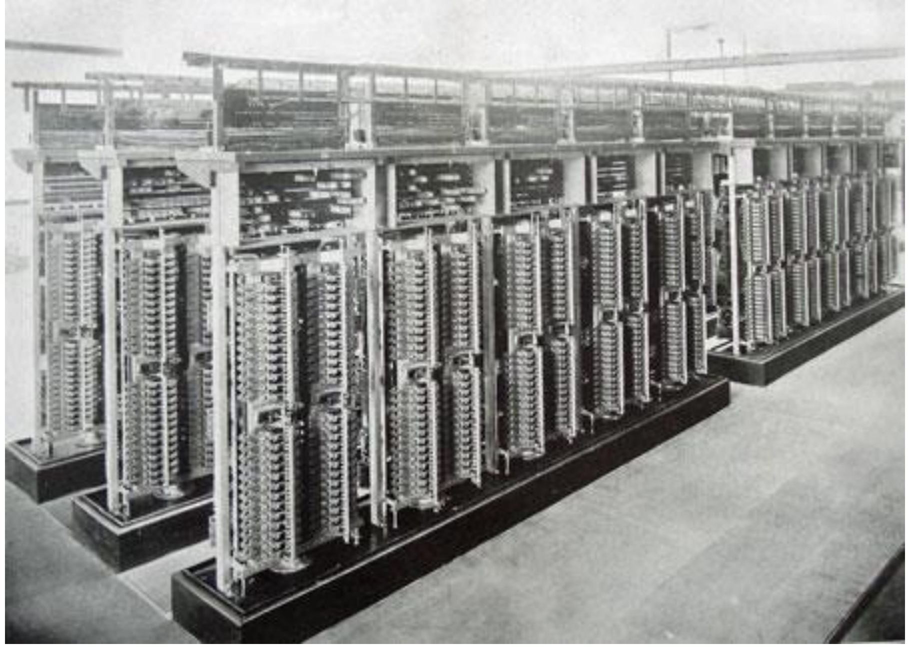


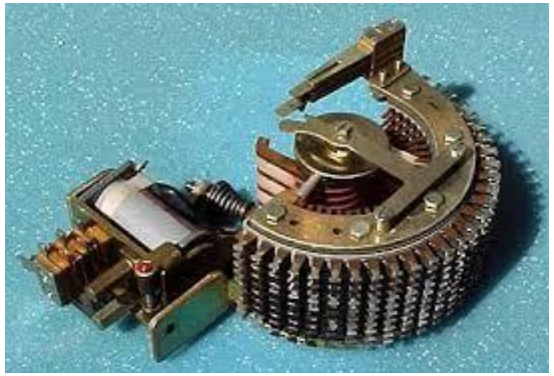
- FDM signal (schematic)
  - Non-ideal devices
  - To reduce the inter-channel influence => gap btw them
  - => 4 kHz



Analog telephone systems

- Initially end-to-end analog transmission
  - Analog terminals
  - Analog multiplexing on trunks (FDM, Frequency Division Multiplexing)
  - (Electro)mechanical switching, continuous galvanic connection inside the switch





The expansion of digital technology

- Digital technology appeared (during/after World War II)
- With many advantages compared to analog:
  - simpler, more reliable, cheaper
  - less space and power requirements
  - the signal/noise ratio is independent of the size of the network (though, the bit error rate may depend)
  - switching is possible without moving parts
  - a higher degree of network intelligence can be realized
  - much more sophisticated signaling is possible
  - data and voice signals can be handled in a uniform way
- It is worth digitalizing the voice signal
  - and send over a digital telephone network this way
  - a huge transformation in the network!

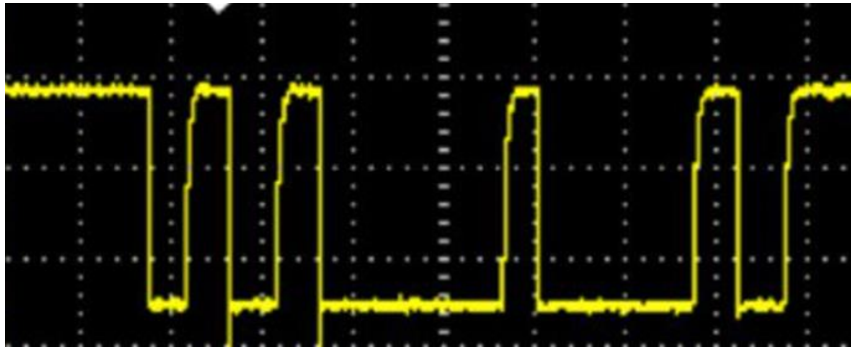
Digital telecommunication networks

- Starting in the 1960s/1970s, network elements in developed countries were gradually replaced by digital ones
  - in Hungary it is finished around 2000
  - today, moreover, they are transferred in IP packets
  - Exception: the majority of *wired terminals* (telephones) remained analog
    - A/D-D/A conversion in the first network device
- How to digitalize voice?

www.postamuzeum.hu

Voice transmission

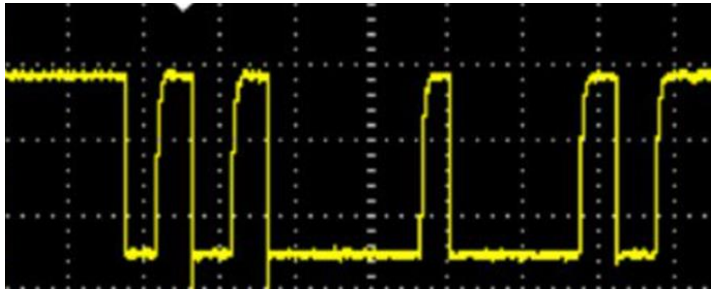
1. Analog voice transmission
2. Analog & digital signals
3. Voice digitalization
4. Codecs



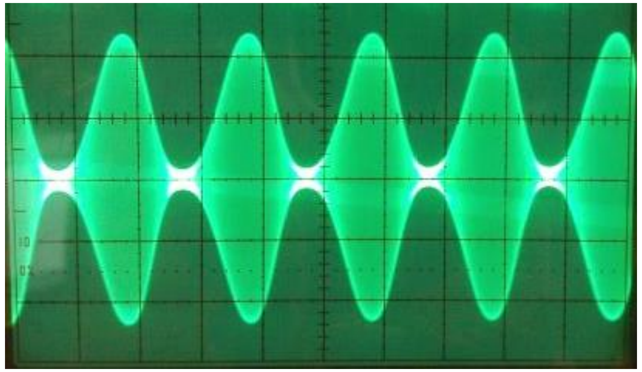
Analog and digital signals

- Are the signals below analog or digital?

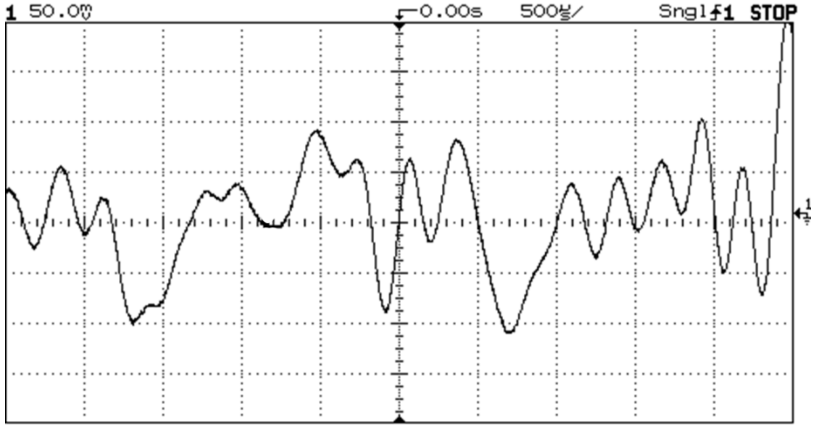
1.



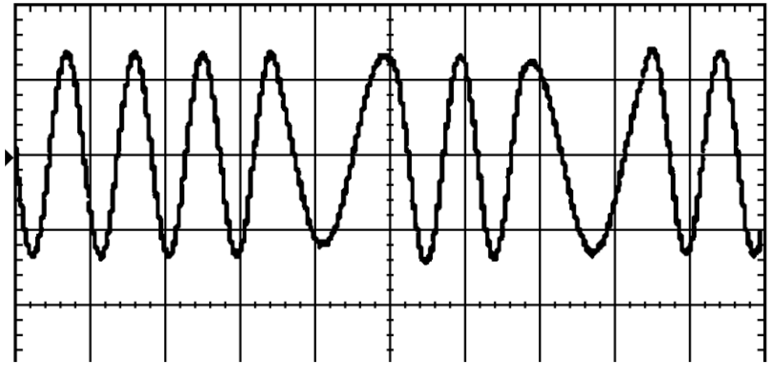
2.



3.

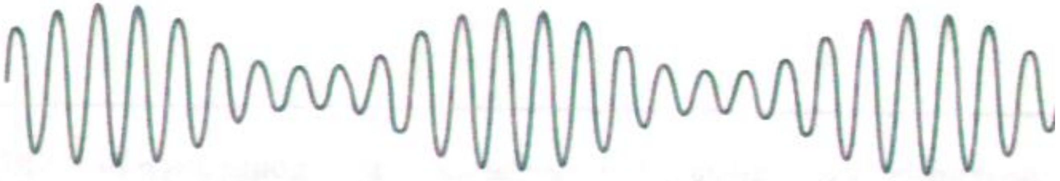


4.



Analog and digital signals

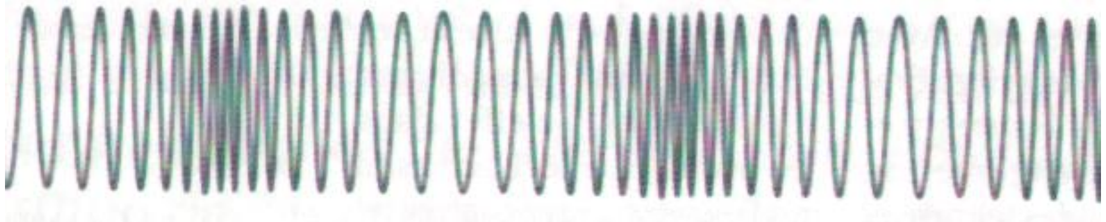
5.



6.

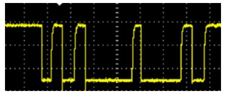
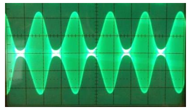
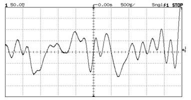
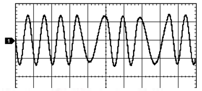
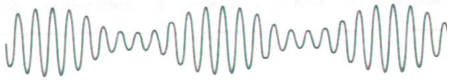
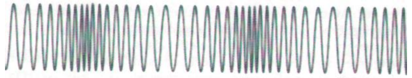
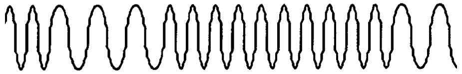
7.

8.



9.

Solution :)

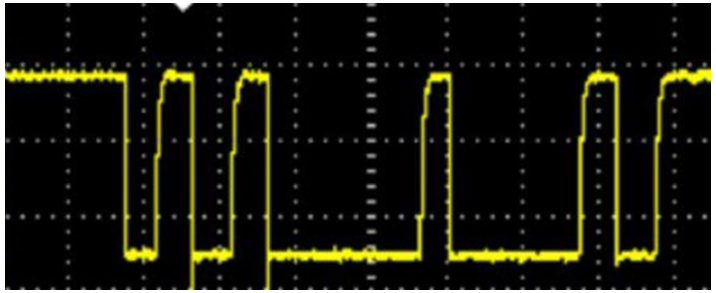
- 1 D 
- 2 A 
- 3 A 
- 4 D 
- 5 A 
- 6 D 
- 7 D 
- 8 A 
- 9 D 
- *In fact, this cannot be determined just by looking at it!*

Analog and digital signals

These were intended to be digital:

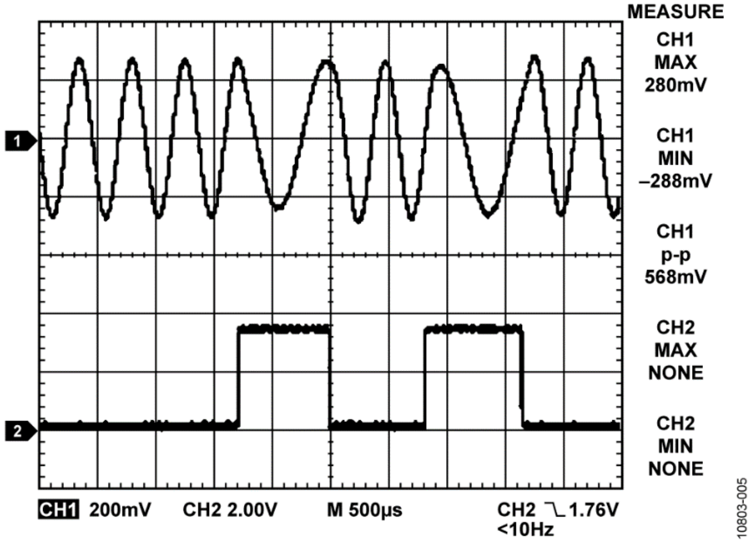
- Baseband signal (level encoding):

1.



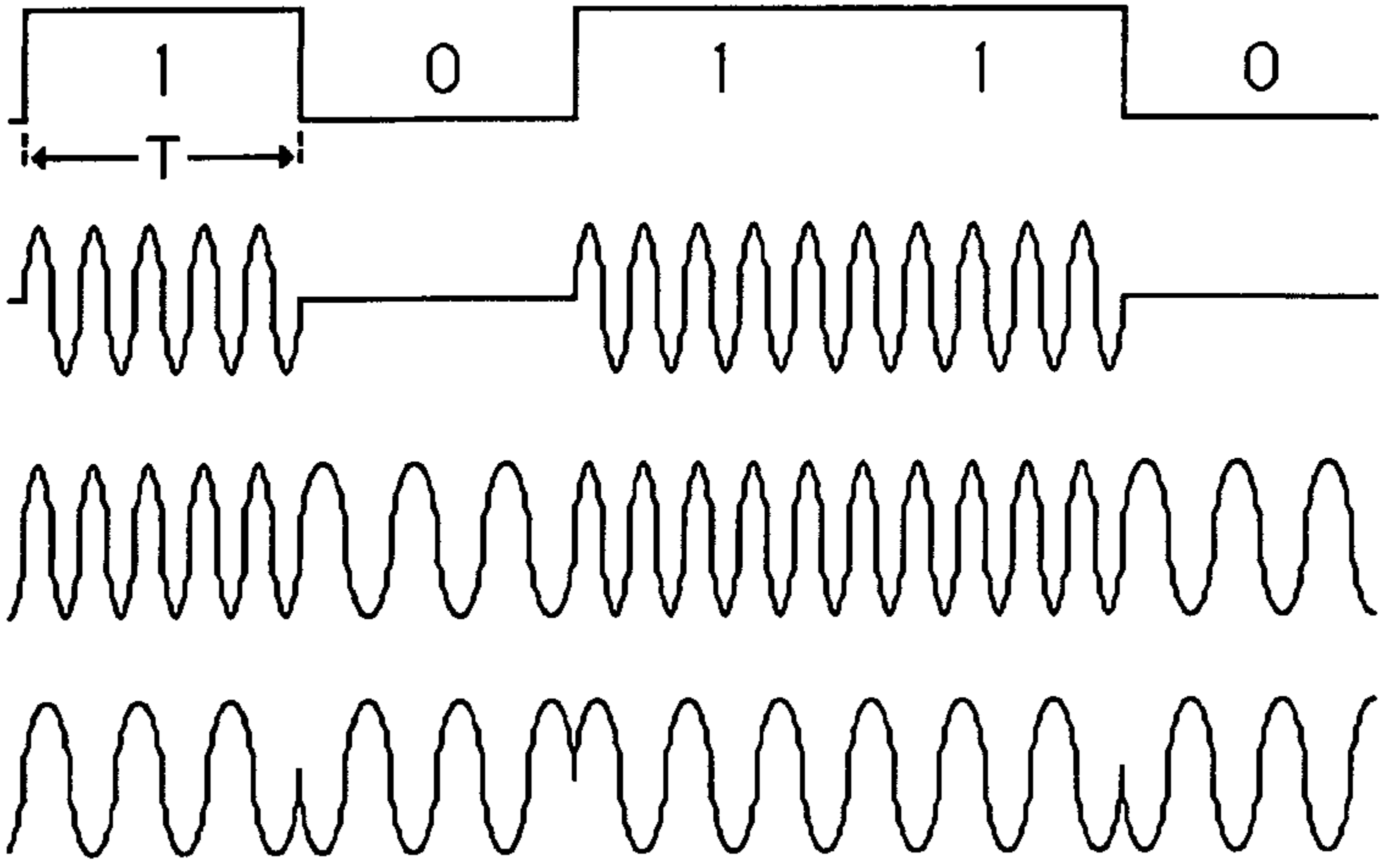
- FSK (Frequency Shift Keying):

4.



Analog and digital signals

These were also intended to be digital :



6. Amplitude-shift keying ASK

9. Frequency-shift keying FSK

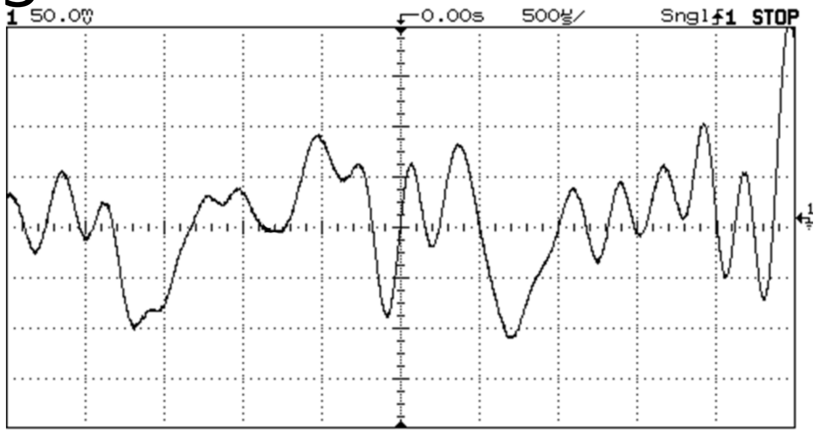
7. Phase Shift Keying PSK

Analog and digital signals

These were intended to be analog:

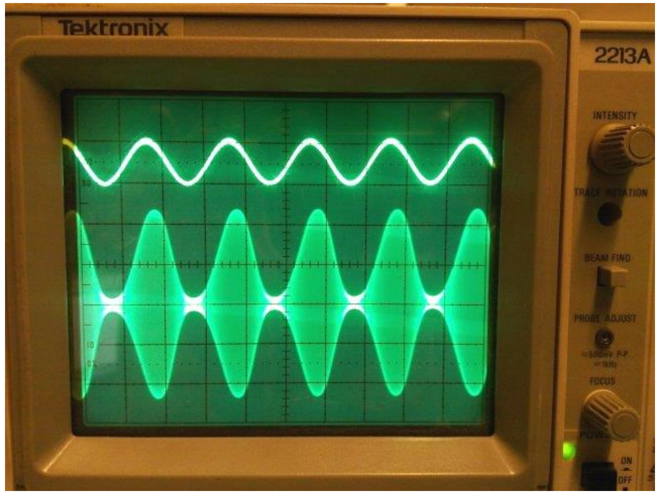
- Analog voice signal:

3.



- AM, Amplitude Modulation :

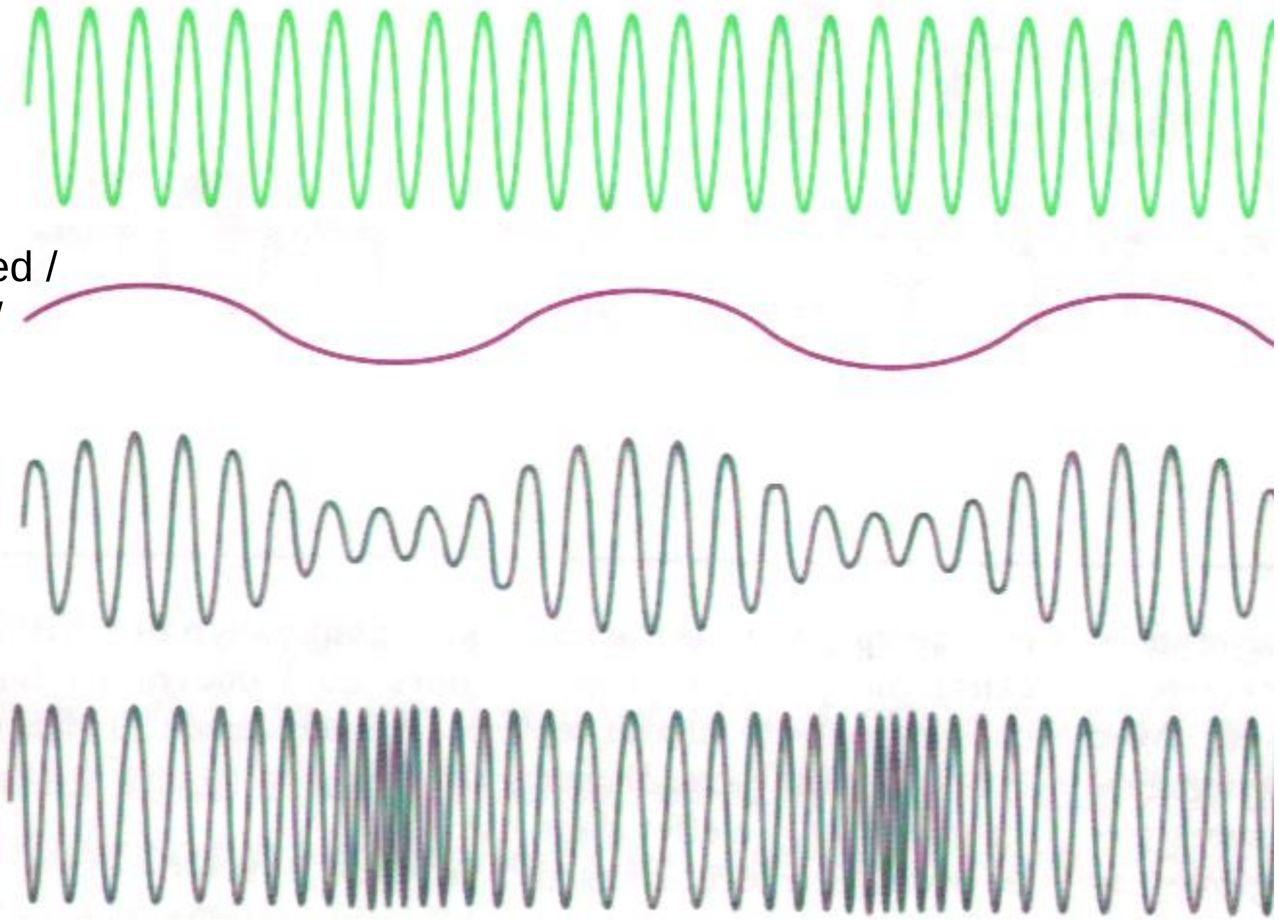
2.



Analog and digital signals

These were intended as a special analog :

- AM, FM (=frequency modulation):



carrier signal

signal to be transmitted /
modulating signal /
analog information

5. AM

8. FM

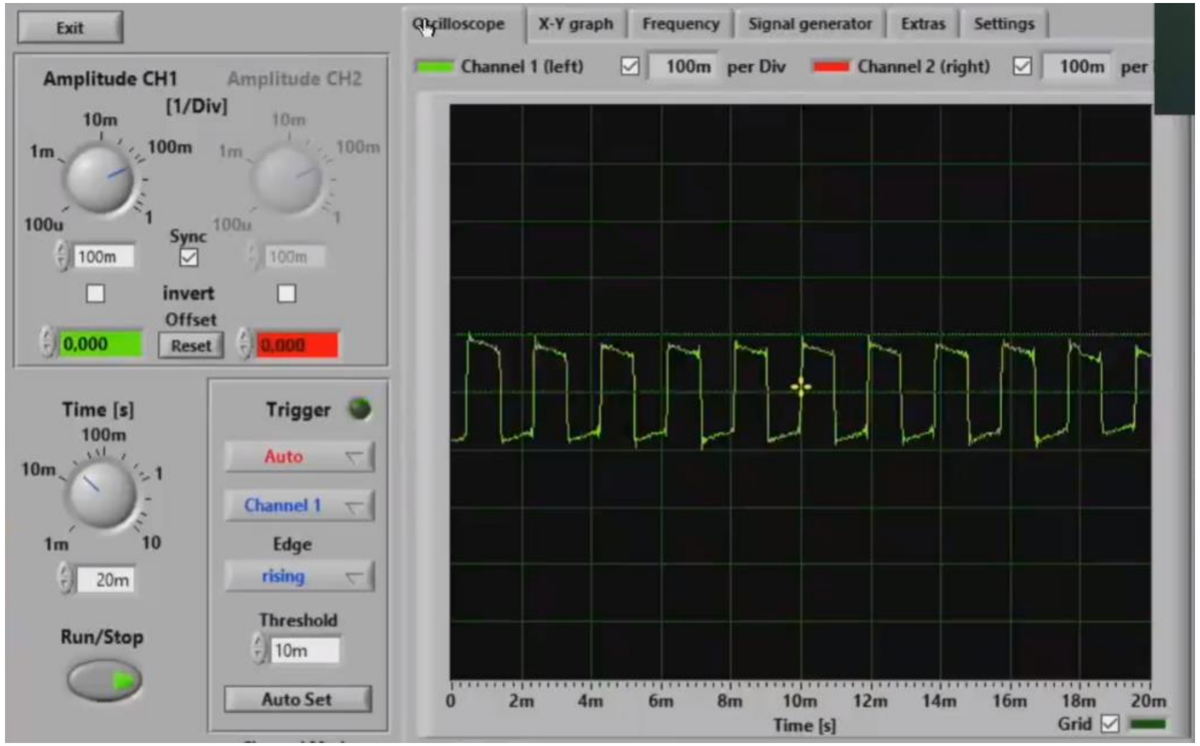
Analog and digital signals

- All signals appearing on a physical medium are “analog”

- *It will be digital if we interpret it as such*
  - not necessarily a square signal, but it can be
  - the important thing is to have some discrete states
  - not necessarily just two
    - What’s the difference btw digital and binary?
  - see later modulation and line encoding

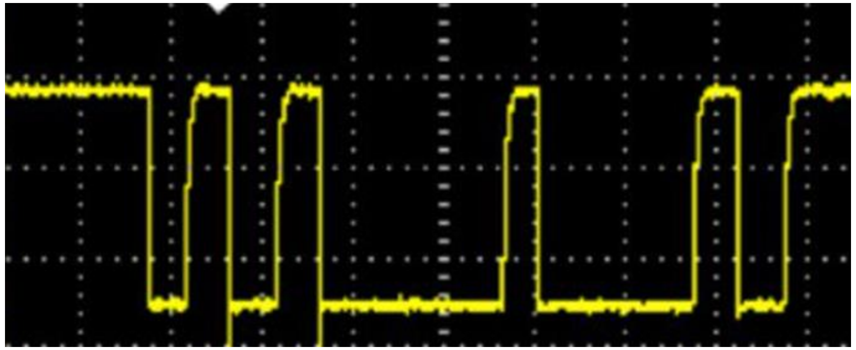
Analog and digital signals

- *It will be analog if we interpret it as such*
  - e.g. music realized with a square signal is considered analog
    - Super Mario’s music
    - https://www.youtube.com/watch?v=THLrhYvtYKg

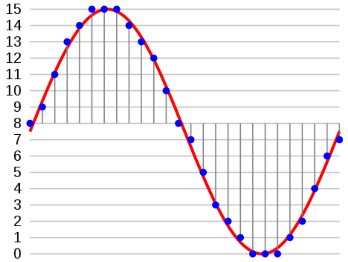


Voice transmission

1. Analog voice transmission
2. Analog & digital signals
3. Voice digitalization
4. Codecs



PCM



- PCM = Pulse Code Modulation
- The first and still widespread voice digitalizer (coder-decoder: codec)
- The principle has been known for a long time
  - It was already used for encryption in World War II
    - after digitalization, it was binary encrypted
- Telephone Network Applications: from the 1970s
- Still today it is the basis of A/D conversion
- There are newer codecs that are better in some ways
  - See later
- Even today, it is still used in many places for voice transmission
  - In fact, to any kind of sound (not only voice)
  - Excellent sound quality
  - In many places, the relatively high data speed ("bitrate") is not a problem today

The First PCM

This slide won't be asked on exam

- SIGSALY
- https://en.wikipedia.org/wiki/SIGSALY
- WW2, The Allies
- Unbreakable coding (random rekeying)
- For this, the speech had to be digitalized
- 30 kW, 40 racks, 50 tons
- About a dozen were made. Pentagon, London, a Pacific ship

PCM

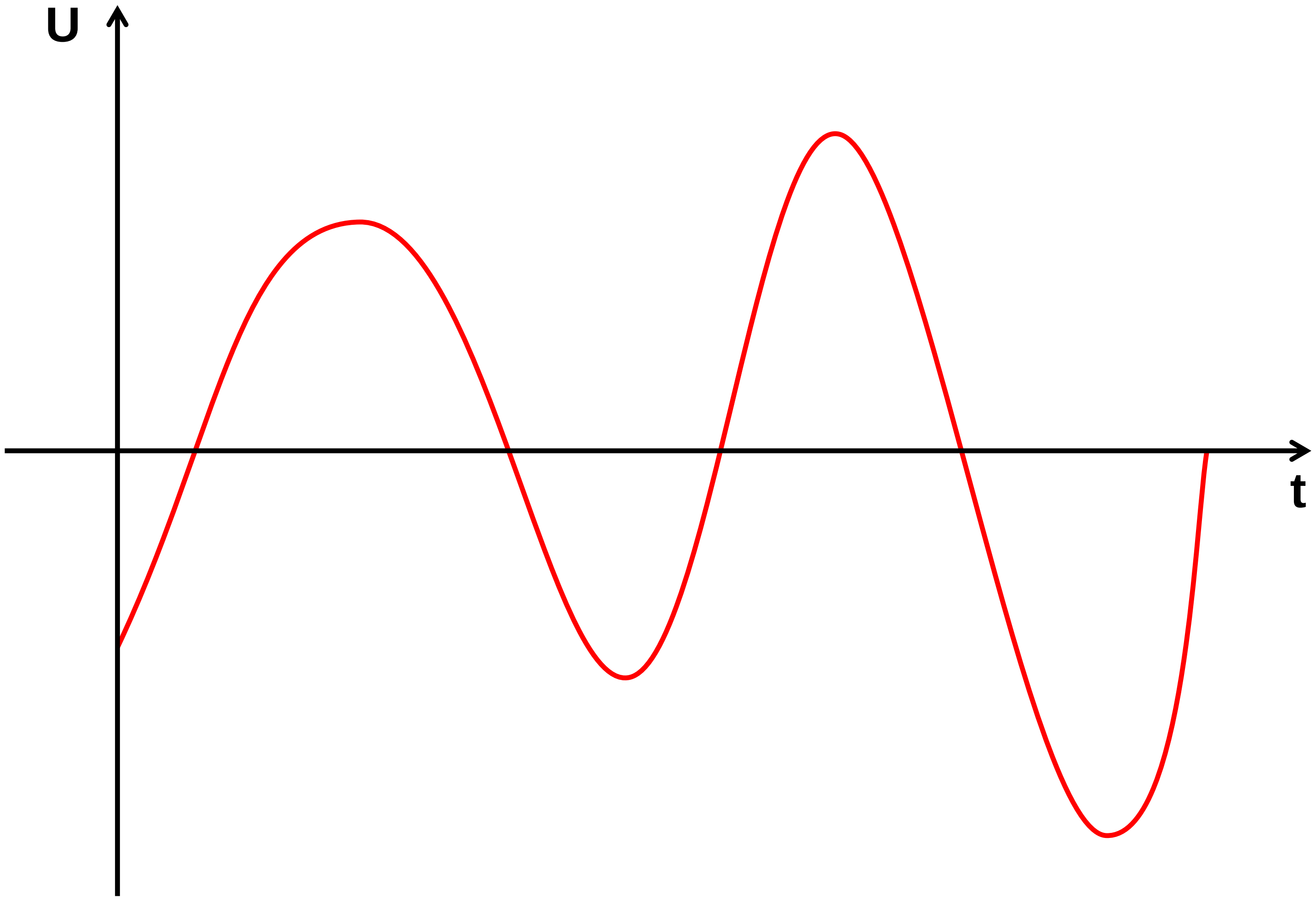
- An analog "signal" is given as a voltage-time function
  - Continuous in time (interpretation range of a function)
  - Continuous in voltage (value set of a function)
- The goal is to produce a sequence of bits
  - From which a signal very similar to the original can be restored
- Steps of A-D conversion
  - Band filtering
  - Sampling
  - Quantization
  - Encoding
- Steps of D-A conversion
  - D-A conversion with the inverse characteristic of quantization
  - Band filtering (smoothing)

Band filtering

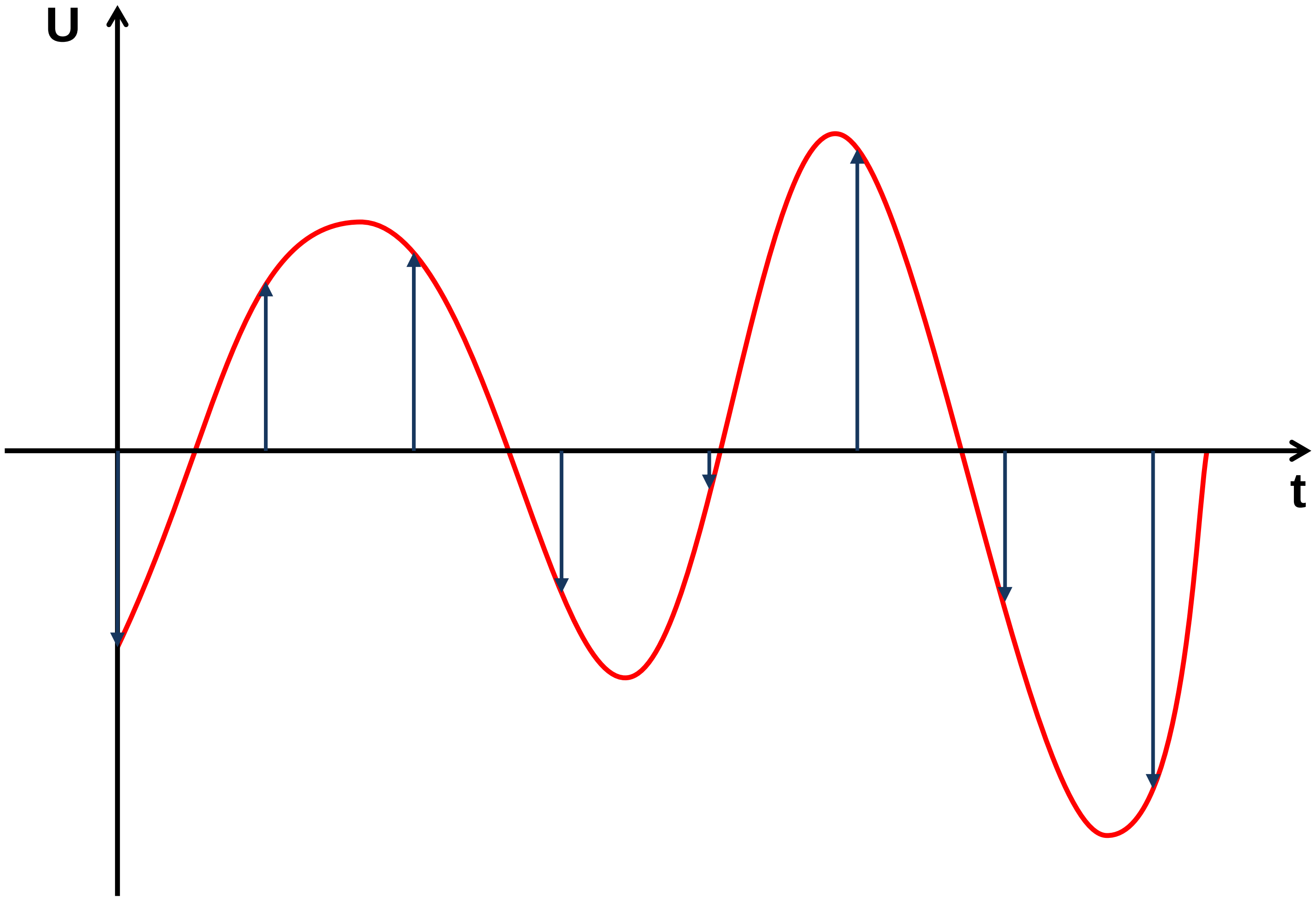
- The first step is to filter the signal to a band between 300-3400 Hz
  - That is, filtering of components outside this range
  - The band-limited signal is important for the sampling, see soon

Example: sampling

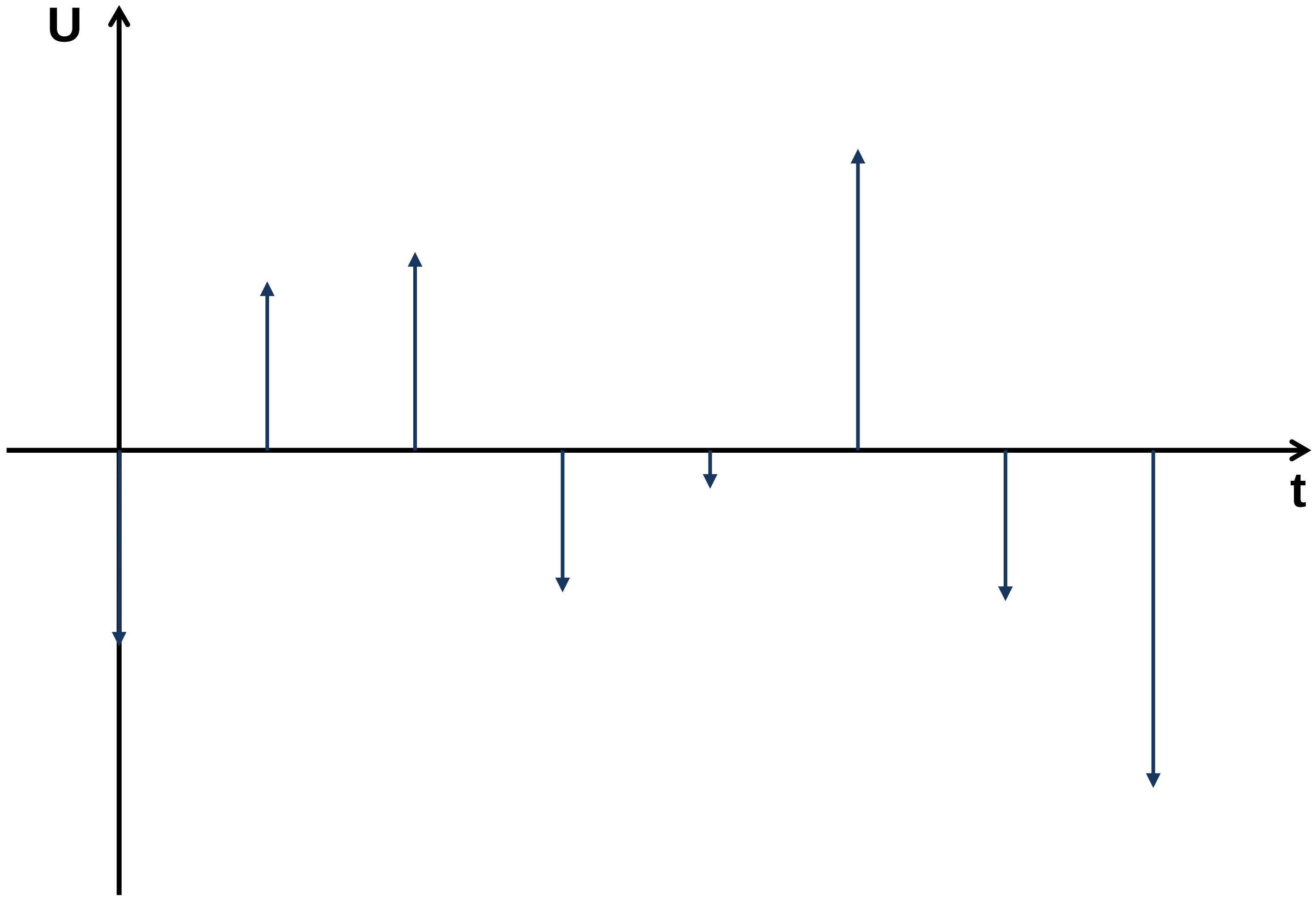
Example: sampling



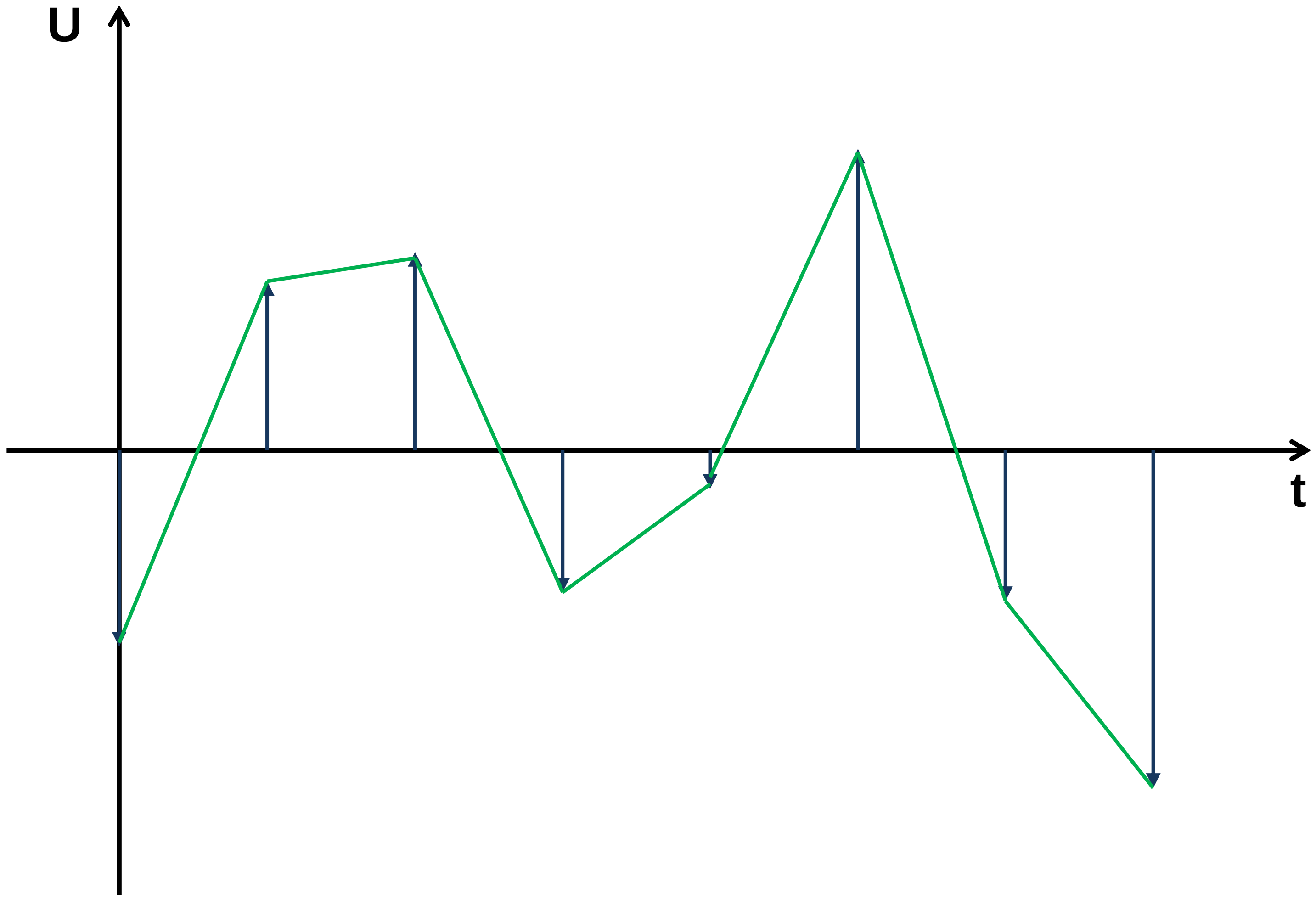
Example: sampling



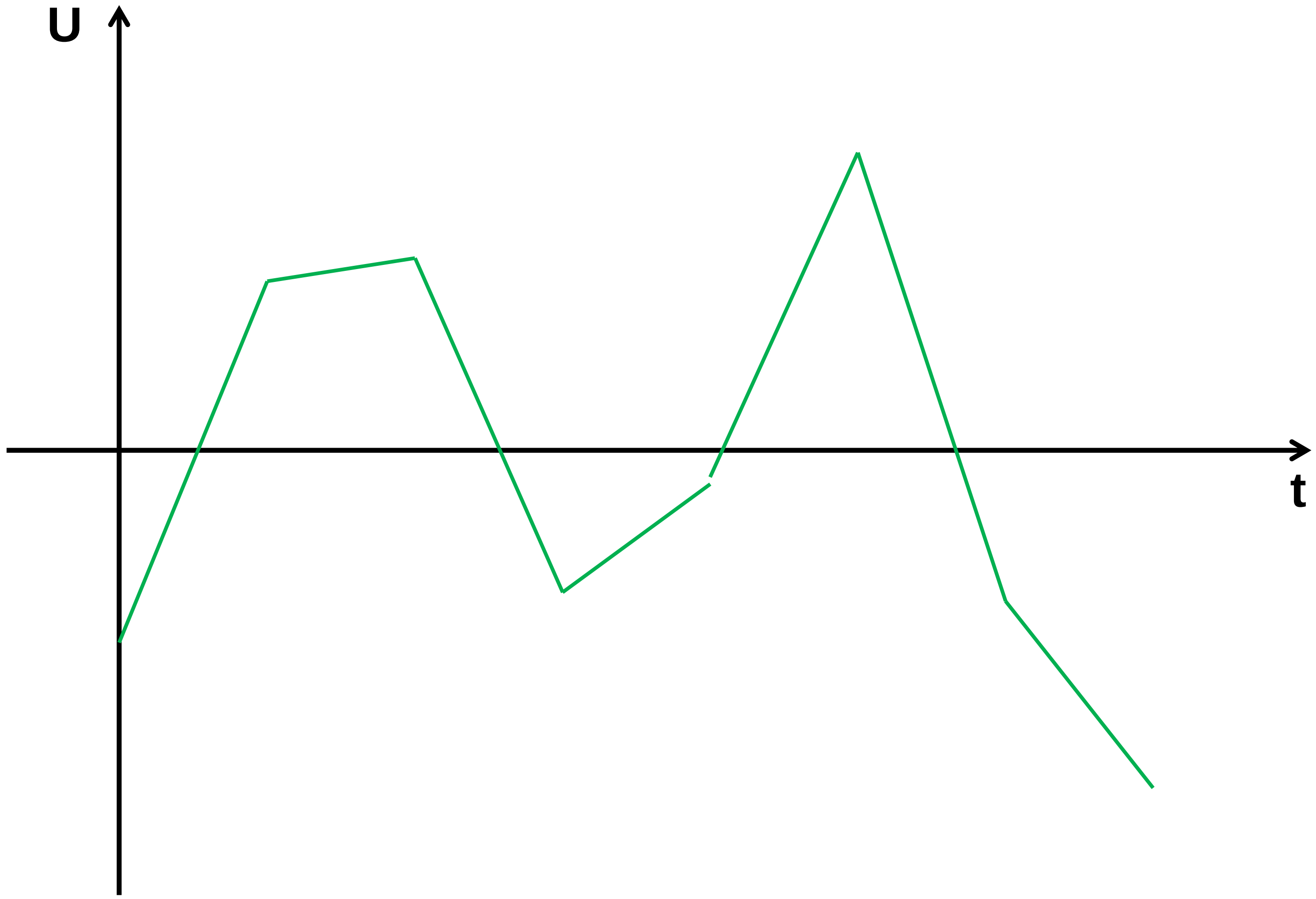
Example: sampling



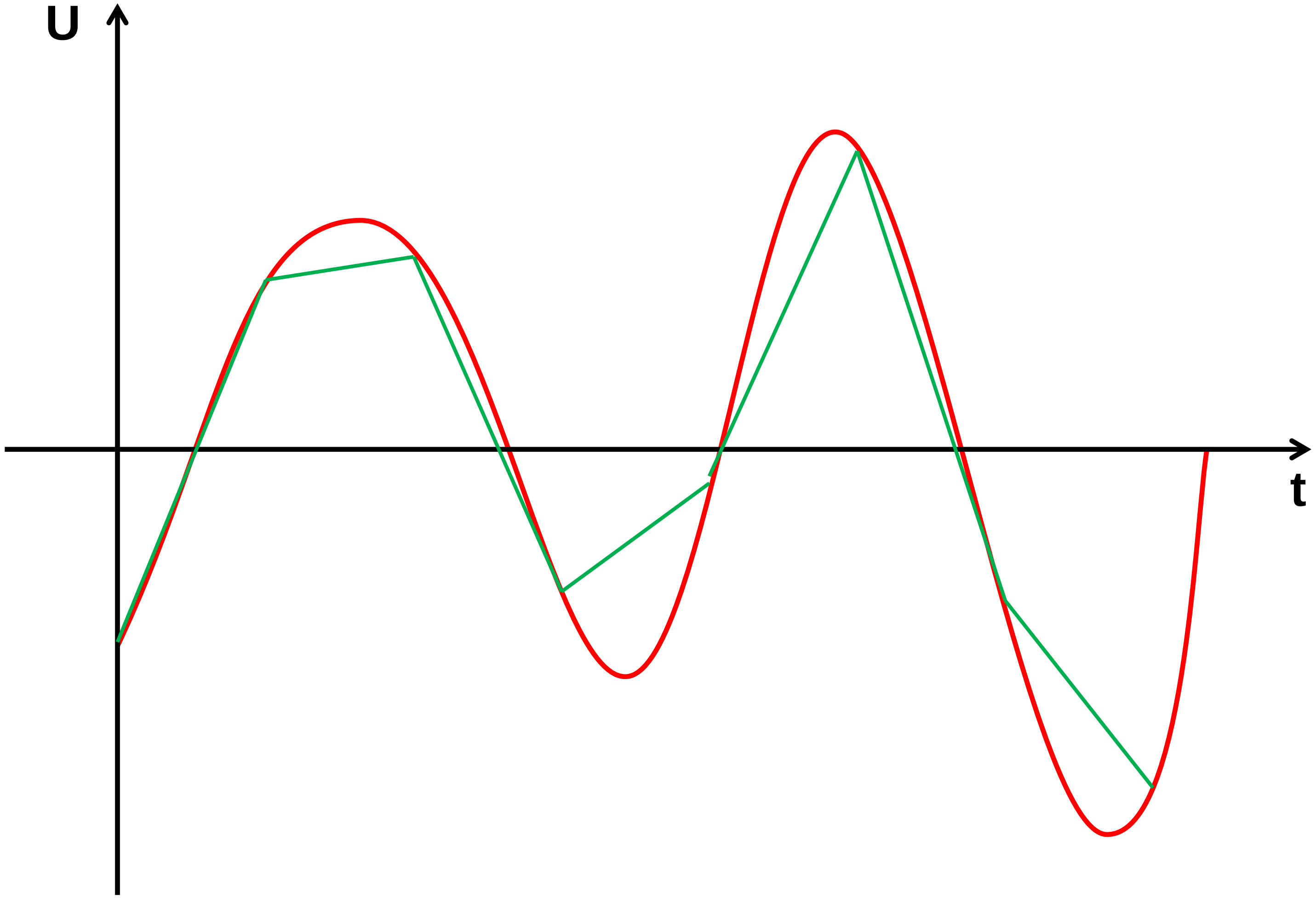
Example: sampling



Example: sampling

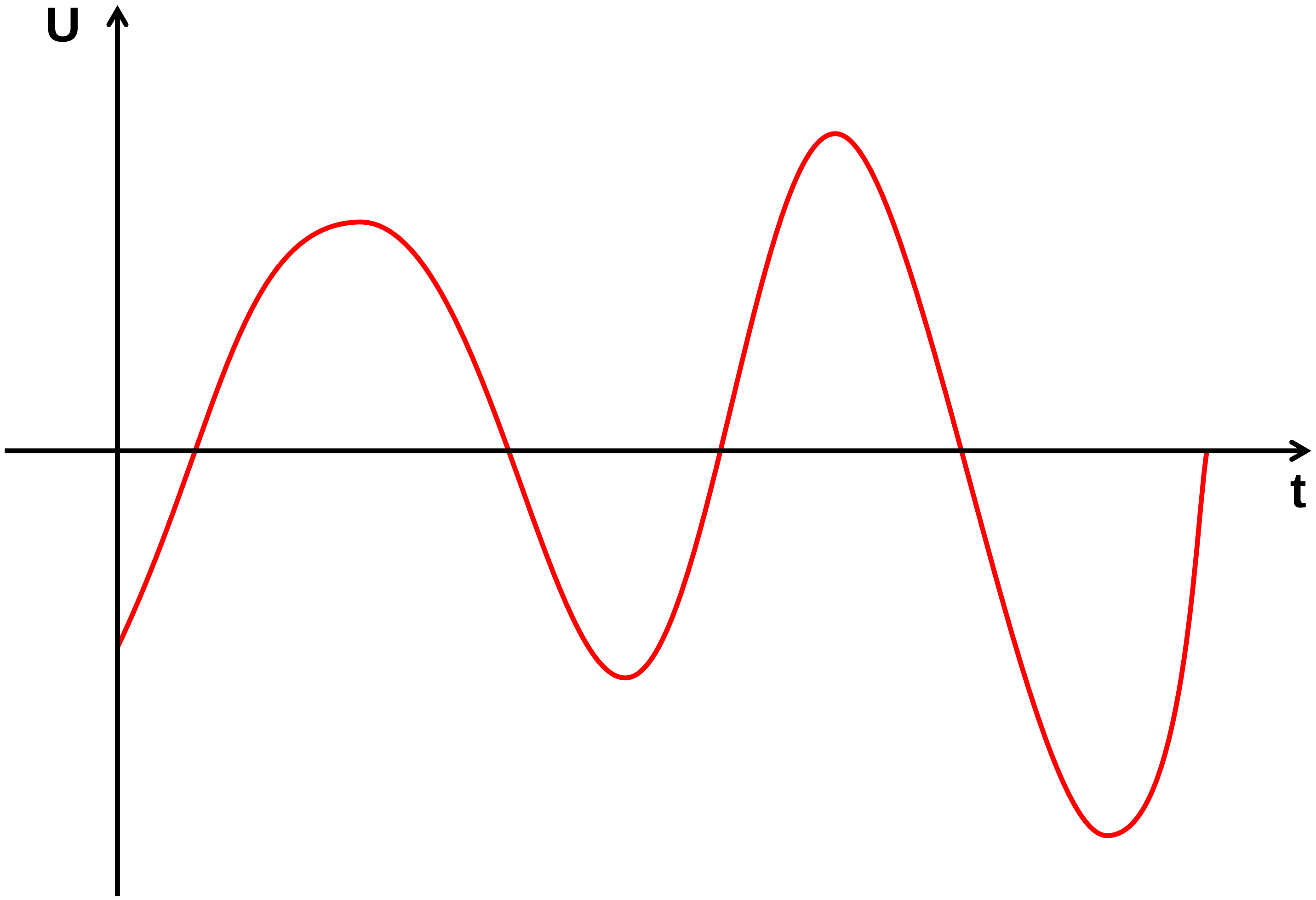


Example: sampling

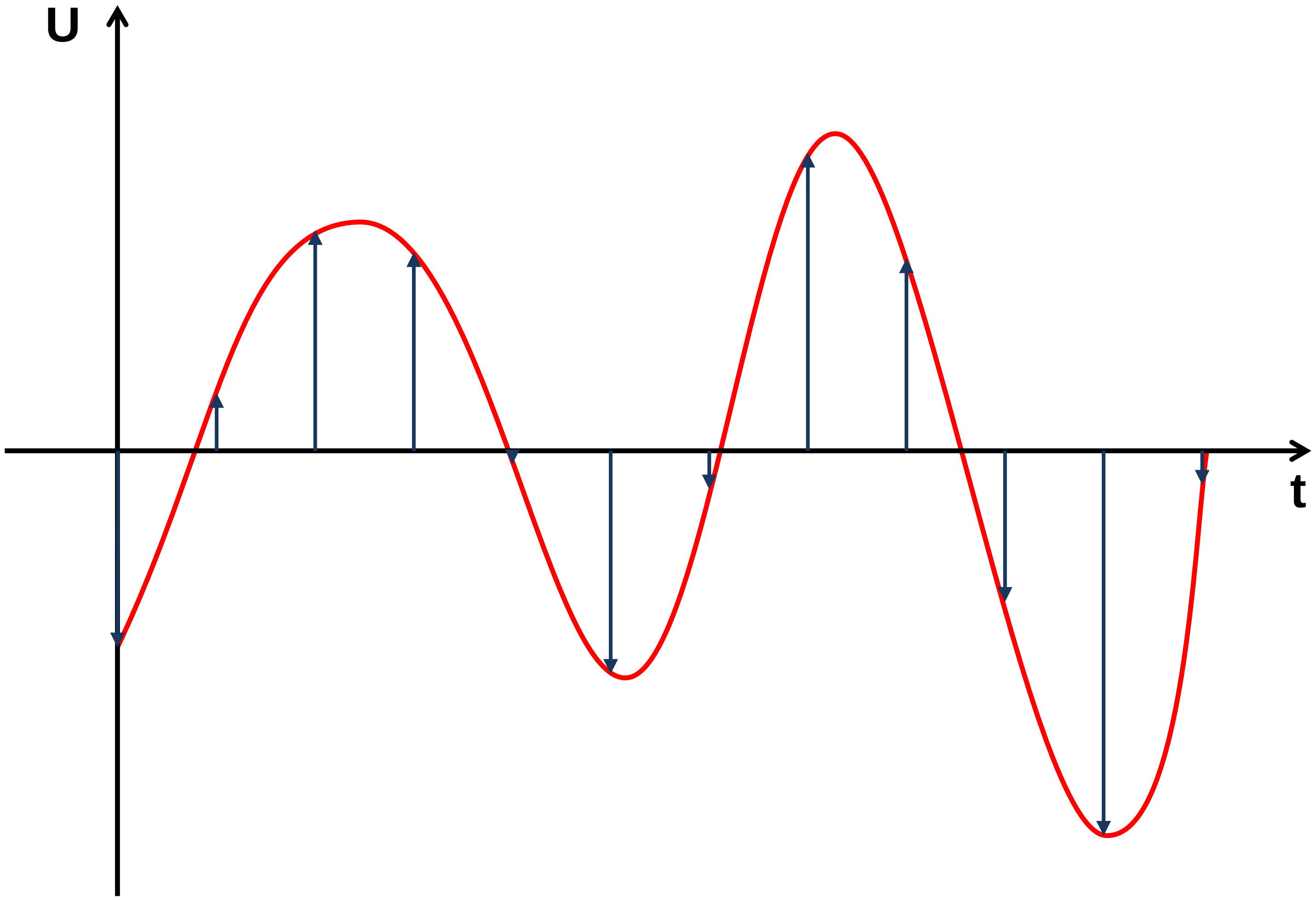


Example: sampling

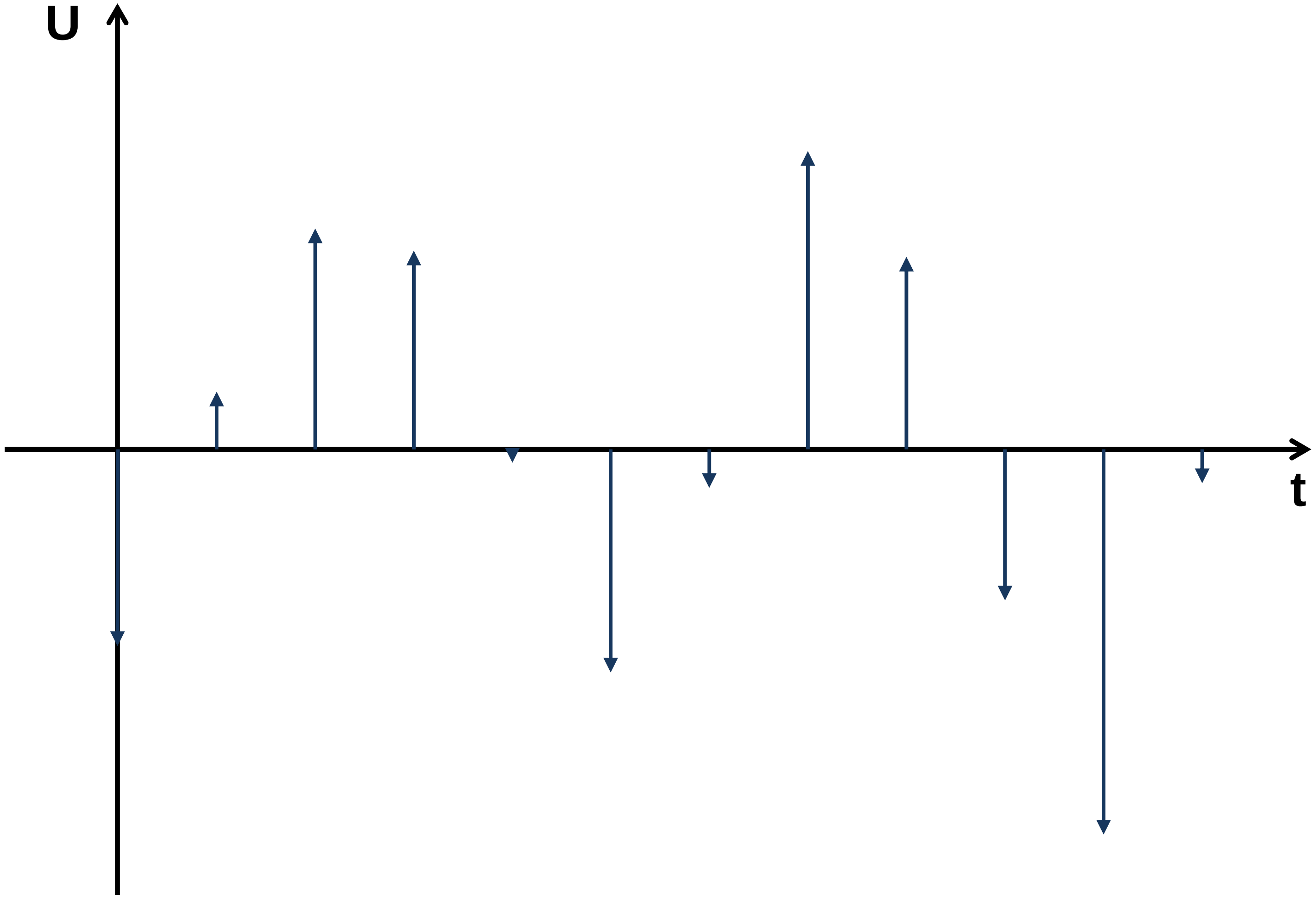
Example: sampling



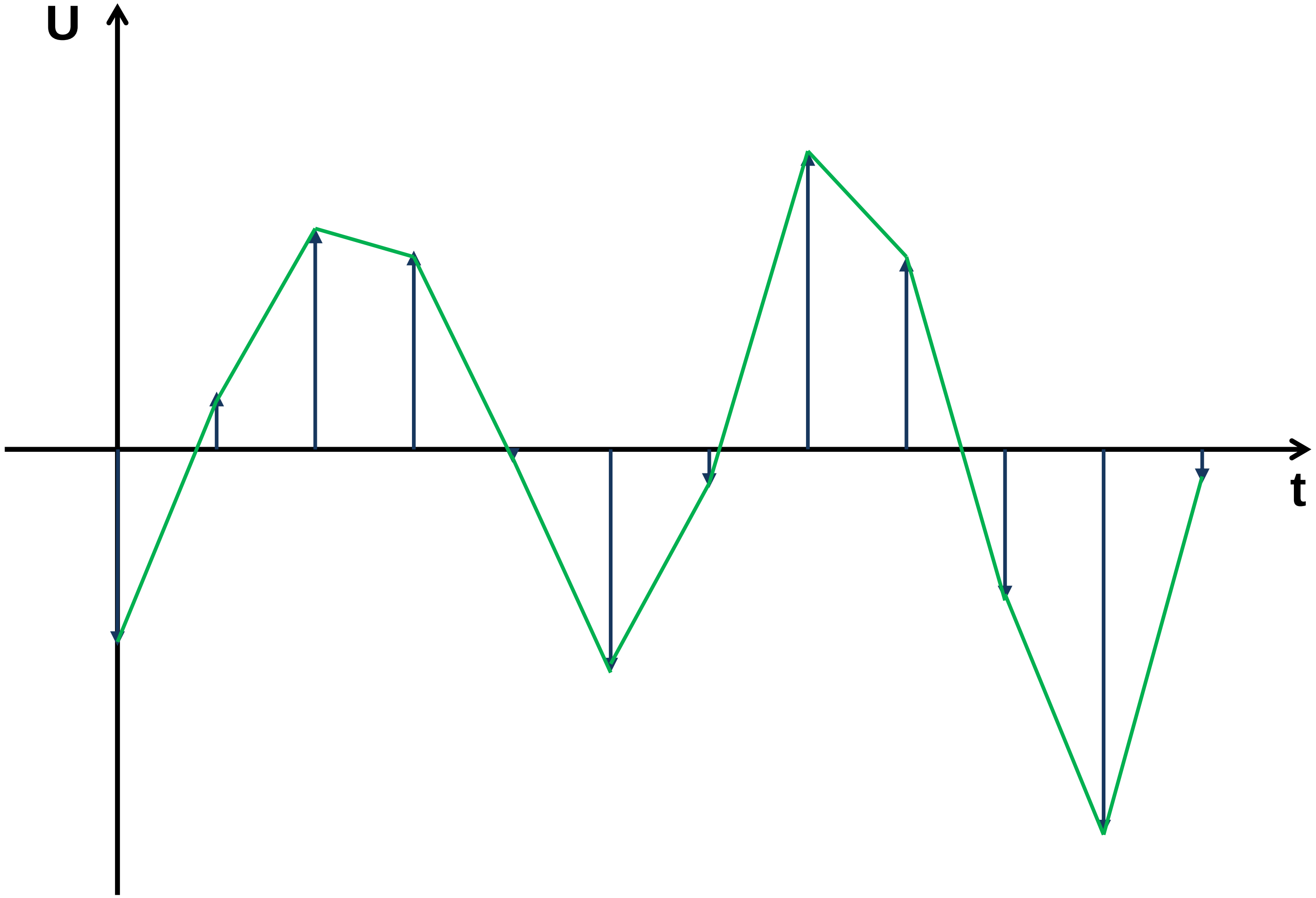
Example: sampling



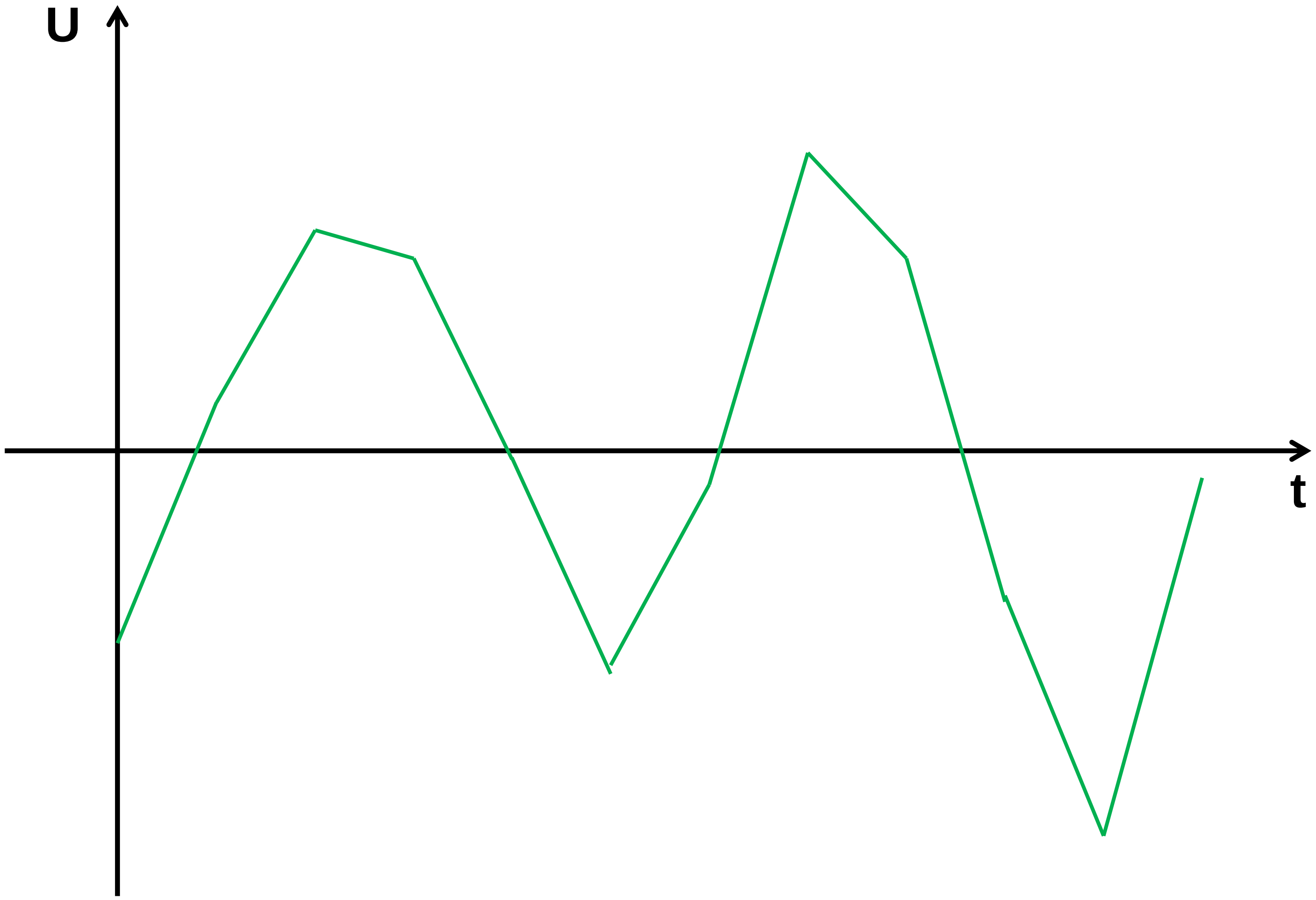
Example: sampling



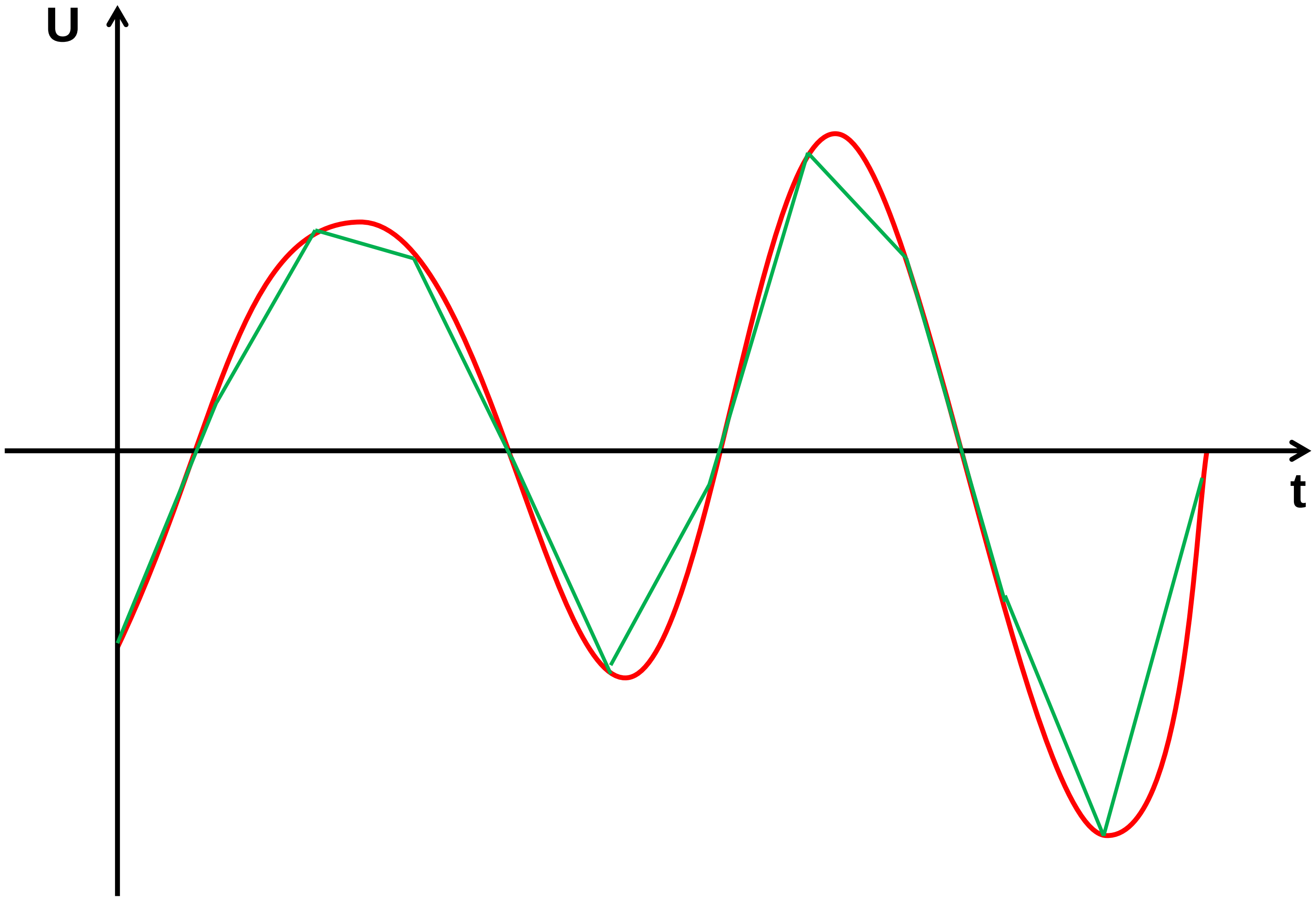
Example: sampling



Example: sampling

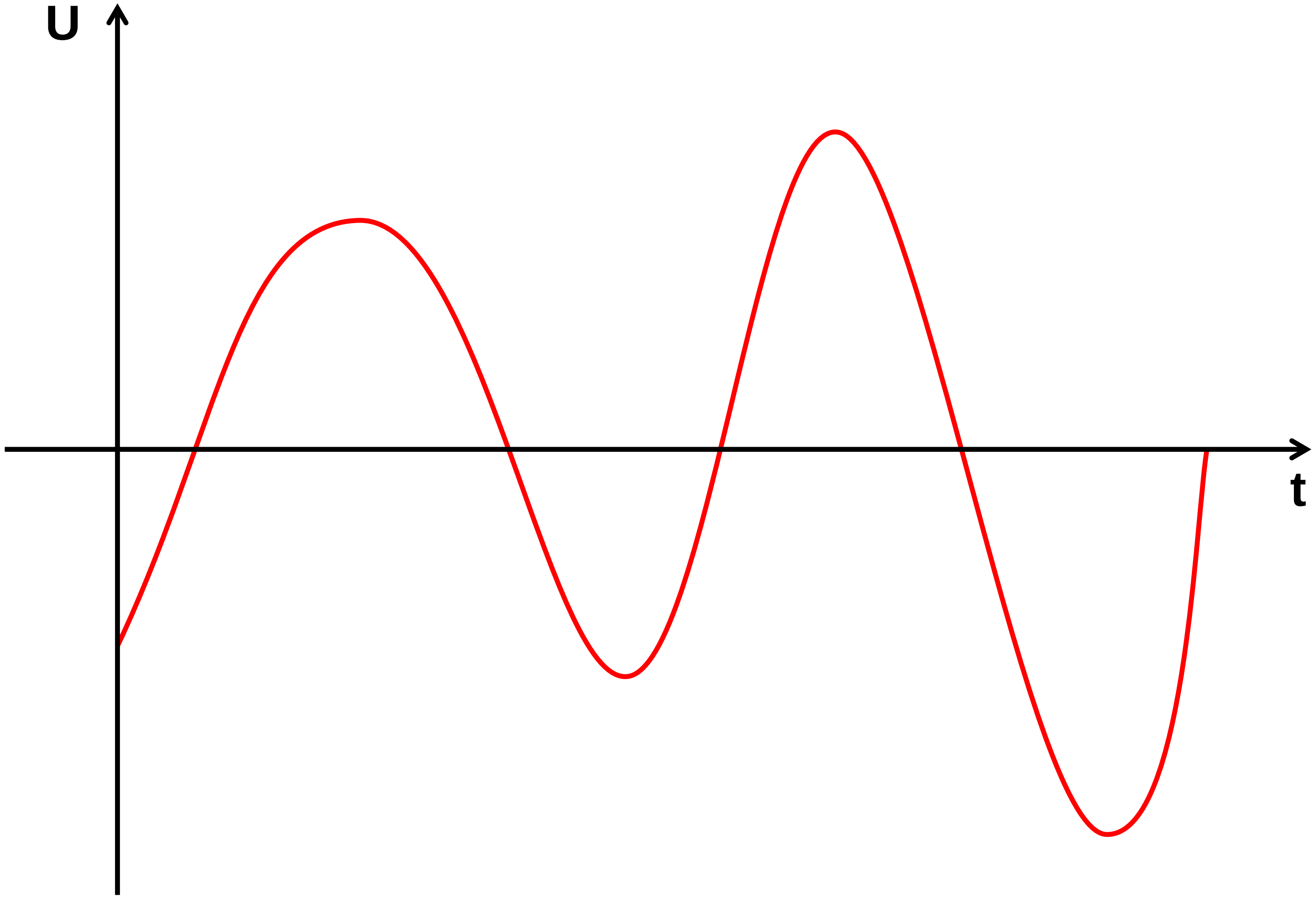


Example: sampling

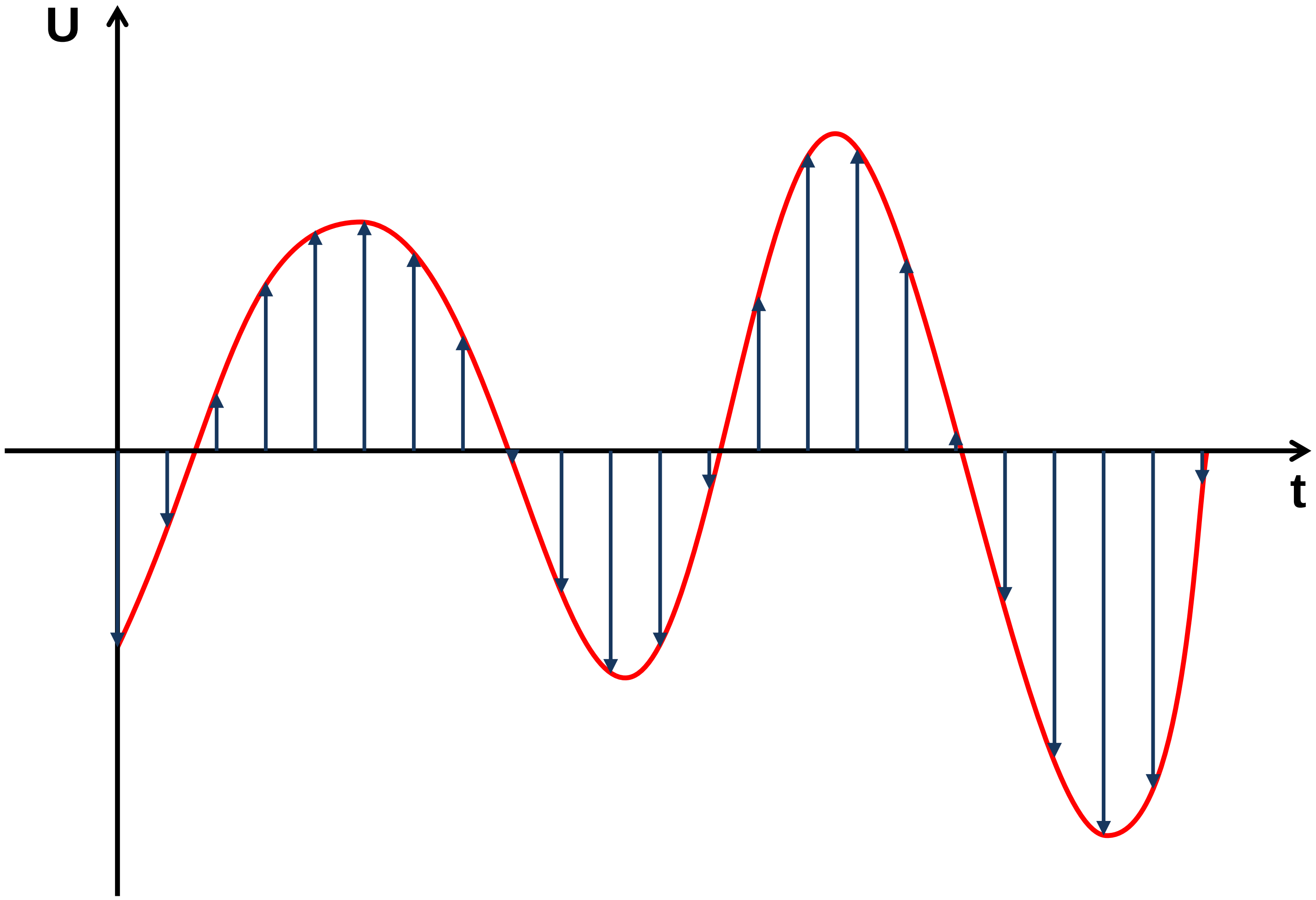


Example: sampling

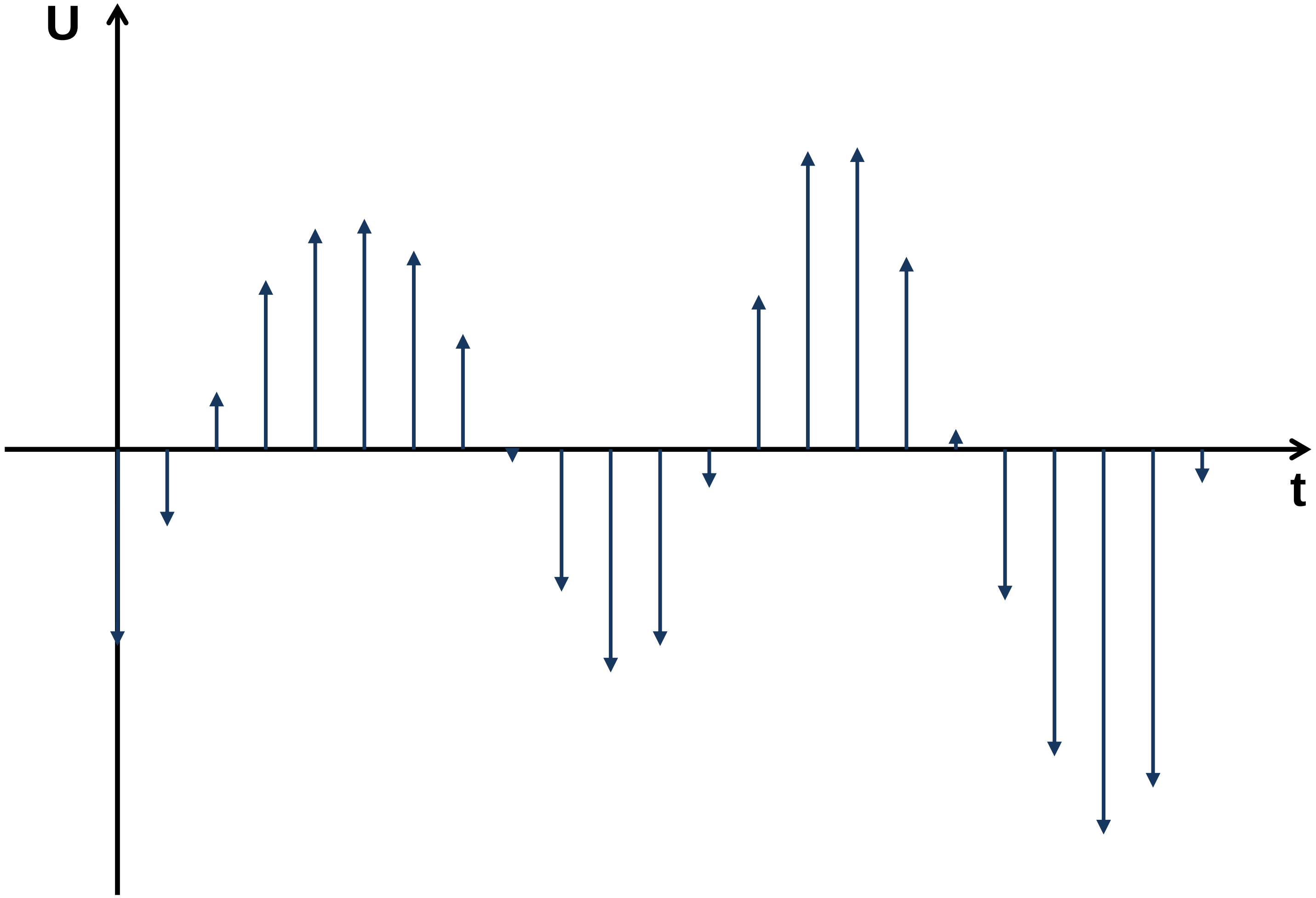
Example: sampling



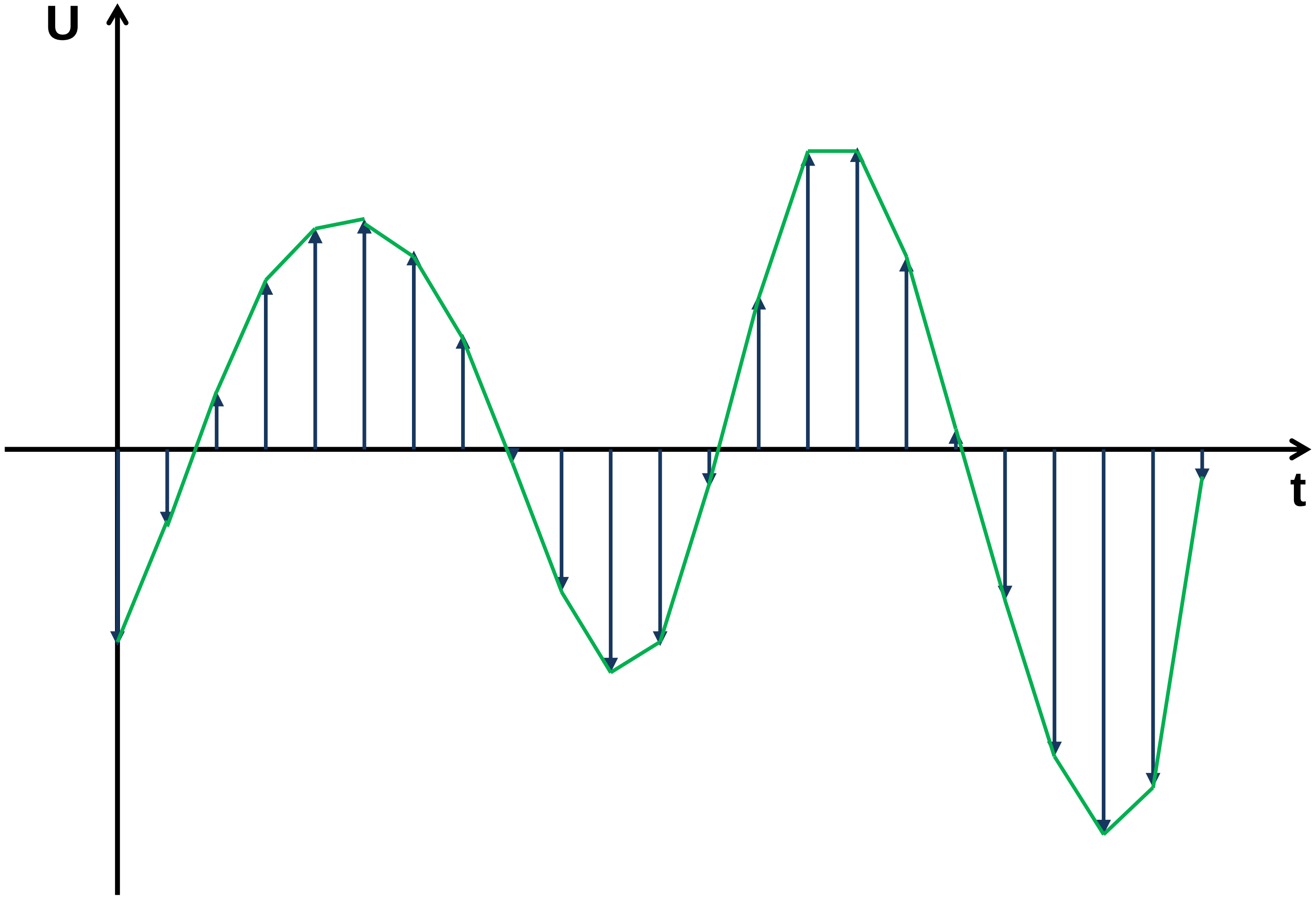
Example: sampling



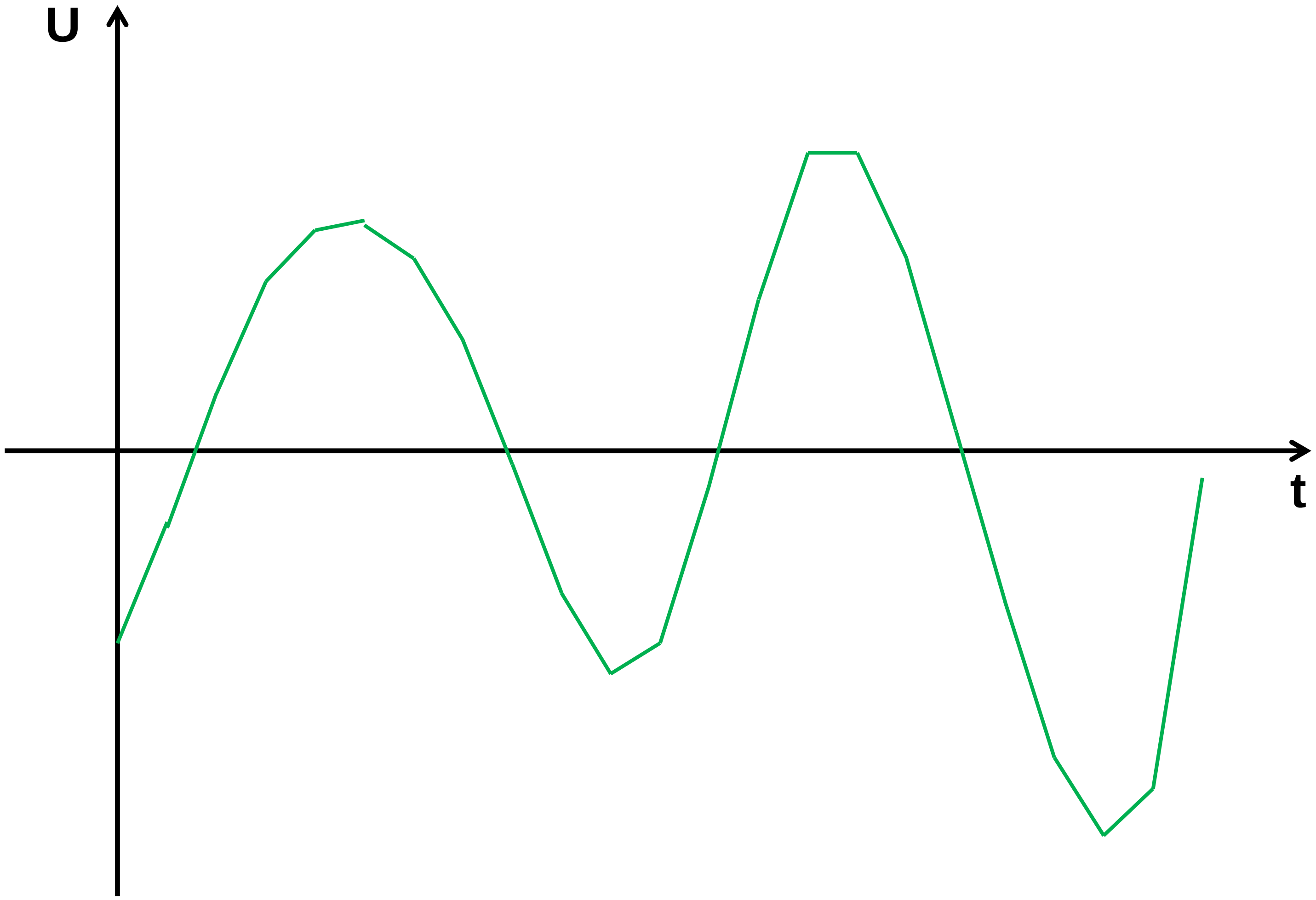
Example: sampling



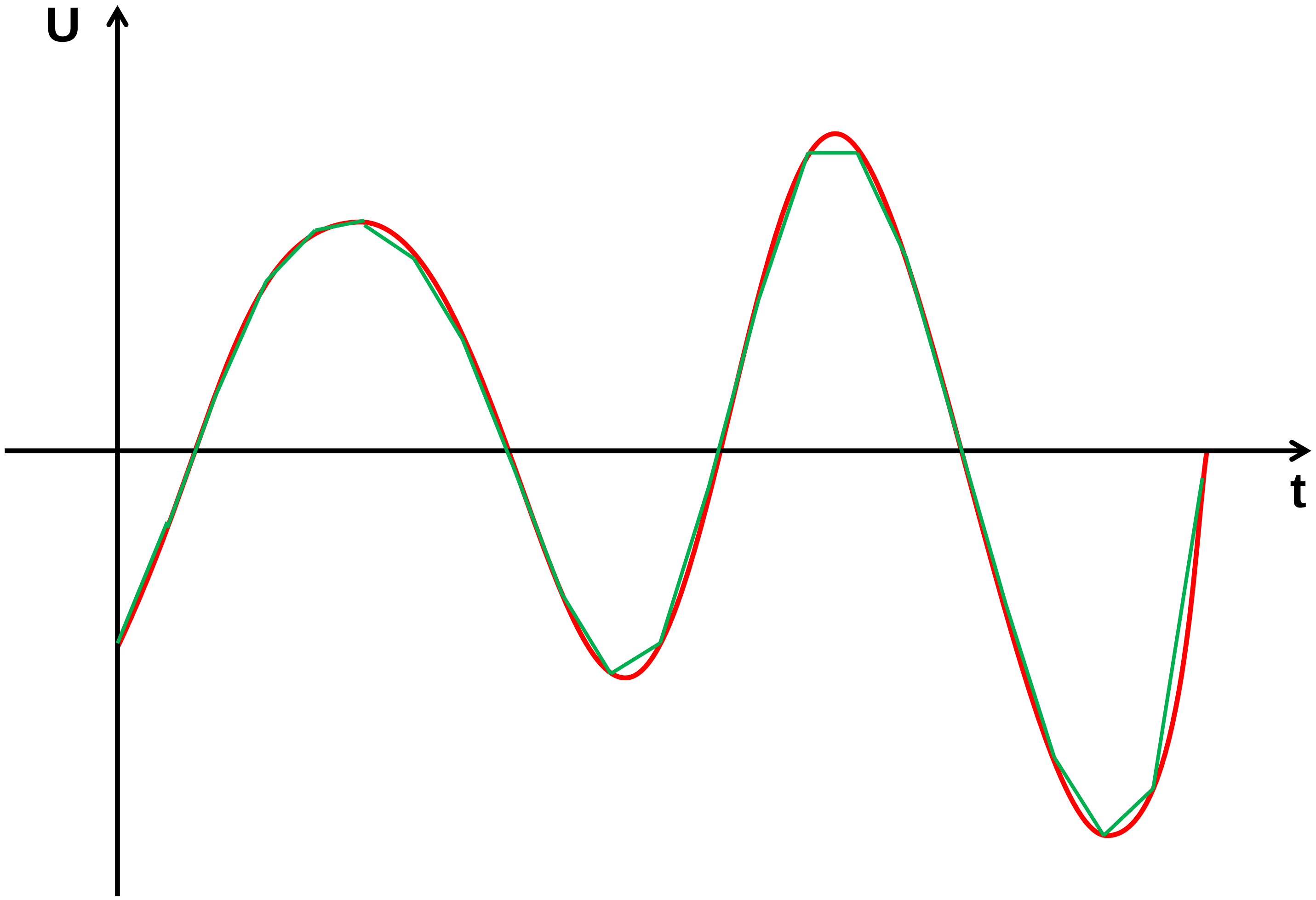
Example: sampling



Example: sampling



Example: sampling



Sampling

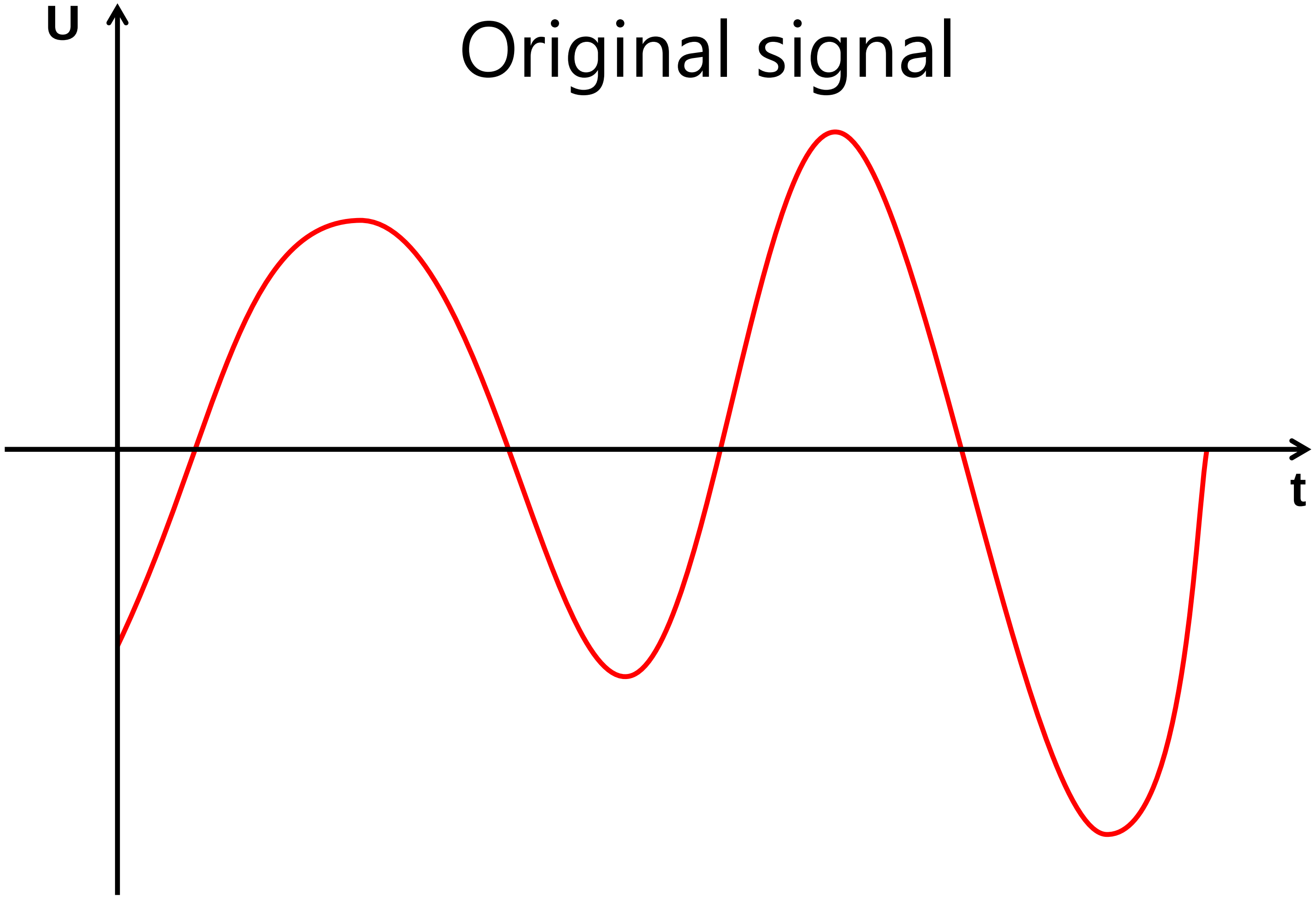
- The more samples we take, the closer we get to the original signal
- Is there such a thing as *enough* samples for a *perfect* reset?
- Nyquist-Shannon sampling theorem:
  - Yes! Double of the maximum frequency of the signal is enough for sampling frequency
    - More precisely a bit more than double
  - Then, upon restoration, after the band filtering (!), we get *exactly the original* signal back
  - Wow :)

Sampling

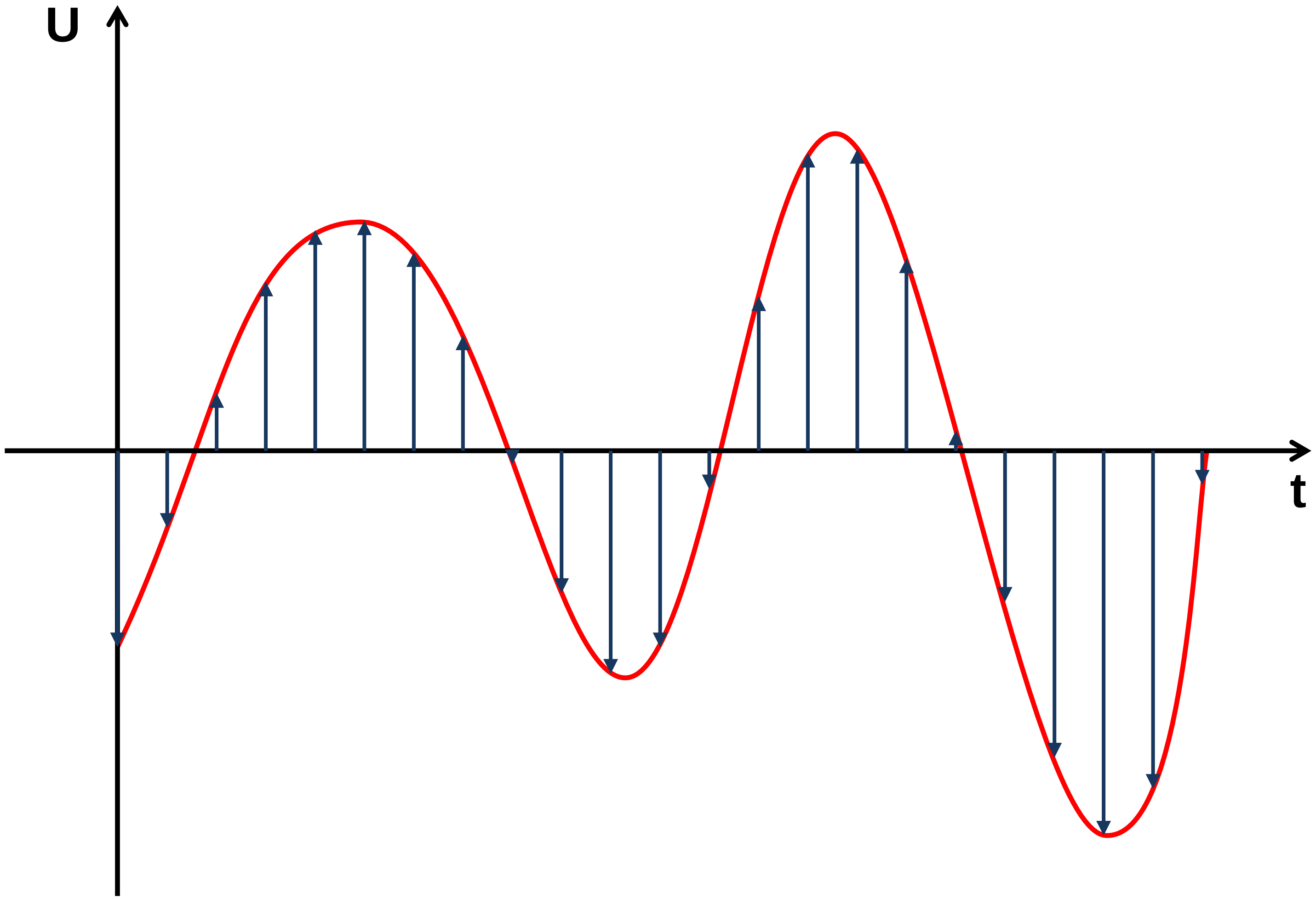
- Bandwidth (B) of the voice channel: 4 kHz
- Sampling frequency: 2*B = 2*4 kHz = 8 kHz
- Number of samples / second = 8000
- Sampling period: 1/2*B = 1/8000 Hz = 125 μs

Restoration of the sampling with bandpass filtering

Restoration of the sampling with bandpass filtering

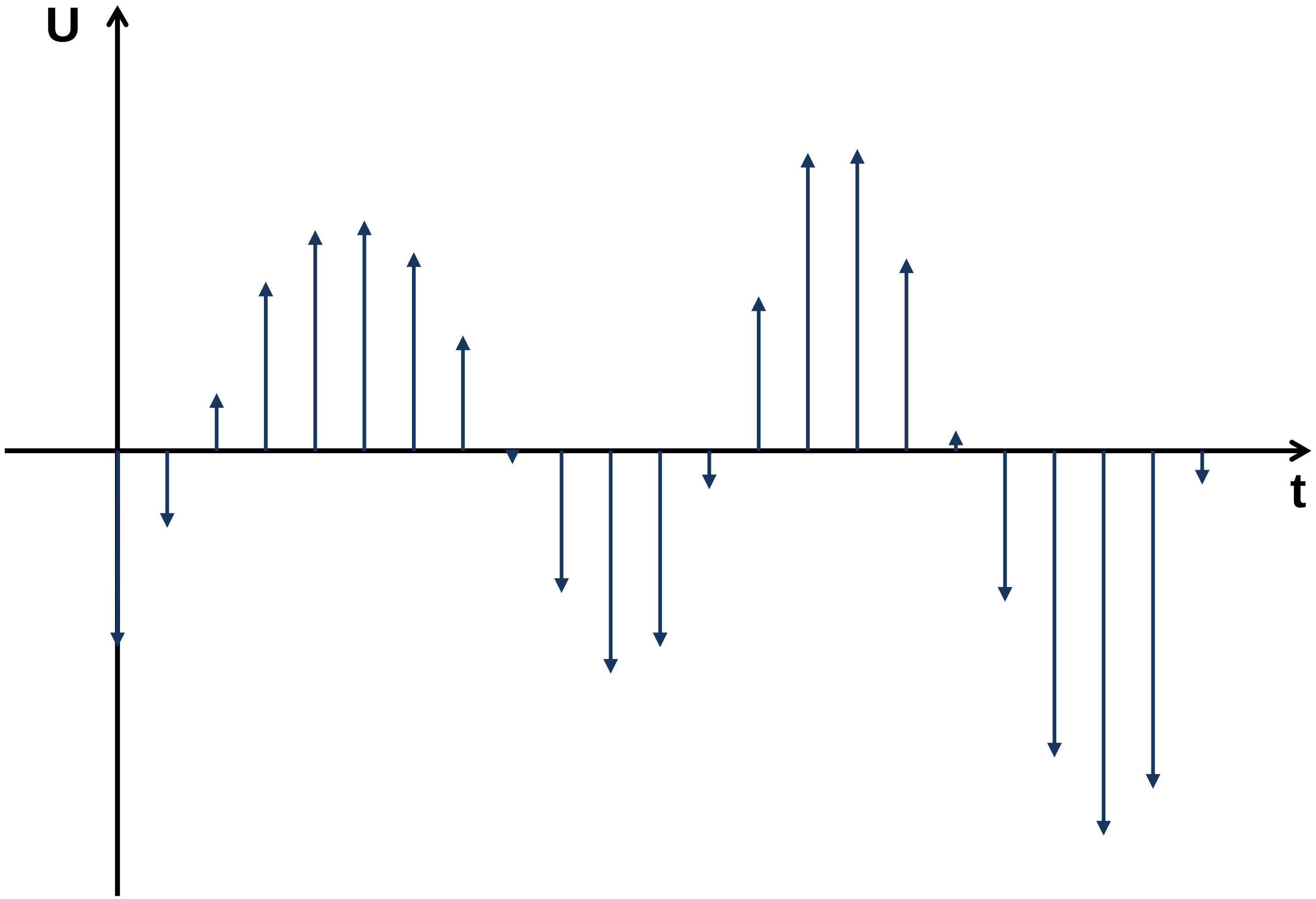


Restoration of the sampling with bandpass filtering



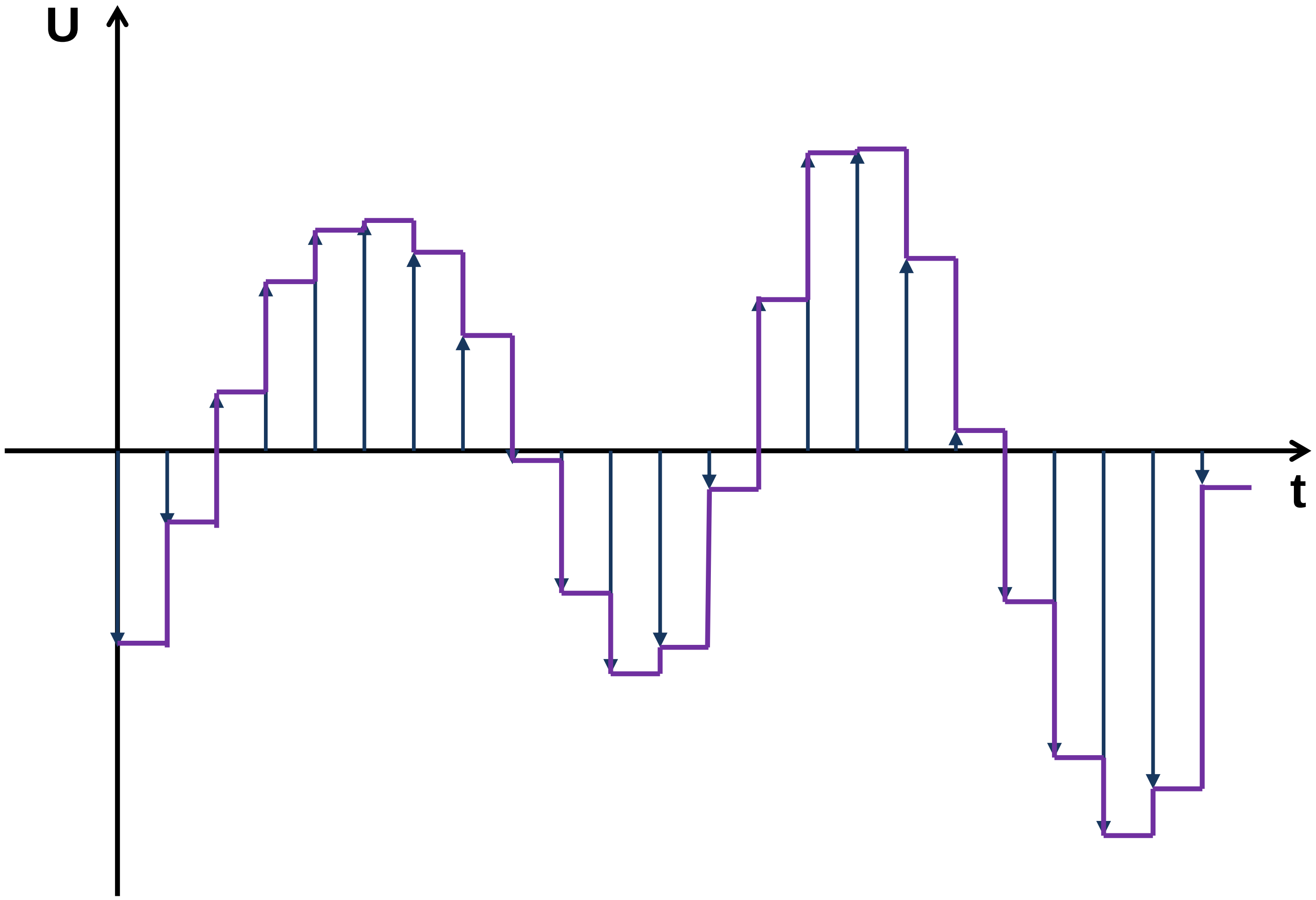
Sampling

Restoration of the sampling with bandpass filtering



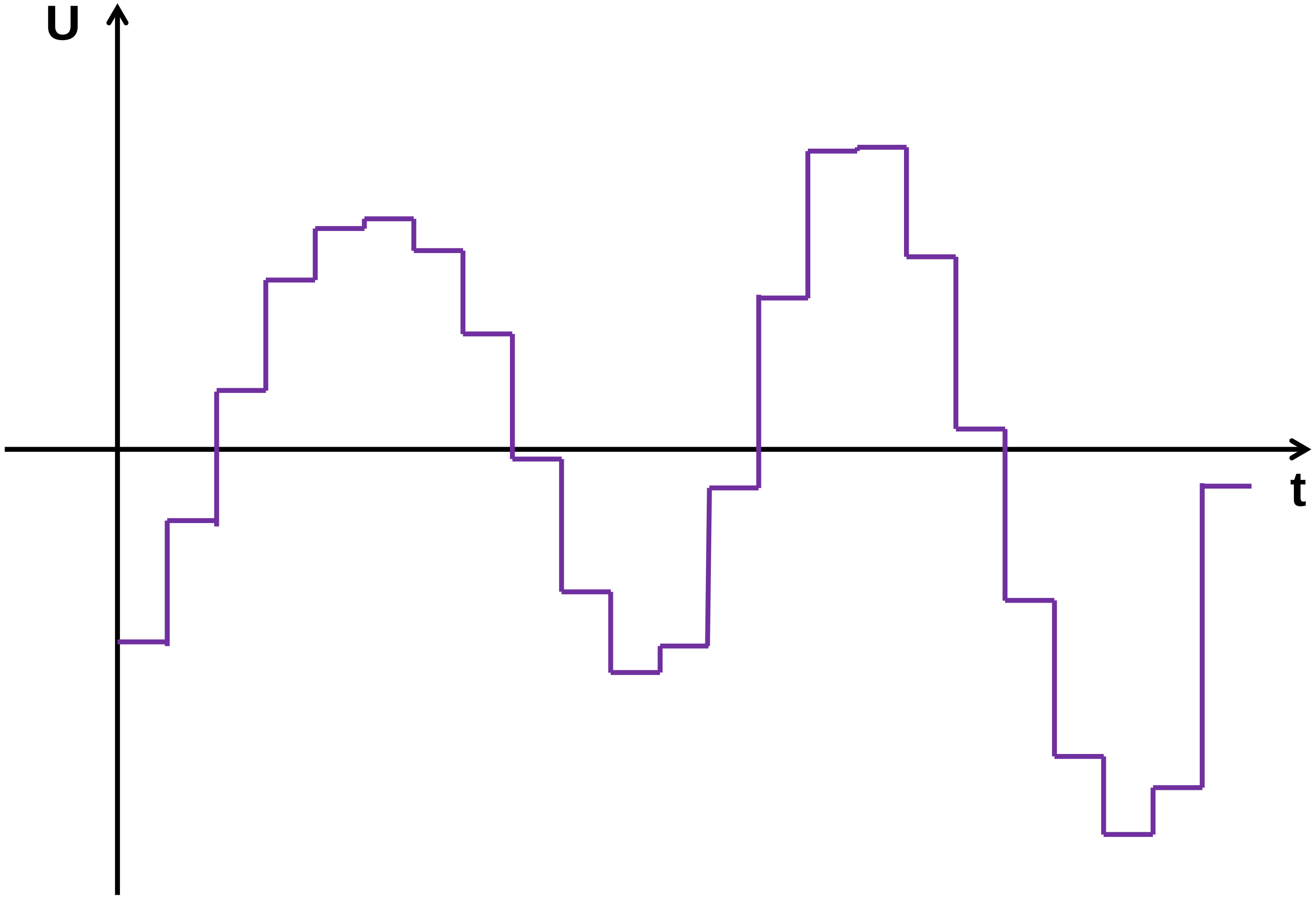
Sampling

Restoration of the sampling with bandpass filtering



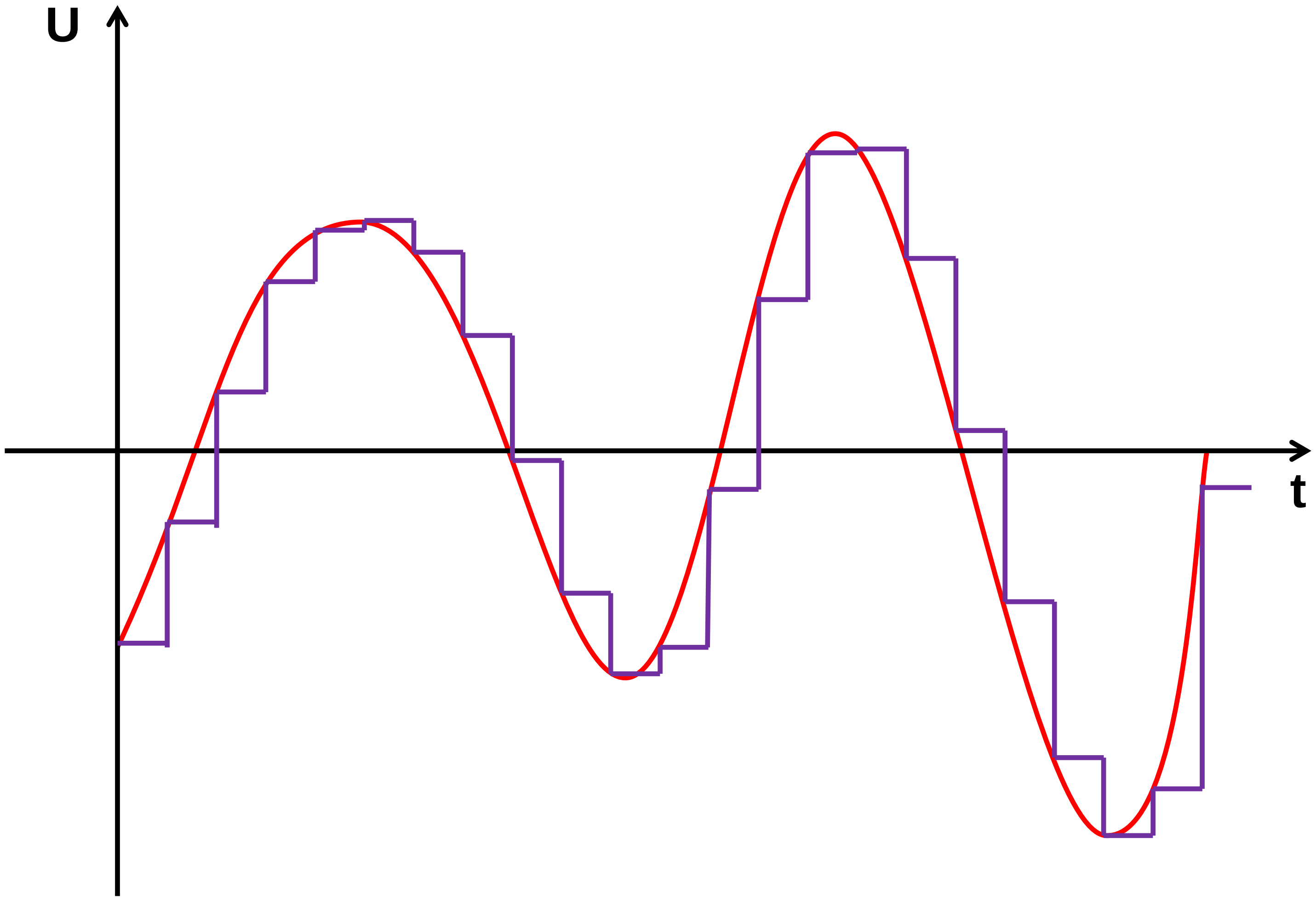
Zero order support (keep the value till the next sample)

Restoration of the sampling with bandpass filtering



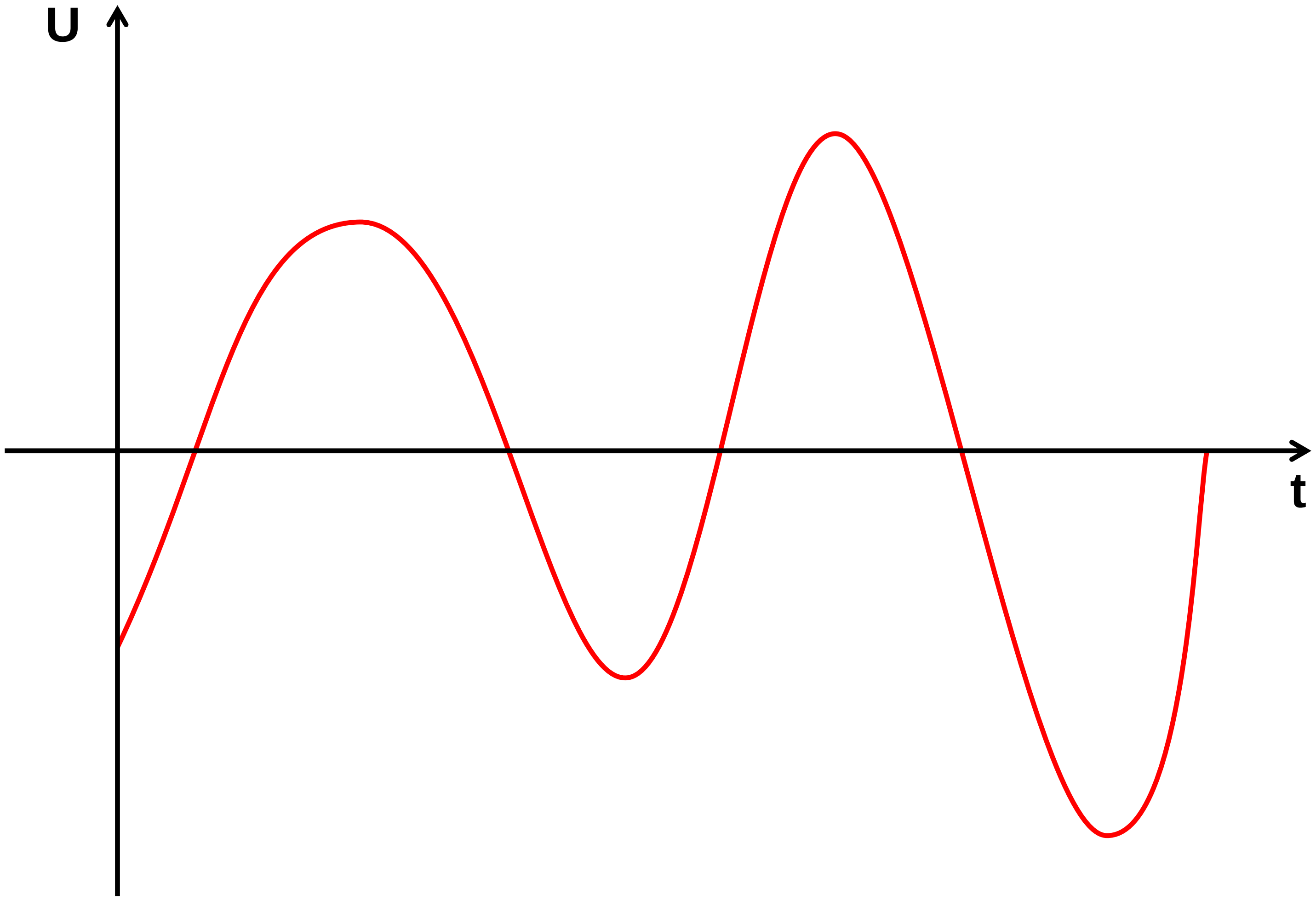
Zero order support (keep the value till the next sample)

Restoration of the sampling with bandpass filtering



Signal after bandpass filtering

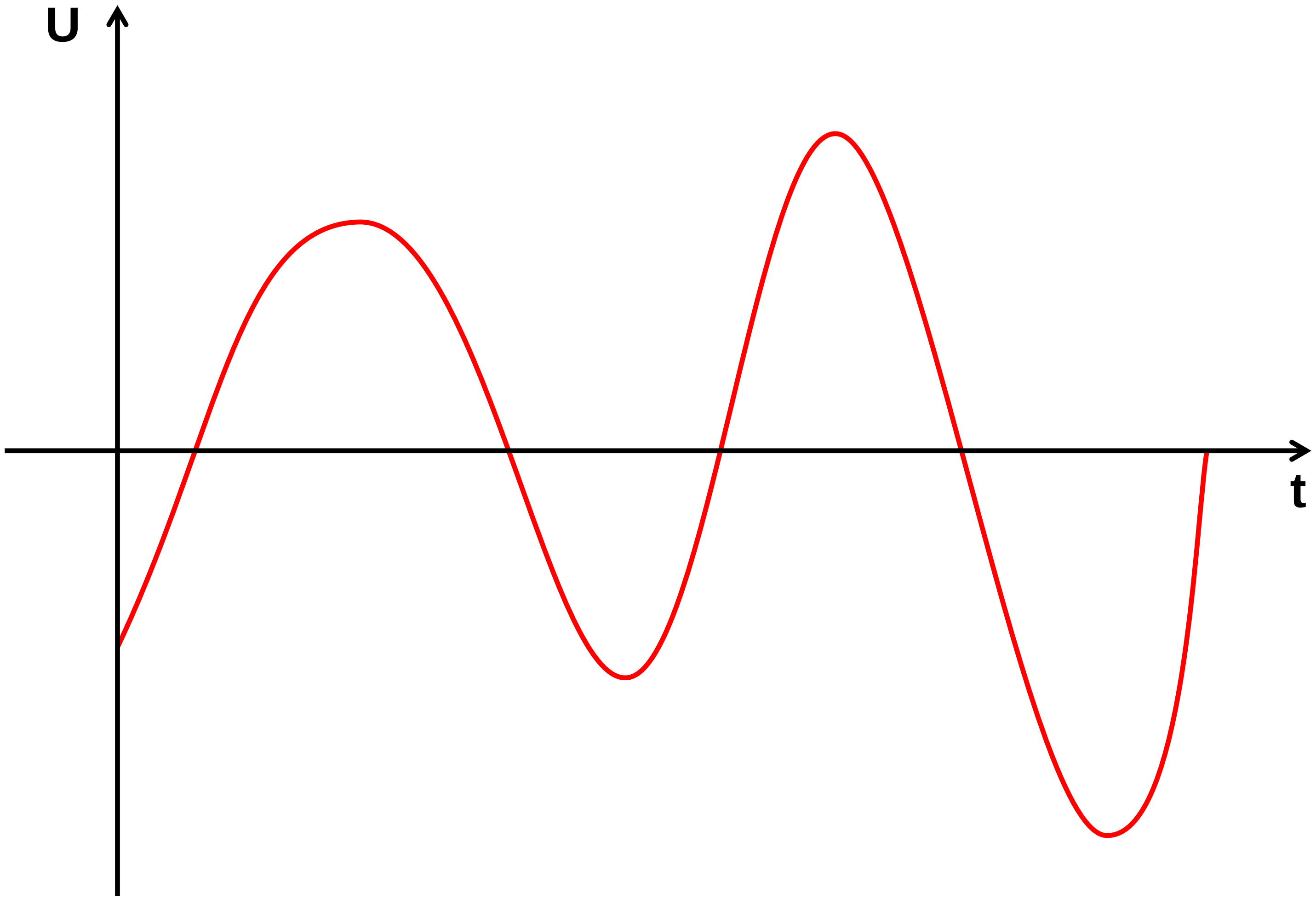
Restoration of the sampling with bandpass filtering



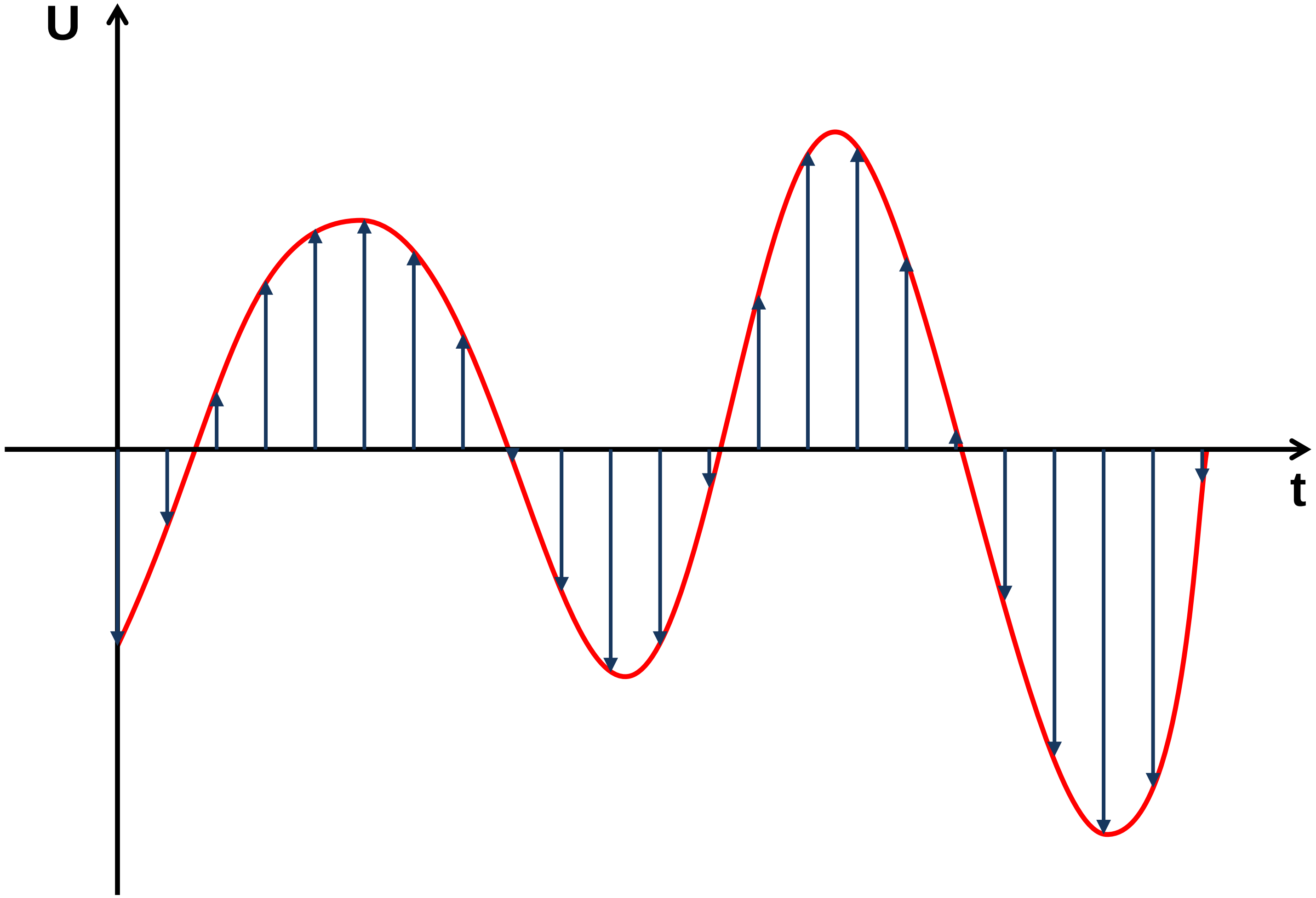
Signal after bandpass filtering

Example: quantizing

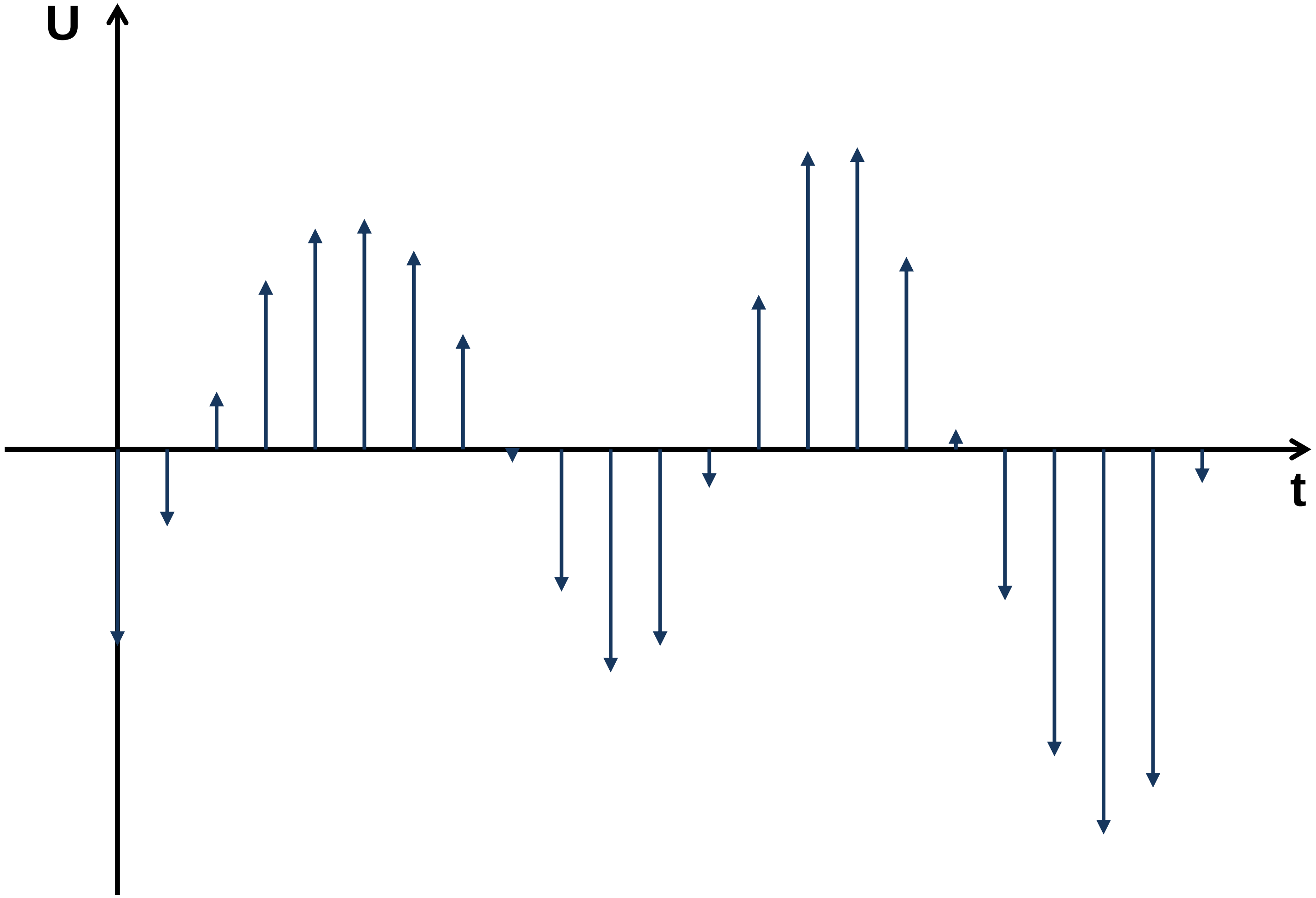
Example: quantizing



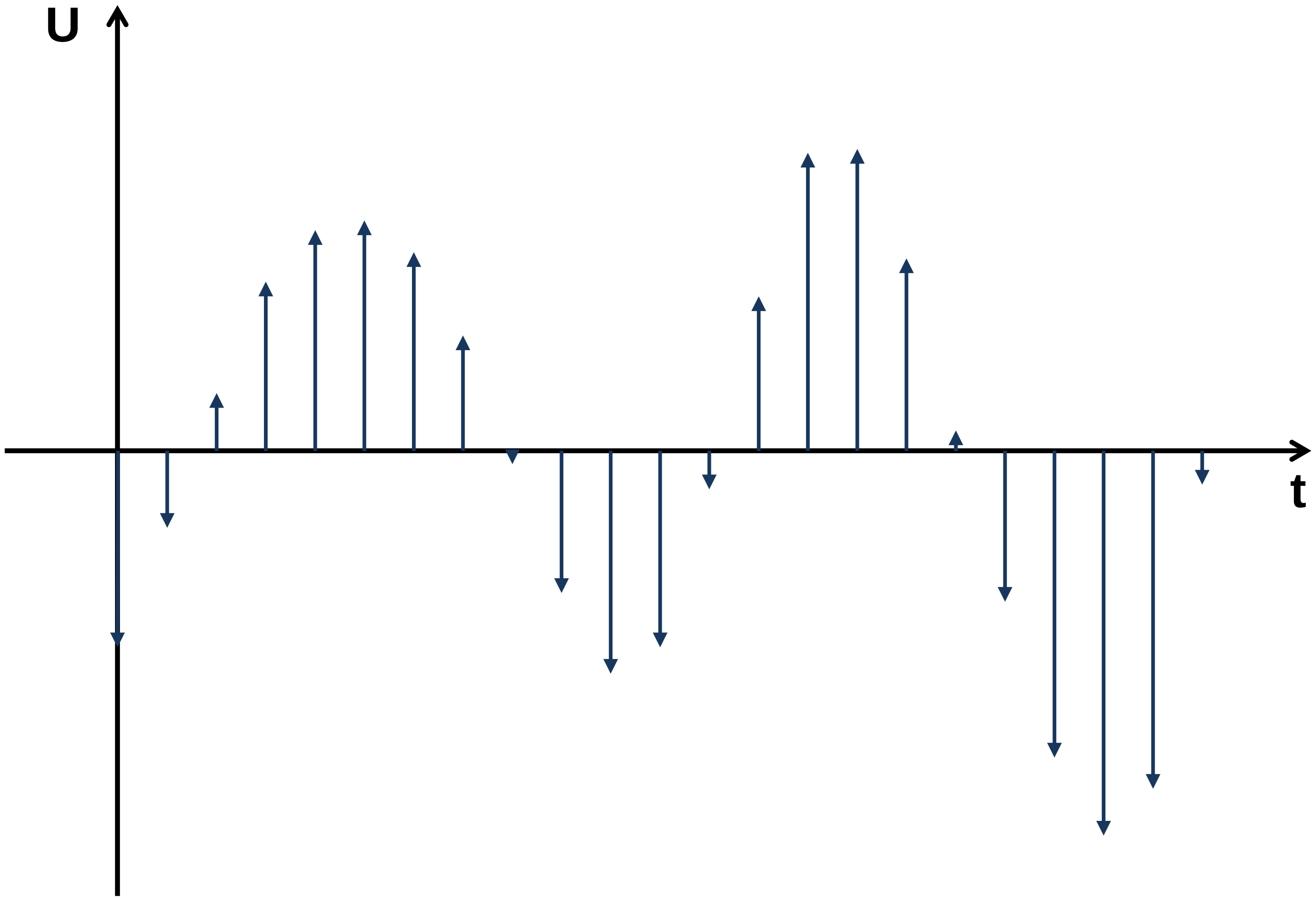
Example: quantizing



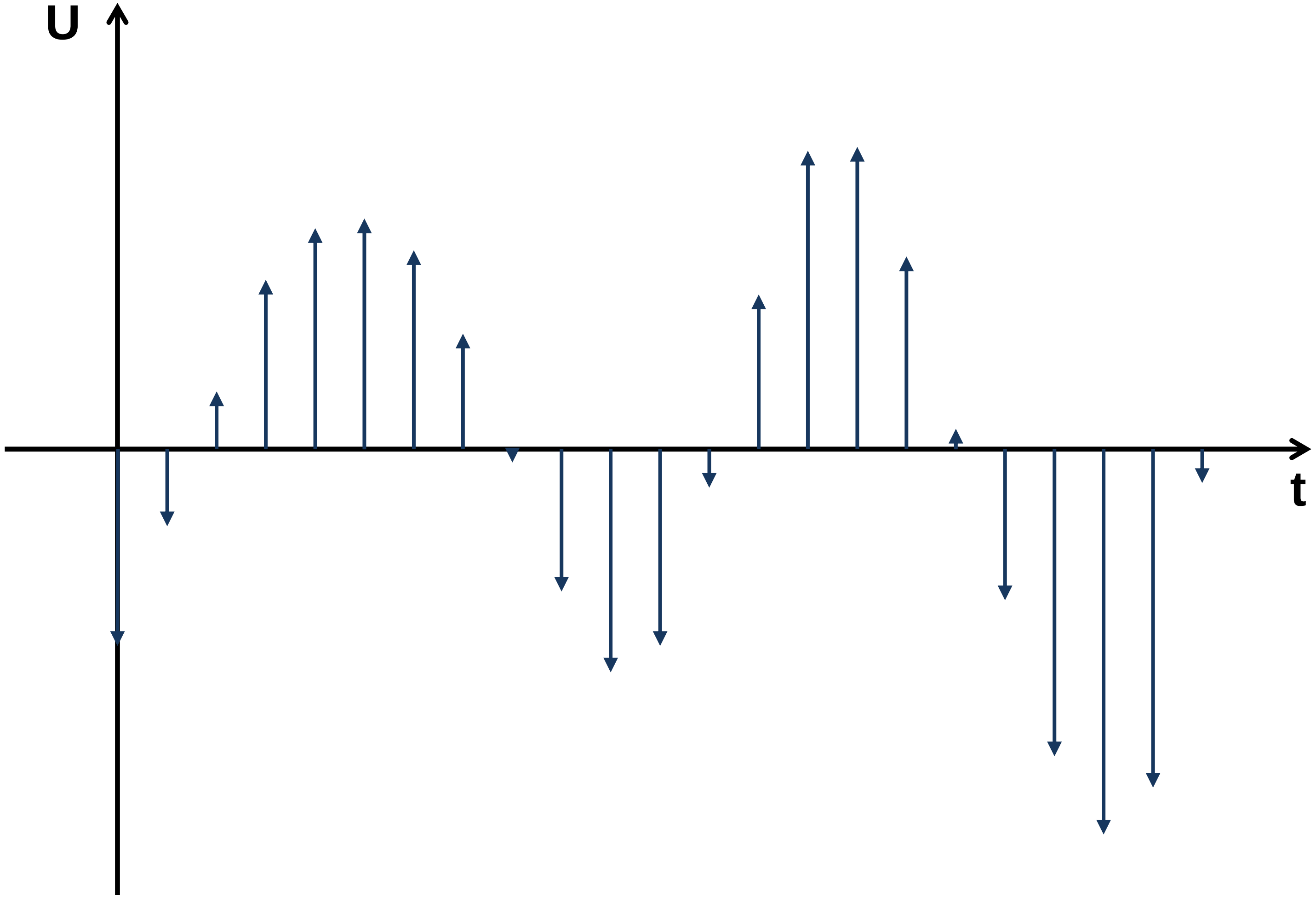
Example: quantizing



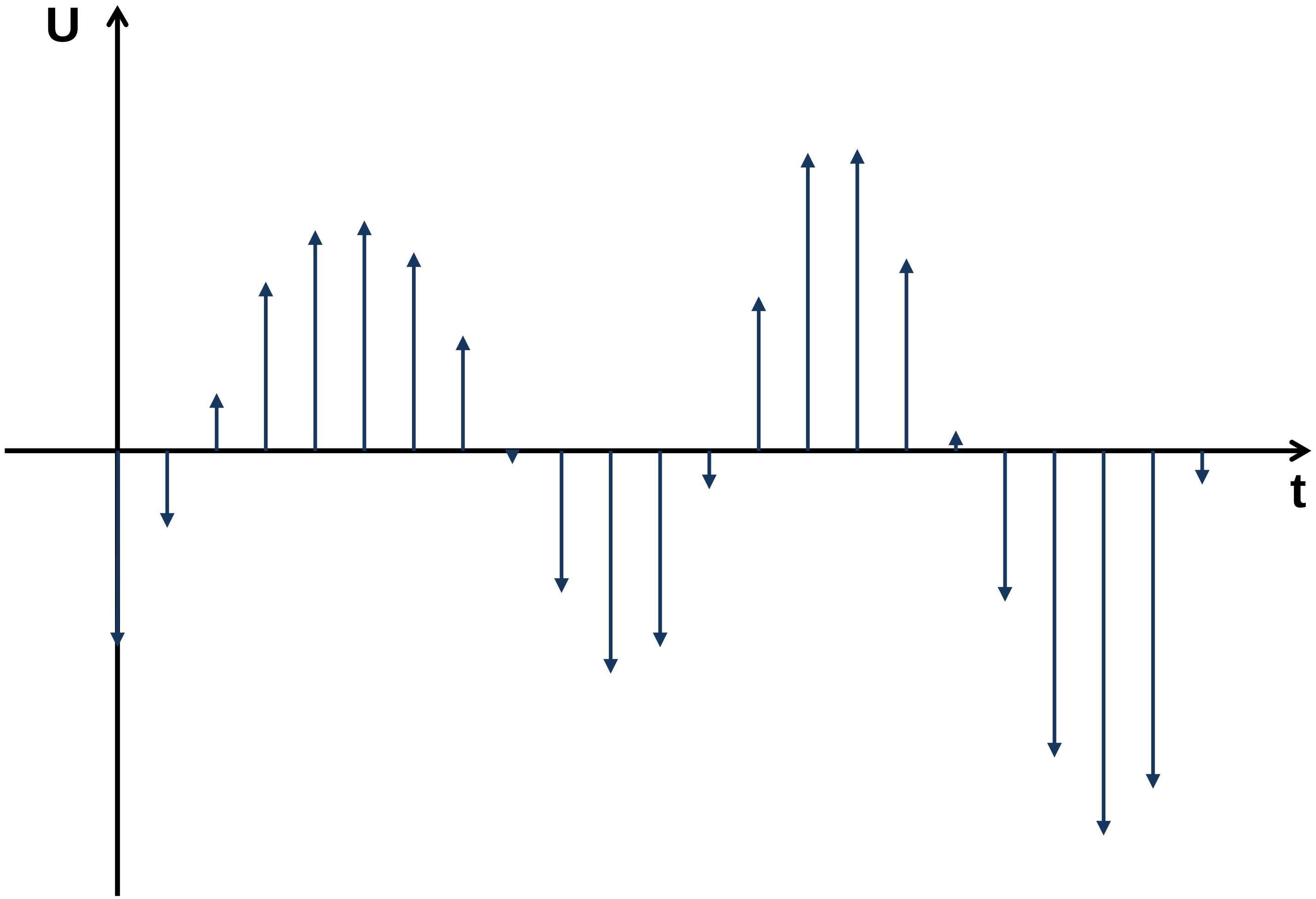
Example: quantizing



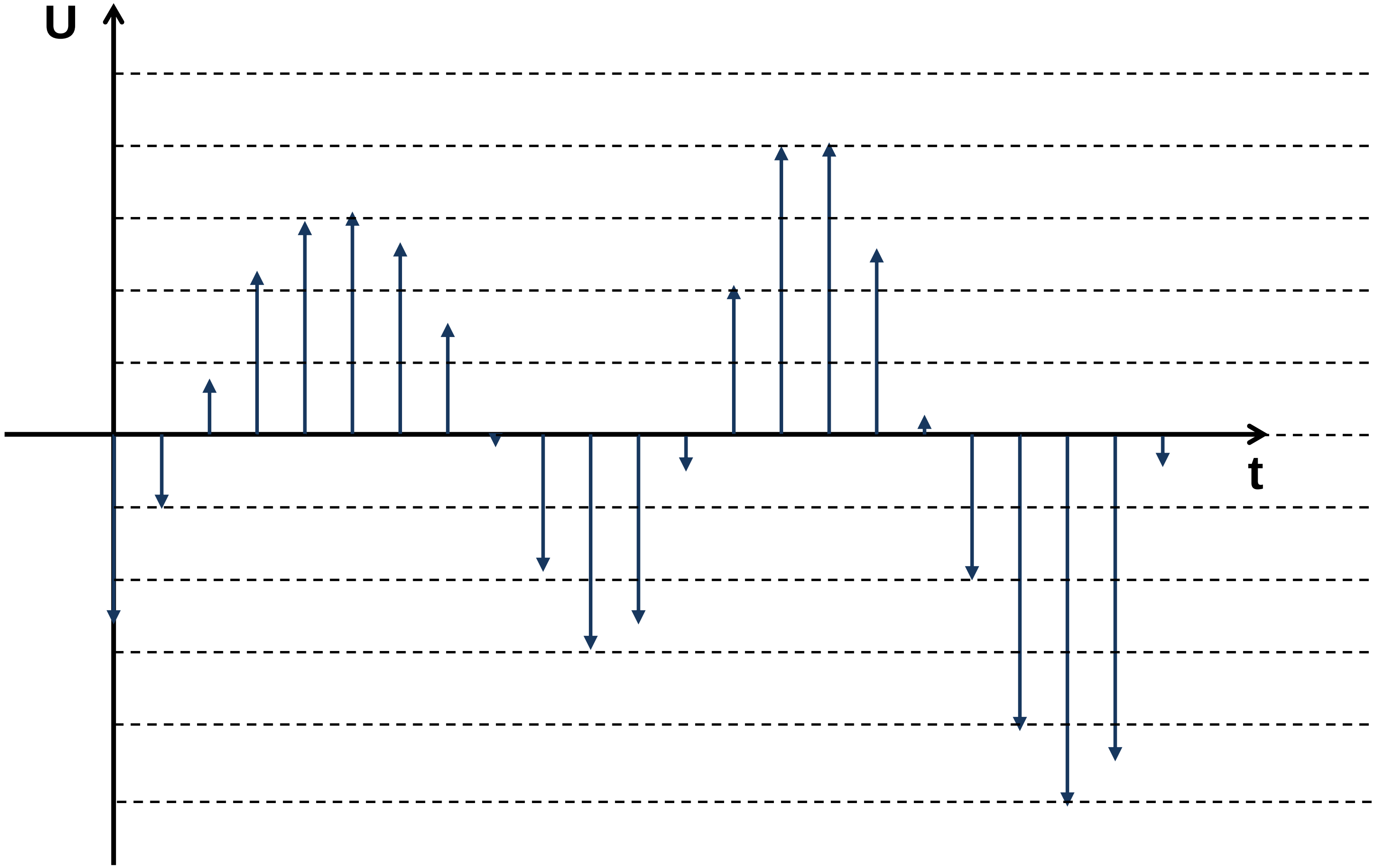
Example: quantizing



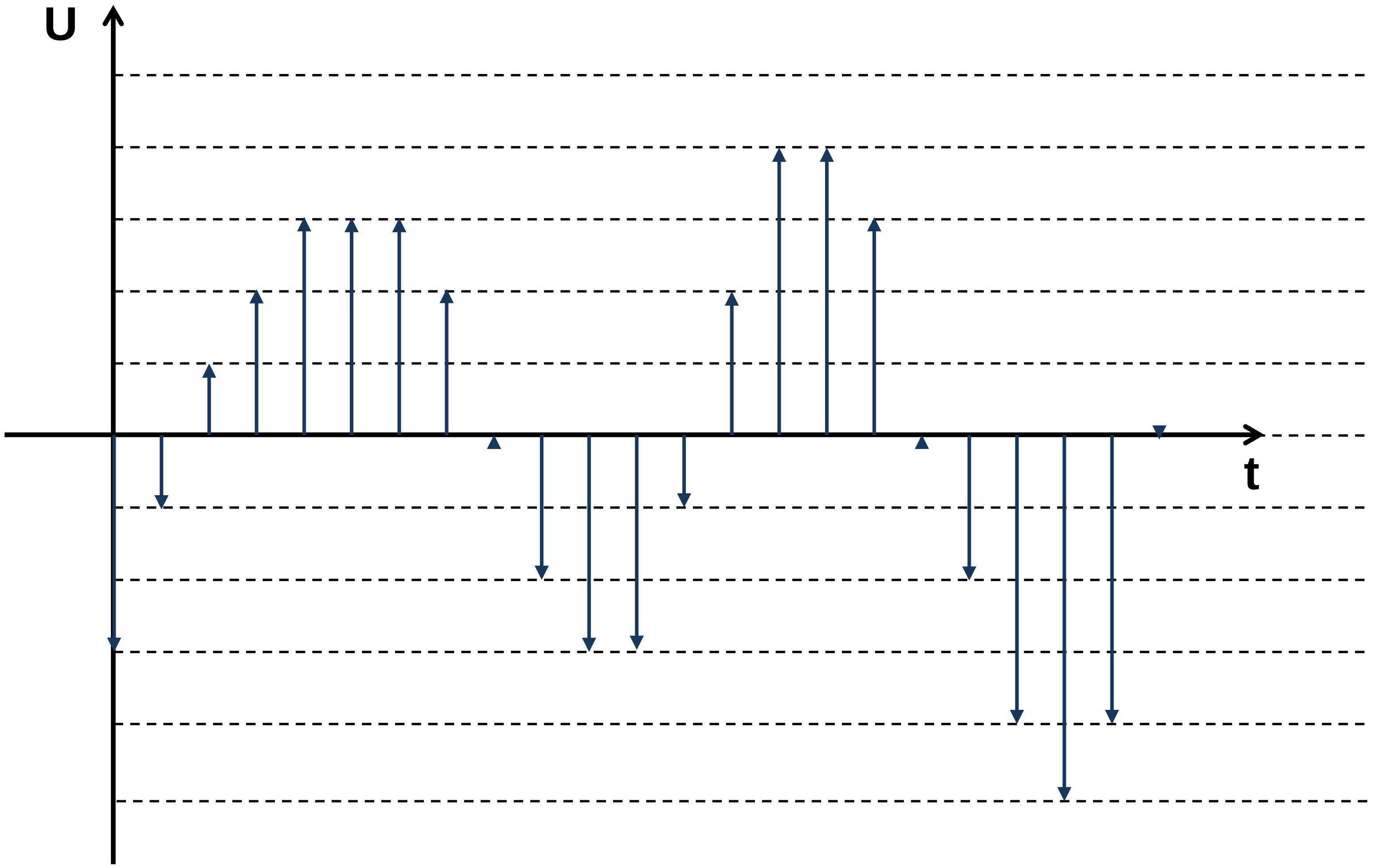
Example: quantizing



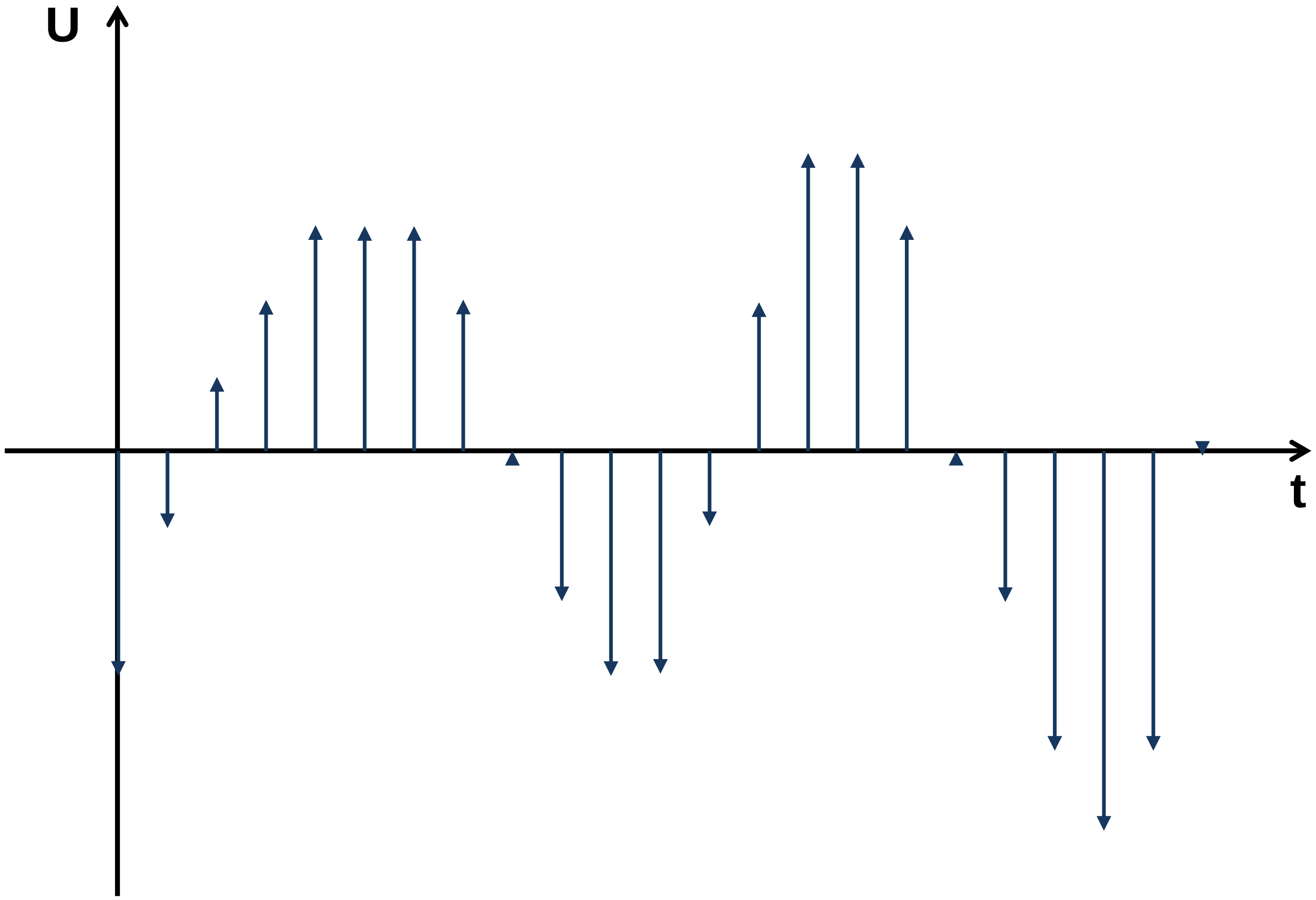
Example: quantizing



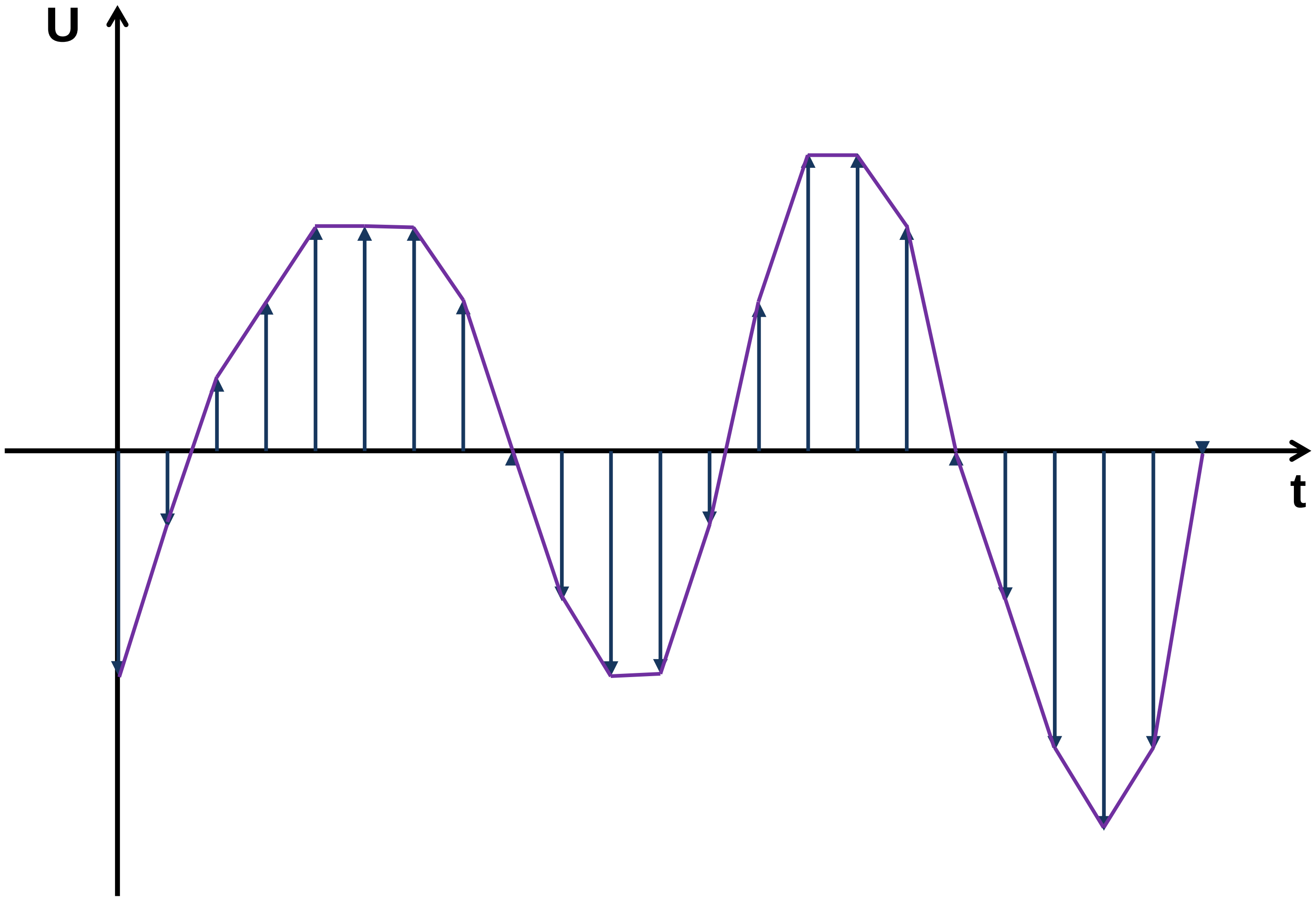
Example: quantizing



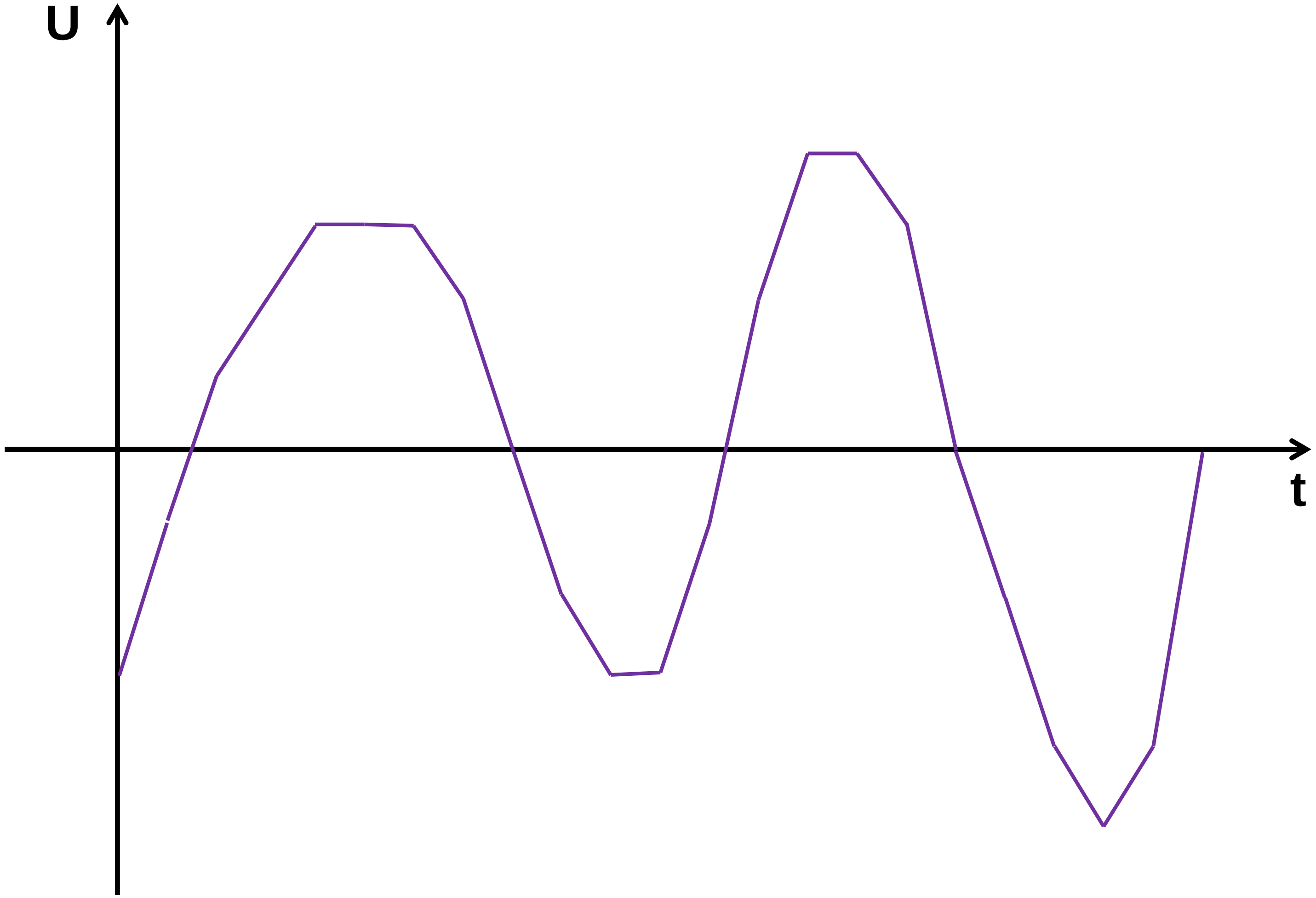
Example: quantizing



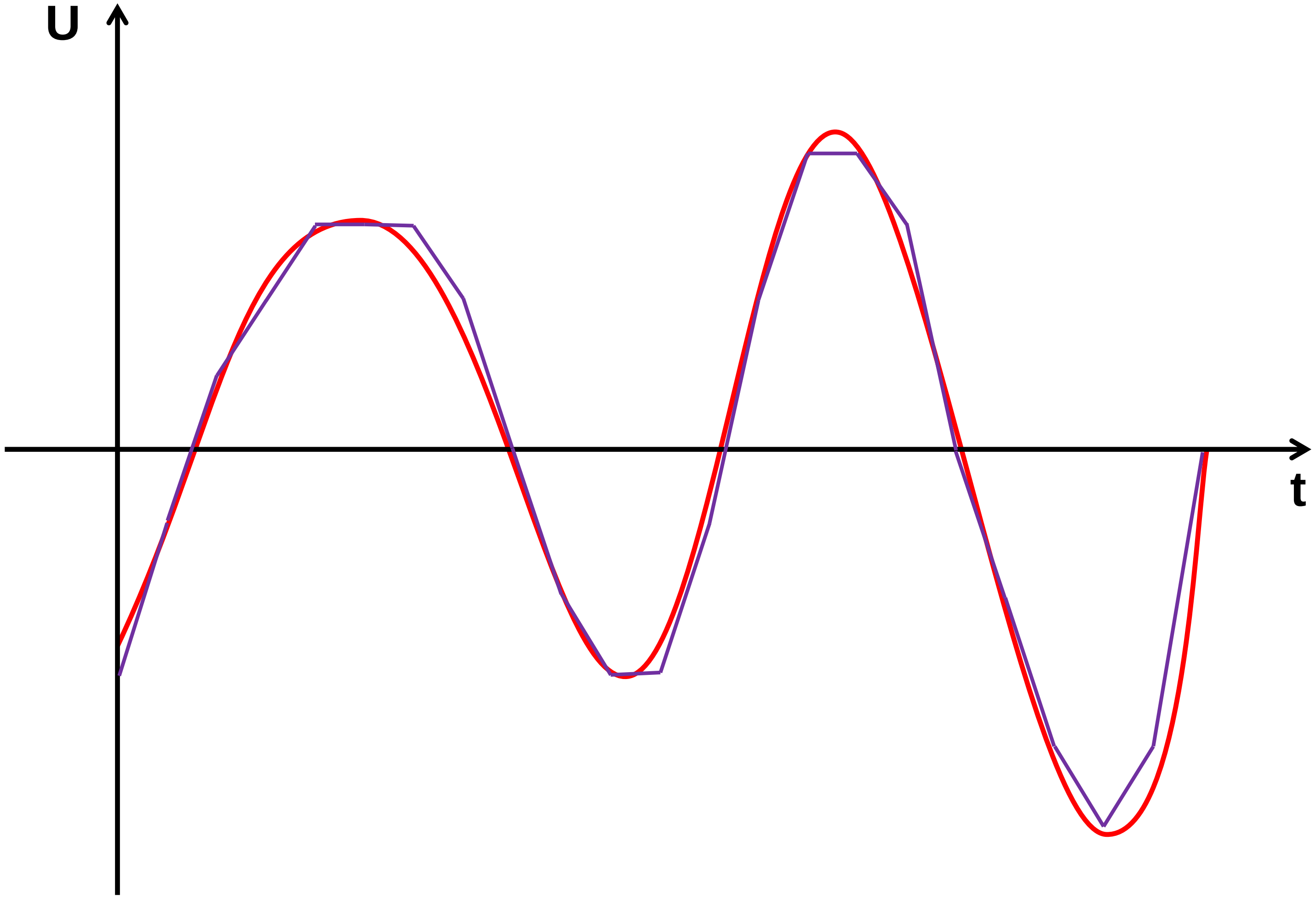
Example: quantizing



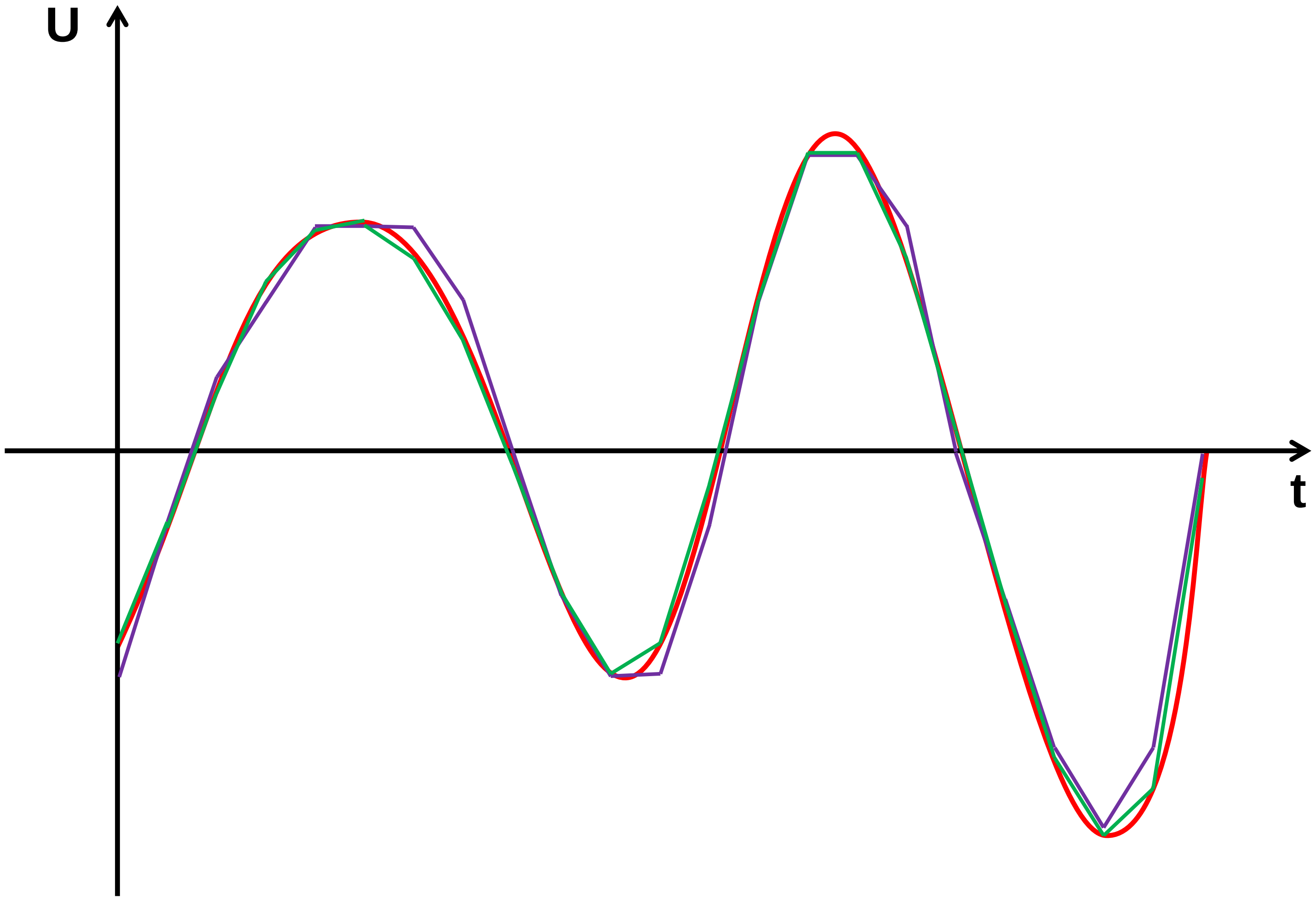
Example: quantizing



Example: quantizing



Example: quantizing



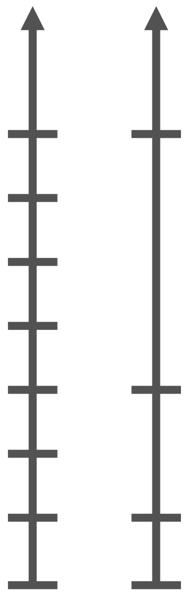
Red: original signal
Green: sampled signal
Purple: sampled and quantized signal

Quantization

- The more levels we quantize, the closer we get to the original signal
- Is there such a thing as *enough* quantization levels for *perfect* restoration?
- No :(
  - There always remain an error: the quantization distortion
    - less precisely: “quantization noise”
  - You will be engineers, some solution is needed!
  - Solution: no need to be perfect, just "good enough".”
    - A lot of quantization levels: storing longer "words" (more bits).
    - A few quantization levels: more distortion, worse quality
    - An application-dependent compromise is required

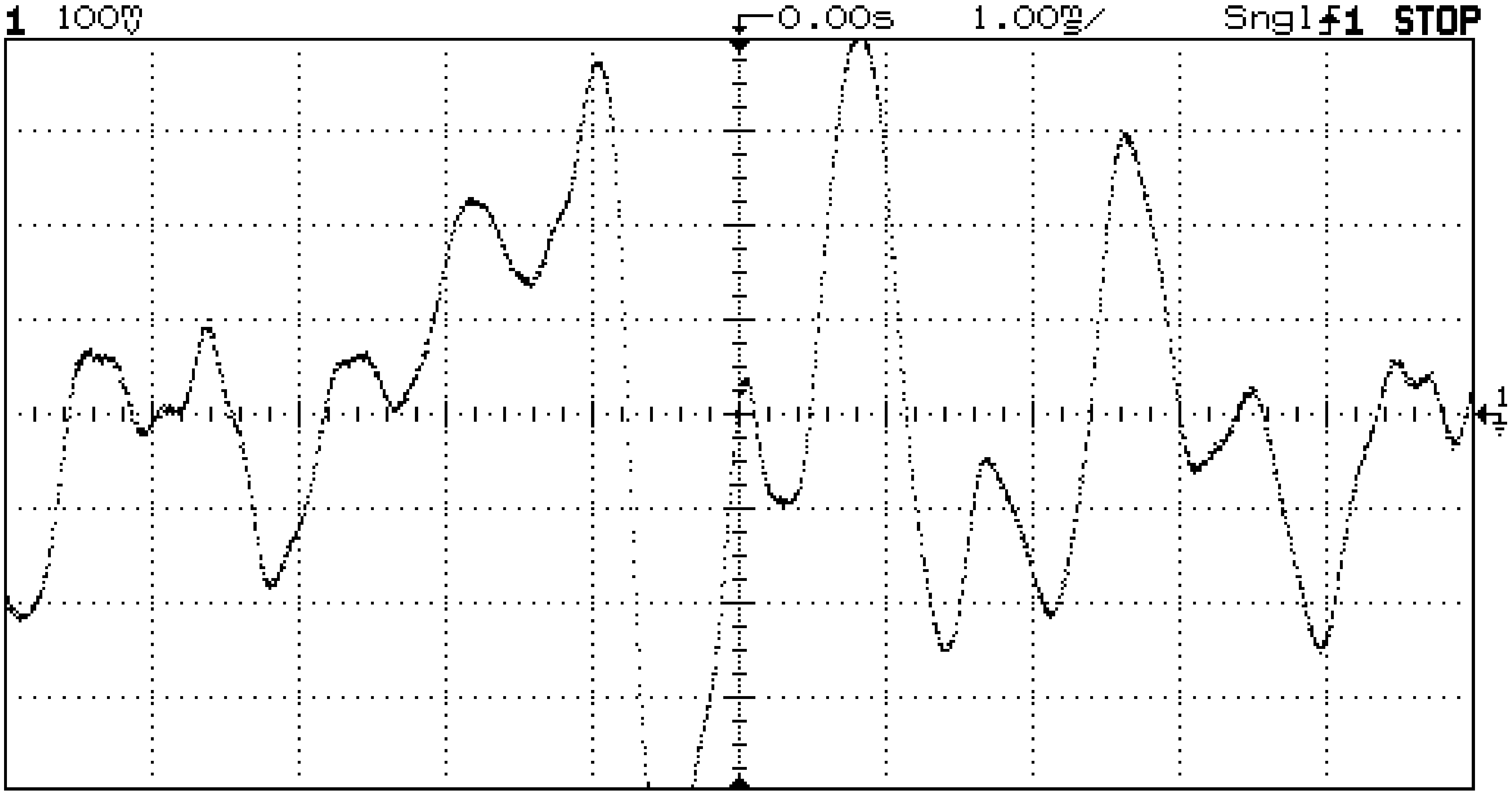
Adopting quantization to a voice signal

- A trick to achieve good results with reducing the quantization levels Voice signal:
  - frequent consonants (m, p, t, etc.) with low amplitude
  - significantly affect voice understandability
  - => at low amplitude, denser quantization levels needed
- For the human ear signal/noise ratio is important
  - the louder the signal, the larger noise it can suppress
- Solution:
  - *Non-linear quantization*
  - With (quasi-)logarithmic characteristics (the human ear is such)
    - At low inputs: denser, at high inputs sparser
    - USA: μ-law quantizer
    - Europe: A-law quantizer
    - Similar, but not compatible, code conversation is required



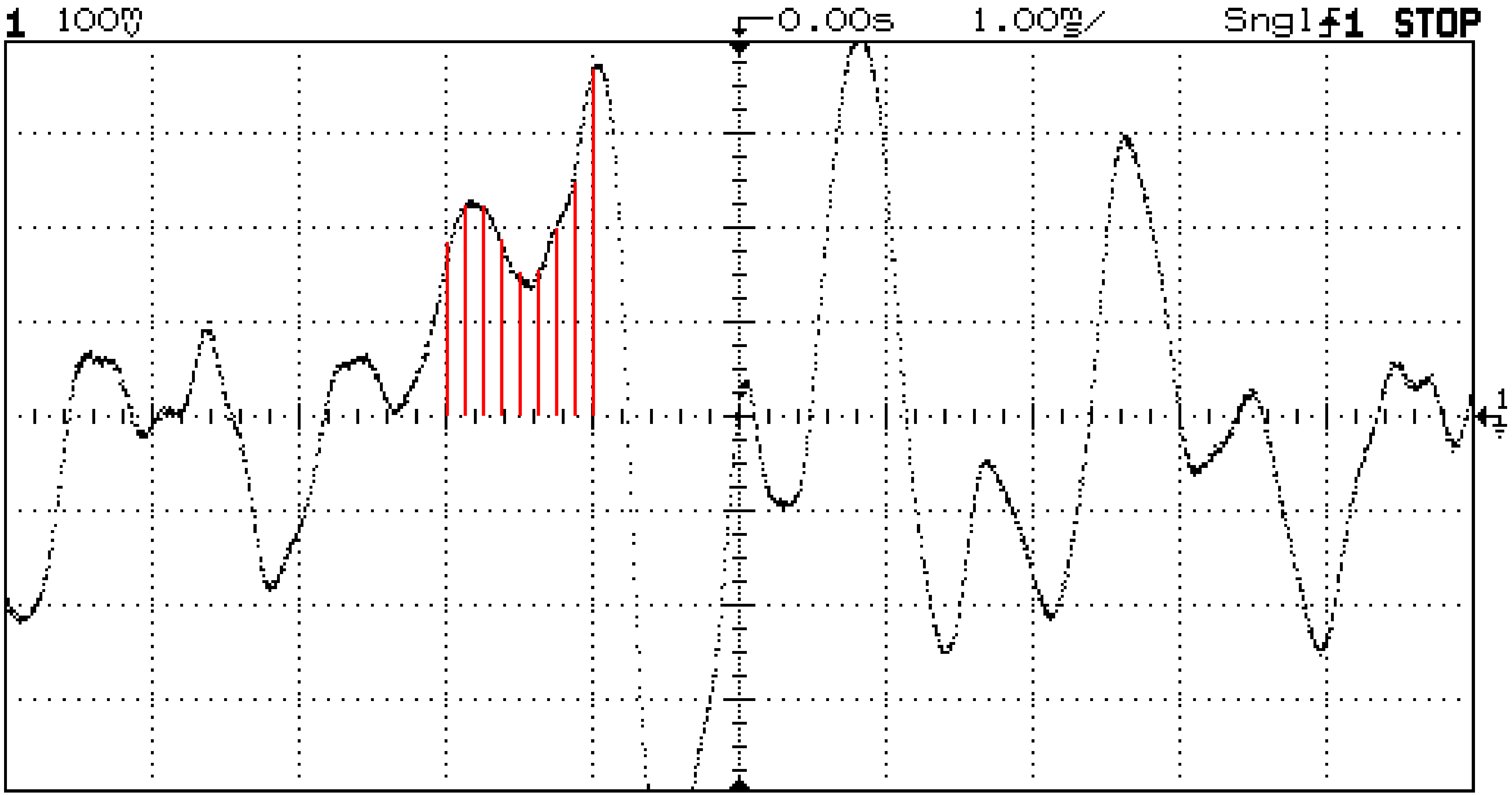
PCM codec through a real example

Human voice signal



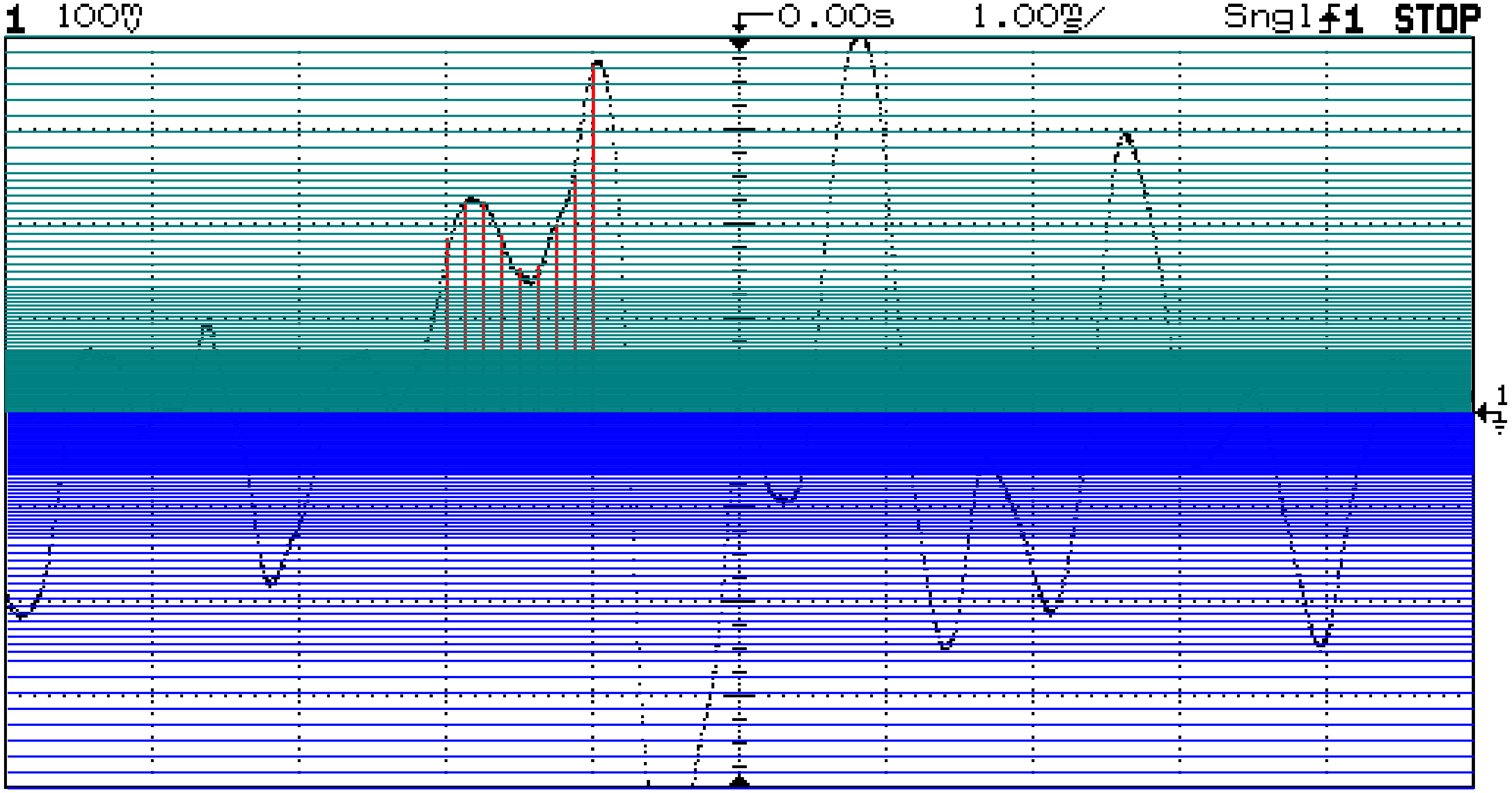
PCM codec through a real example

Human voice signal



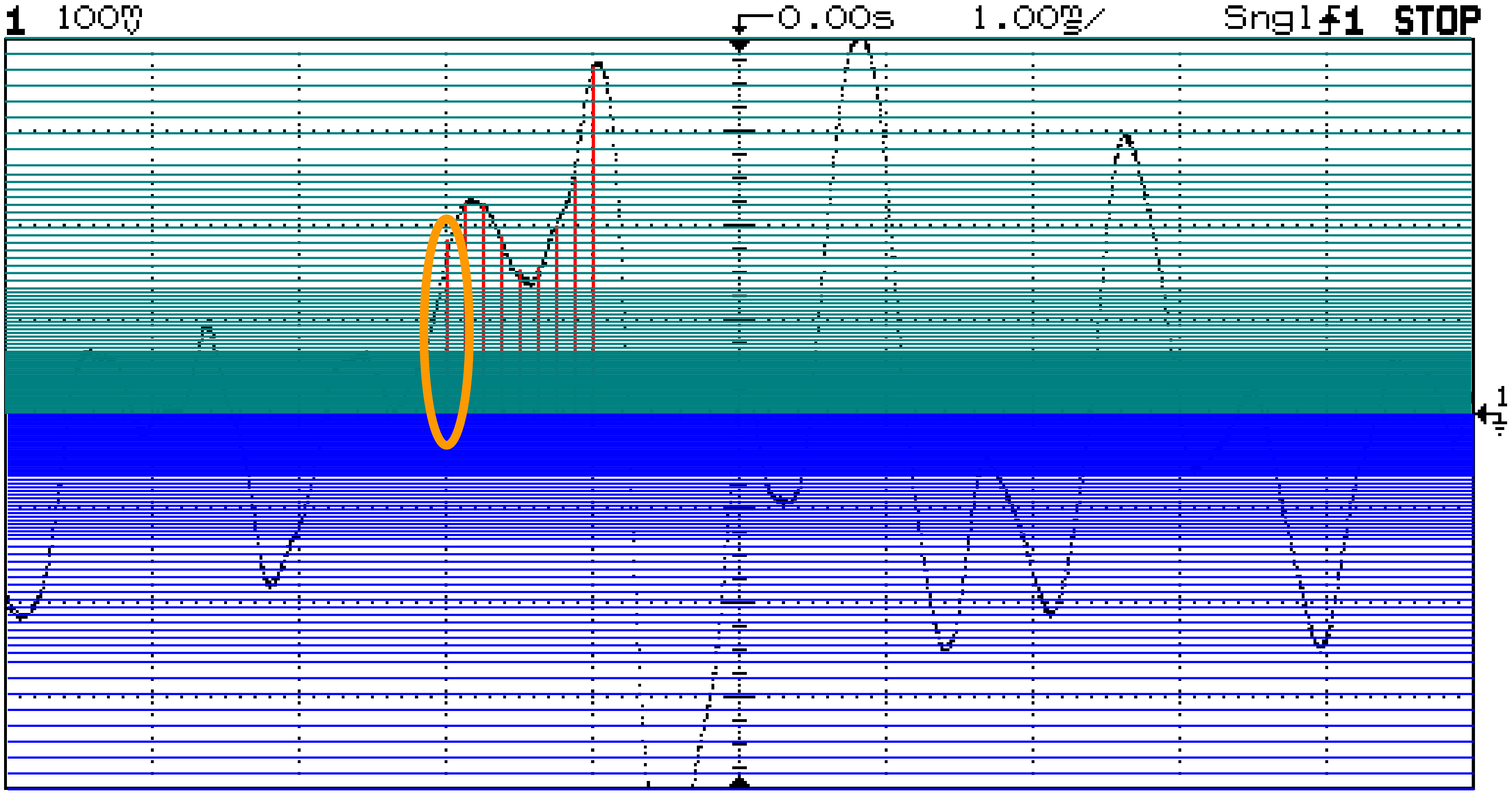
PCM codec through a real example

Human voice signal



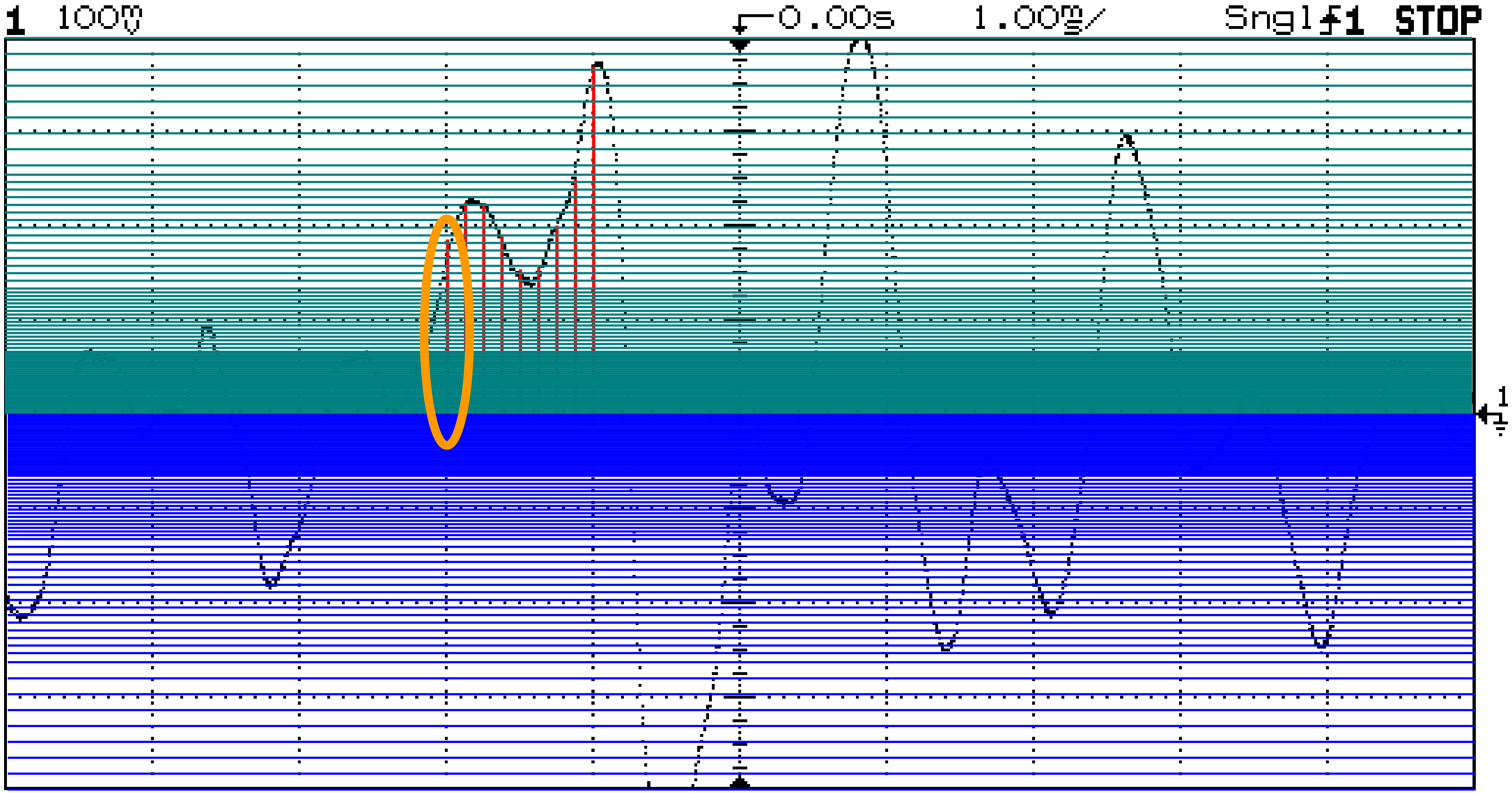
PCM codec through a real example

Human voice signal



PCM codec through a real example

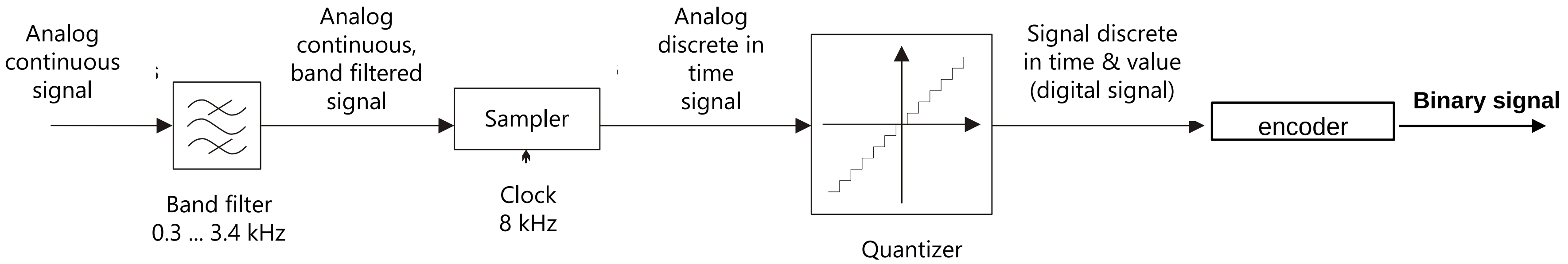
Human voice signal



Related bit sequence (appr. 180 mV): 10100110 (MSB is the sign – 1 positive, 0 negative)

PCM

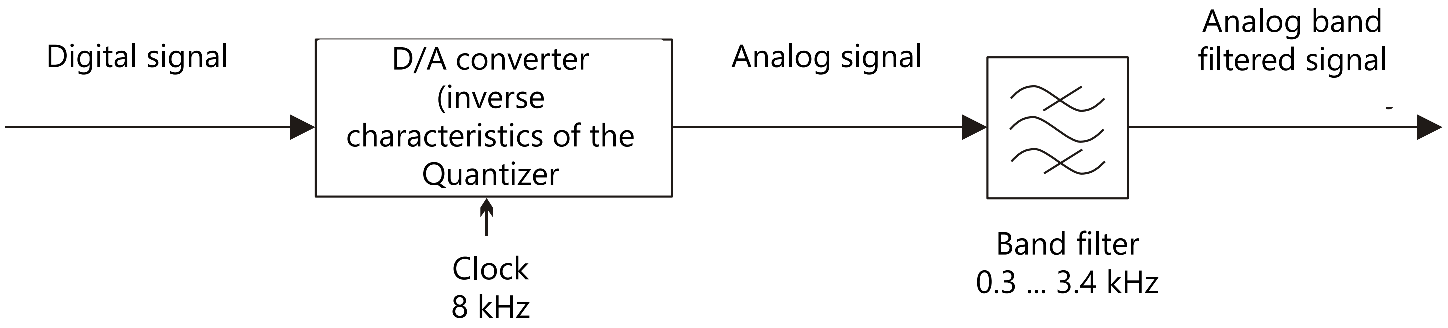
- A/D conversion (PCM encoding):



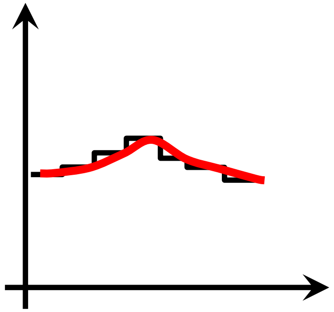
- Band filter
  - Frequencies above the band would appear as in-band noise after sampling
  - Based on the previous: 300-3400 Hz are let through
- Sampler
  - 8 kHz is the sampling frequency (= we take 8000 samples per second)
  - Double of the max. signal ftequency (Nyquist-Shannon theorem)
- Quantization
  - We quantize to 256 levels, this gives an acceptable result
  - with a (quasi-)logarithmic characteristic
  - 256 levels a=> 8 bit a sample
- 8000 sample /sec • 8 bit/sample = ***64 kb/s***

PCM

- D-A conversion (PCM decoding):



  – Role of Band Filter:
    smoothing the output:



PCM: CD

- PCM is also used on Audio CDs
- Sampling: 44.100 Hz
- Quantization: 16 bit (65.536 levels)
- Stereo: 2 independent channels
- Bitspeed:
  - 44.100 samples/s/channel * 16 bit/sample * 2 channel =
    = 1.411.200 bit/s ≈ 1,4 Mb/s ≈ 176,4 kB/s

Refresh

- Voice bandwidth 300-3400 Hz
  - Understandability (quality) <-> Economics
  - FDM – Gap -> Voice channel 4 kHz
- Analog vs Digital signals
  - How it is interpreted
    - Basically, all analog
    - Digital: if only at predefined points in time matters

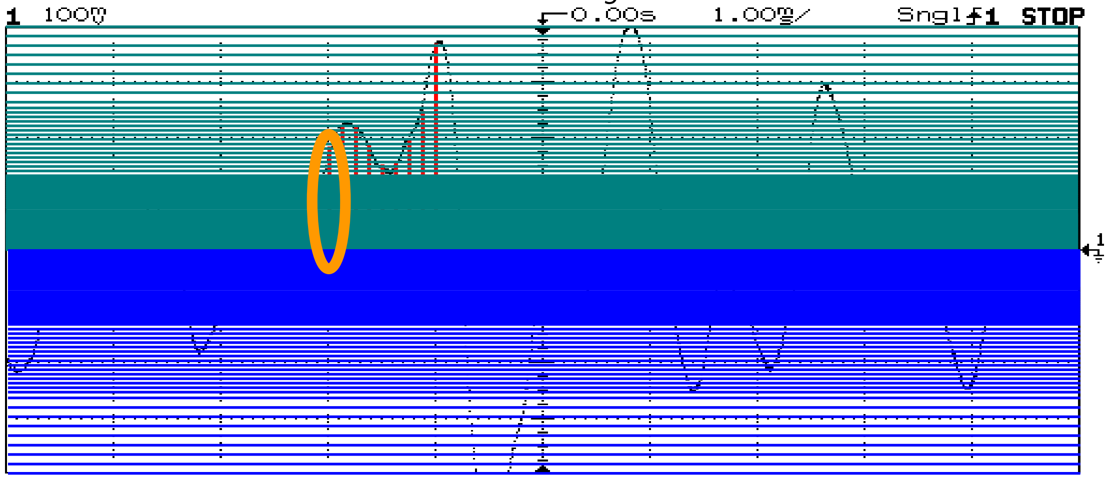
Refresh

- Voice Digitalisation
  - Band filtering
  - Sampling
    - Nyquist-Shannon: $T_{sampling}$ = 1/2B – 1/(2*4000Hz) = 125 μs
  - Quantisation
    - Quantisation distortion -> always info loss
    - Linear / Logarithmical
  - Encoding
  - Line Encoding

Refresh

- Voice Restoration
  - D-A conversion with the inverse characteristic of quantization
  - Band filtering (smoothing)
- PCM
  - μ-law/A-law
  - 8000 sample /sec • 8 bit/sample = ***64 kb/s***
- Digital
  - Digitalisation: loss (quantisation)
  - Transmission: Amplification -> Repeater -> can restore orig. signal -> “filters the noise”
    - (Analog: enlarges the noise)

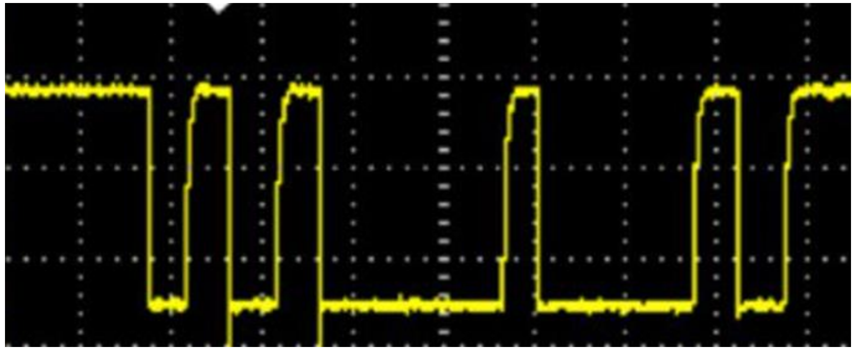
Human voice signal



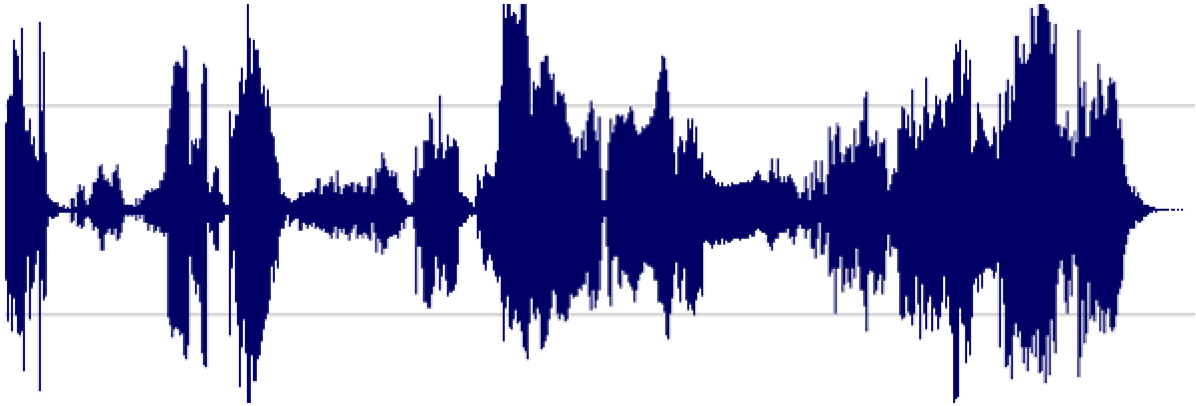
Related bit sequence (appr. 180 mV): 10100110
(MSB is the sign – 1 positive, 0 negative)

Voice transmission

1. Analog voice transmission
2. Analog & digital signals
3. Voice digitalization
4. Codecs

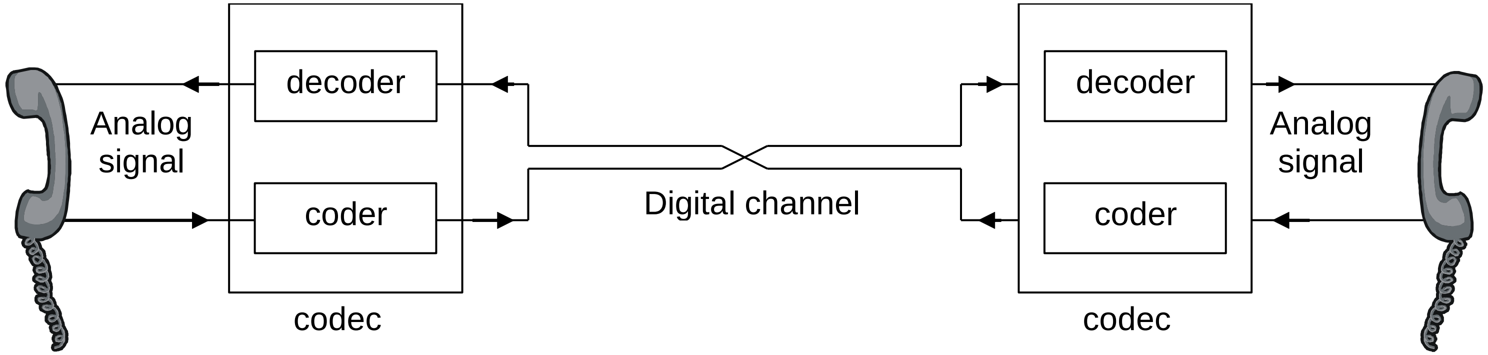


Additional voice encoders



Additional voice encoders

- Digitalization of voice: codec (COder, DECoder)



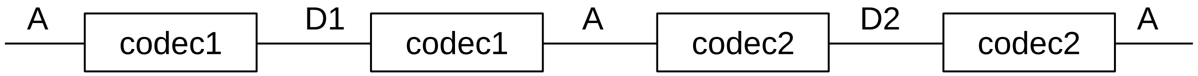
  - In general, a codec can be any coder – decoder, e.g. for film (mpeg4), music (mp3)
  - We only deal with voice encoders now
- The same encoder must be used on both sides, or intra-network conversion needed
  - More codec types
    - Negotiation needed
  - Conversion
    - Radio/Cable
    - PCM/VoIP

Codec features

- Bitspeed / Bitrate
  - 2,4 -- 64 kb/s
  - Smaller the better <-> quality
- Quality of voice signal
  - difficult to measure objectively
  - MOS (Mean Opinion Score):
    - 15-40 people score several samples, the whole is averaged
    - 1: unacceptable, 2: weak, 3: average, 4: good, 5: perfect
    - Above 4: considered very good
- Encoding delay
  - the larger time slice we process at the same time, the better we can compress - at the cost of a greater delay
  - 0.125 – 80 ms
- Complexity
  - it was especially important for older mobile devices
  - unit of measure: MIPS (Million Instructions Per Second)

Codec features

- Robustness
  - in case of error, there is no time to retransmit
  - the error rate of the radio transmission appr. 10^{-3}
  - FEC (Forward Error Correction)
    - Introducing redundant data; error correcting code / parity
- Tandemizability and recoding
  - with itself or with other codecs one after the other

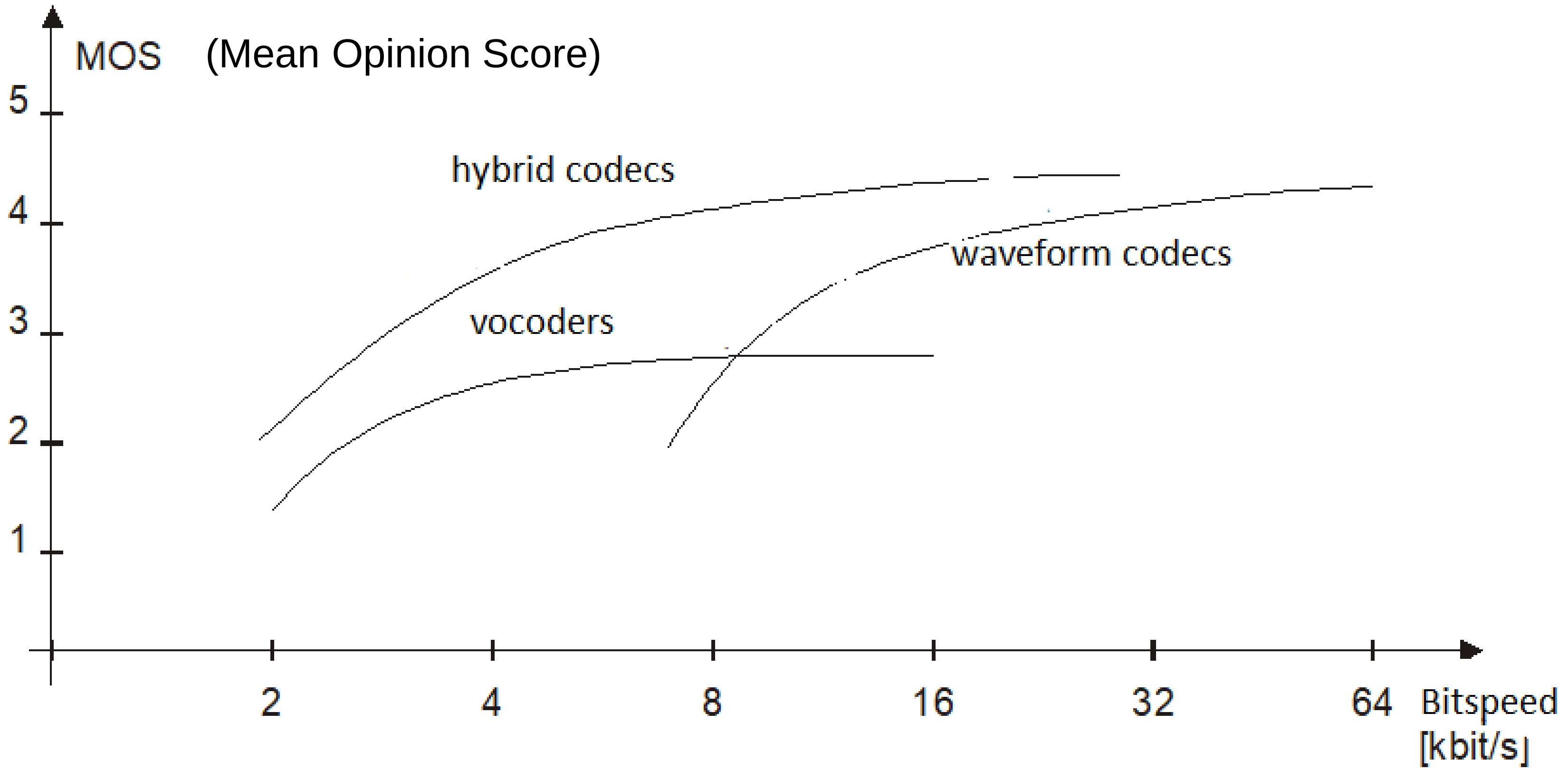


  - how does it tolerate?
- Transparency
  - Can other than voice sound (between 300-3400 Hz) be transmitted?
  - For example, DTMF (Dual Tone MultiFrequency) is a sinusoid signal, but it is needed for the operation of call centers (e.g. telebank). The question is whether it passes through the codec with a sufficiently small distortion to remain decodable on the other side.
- Adaptivity
  - lower signal speed in case of heavy traffic
    - but worse quality

Codec types

- Waveform encoder
  - preserving the shape of an analog signal
  - good quality
  - high bitrate (this is a disadvantage, not an advantage! :))
  - transparency
- Vocoder
  - on the transmitter side: filtering out typical parameters from speech
  - on the receiving side: speech synthesis based on these
  - lower bitrate
  - voice not very similar to the original
- Hybrid codec
  - a mixture of the above

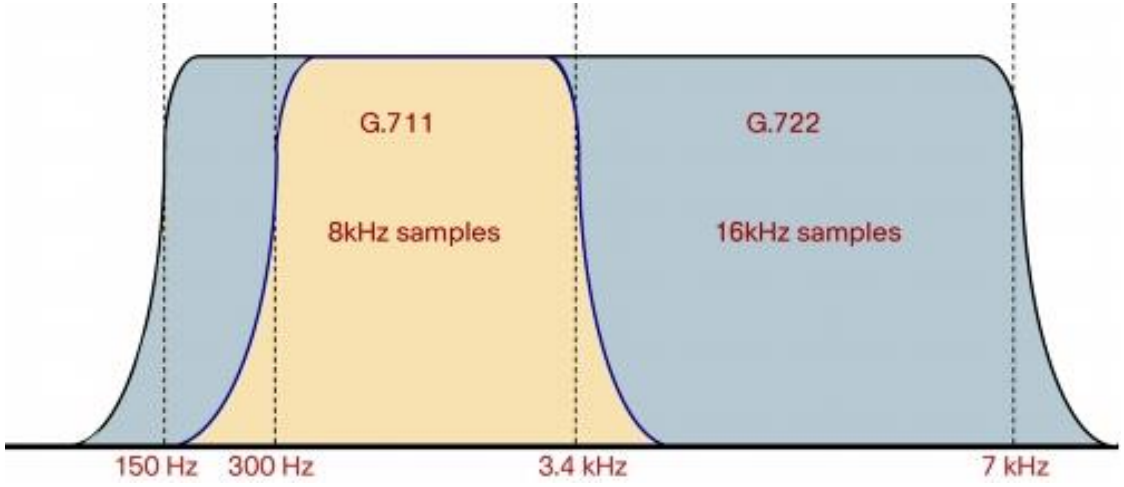
Codec types



Log. scale!

AMR-WB (G.722.2)

- AMR-WB: Adaptive Multirate, Wideband
- Marketing name: „HD Voice"
- 10+ -year-old standard, it was introduced in practice in the last few years
- Greater transmitted (analog) bandwidth: approx. 150-7000 Hz
- 16 kHz sampling + other fixes on the encoder
- Result: higher bit rate, but better sound quality



http://www.voipsupply.com/cisco-hd-voice

Codec Types

The framed part is the exam material
Numerical values only need to be known in order of magnitude

Name of Encoder	Standard	Main application	Introduc-tion	Bitrate (kb/s)	Speech quality (MOS)	Encoding delay (ms)	Calculation complexity (MIPS)
PCM	G.711	Fixed telephony, ISDN	1972	64	4,5	0,125	0,52
ADPCM	G.721/ G.726	Fixed telephony, ISDN	1984* / 1990	16/24/32*/ 40	4,1*	0,125	7,2
(GSM) FR	GSM 06.10	GSM	1989	13	3,7	20	4,5
(GSM) HR	GSM 06.20	GSM	1994	5,6	3,5	24,4	17,5
(GSM) EFR	GSM 06.60	GSM	1995	13	4,0	20	14,4
AMR	GSM 06.90	3G mobile	1998	4,75-12,2	3,5-4,0	20	15-25
AMR-WB**	G722.2	3G mobile	2004	6,6-23,85	3,6-4,2	25	20-30
G.723.1	G.723.1	VoIP	1996	6,3 5,3	3,9 3,6	30 30	15 20
G.729	G.729	VoIP	1996	8	4,0	15	11
LPC-10	LPC-10	military	1976	2,4	2,3	≥ 22,5	7

FR: Full Rate
HR: Half Rate
EFR: Enhanced Full Rate
AMR(-WB): Adaptive Multirate (-Wideband),