Digital Audio

Sampling, quantization, and digital representation of sound

Lorenz Schwarz
Karlsruhe University of Arts and Design (HfG)

Winter Semester 2024/25

Course info

← Chapters · Download PDF ↓

DIGITAL REPRESENTATION OF SOUND

Sound (physical domain)

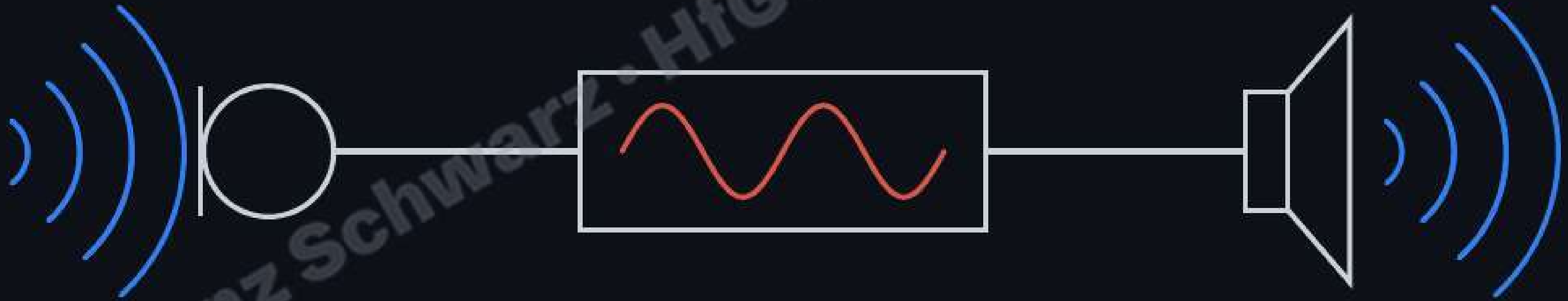
Sound is a physical phenomenon consisting of pressure variations in a medium (e.g. air) over time.

- Exists as acoustic energy
- Described by sound pressure, particle velocity, and intensity
- Continuous in time and amplitude

Transduction of sound

A transducer (e.g. a microphone) converts sound pressure into a corresponding electrical voltage.

- Energy changes from acoustic to electrical

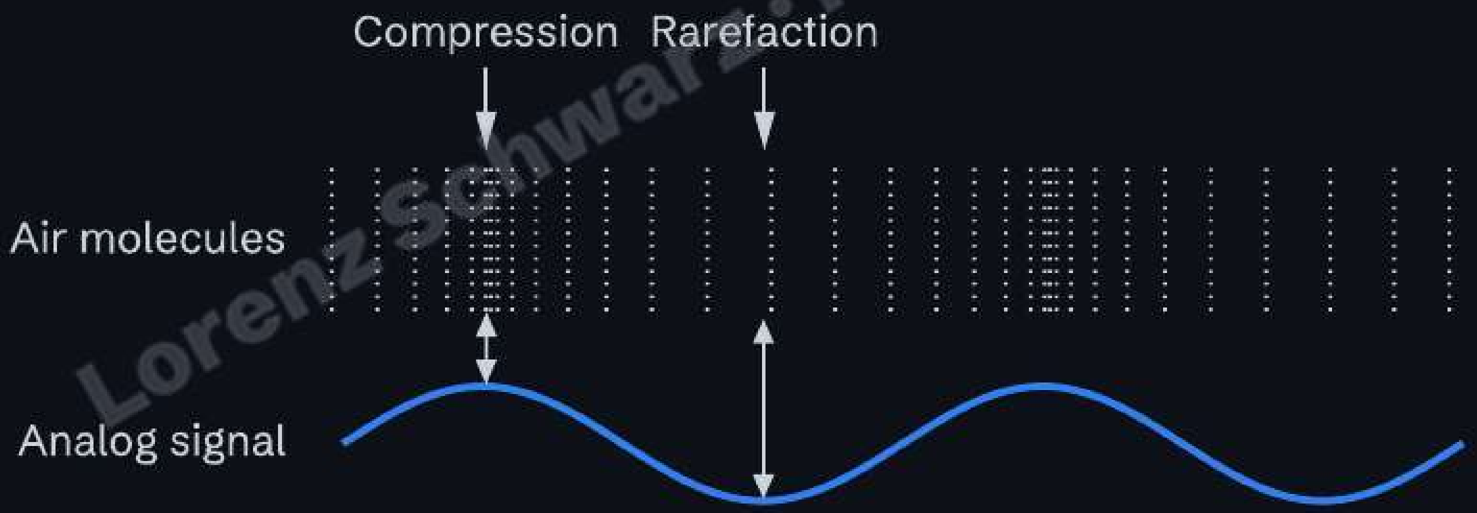


acoustic sound - electrical signal - acoustic sound

Analog signal

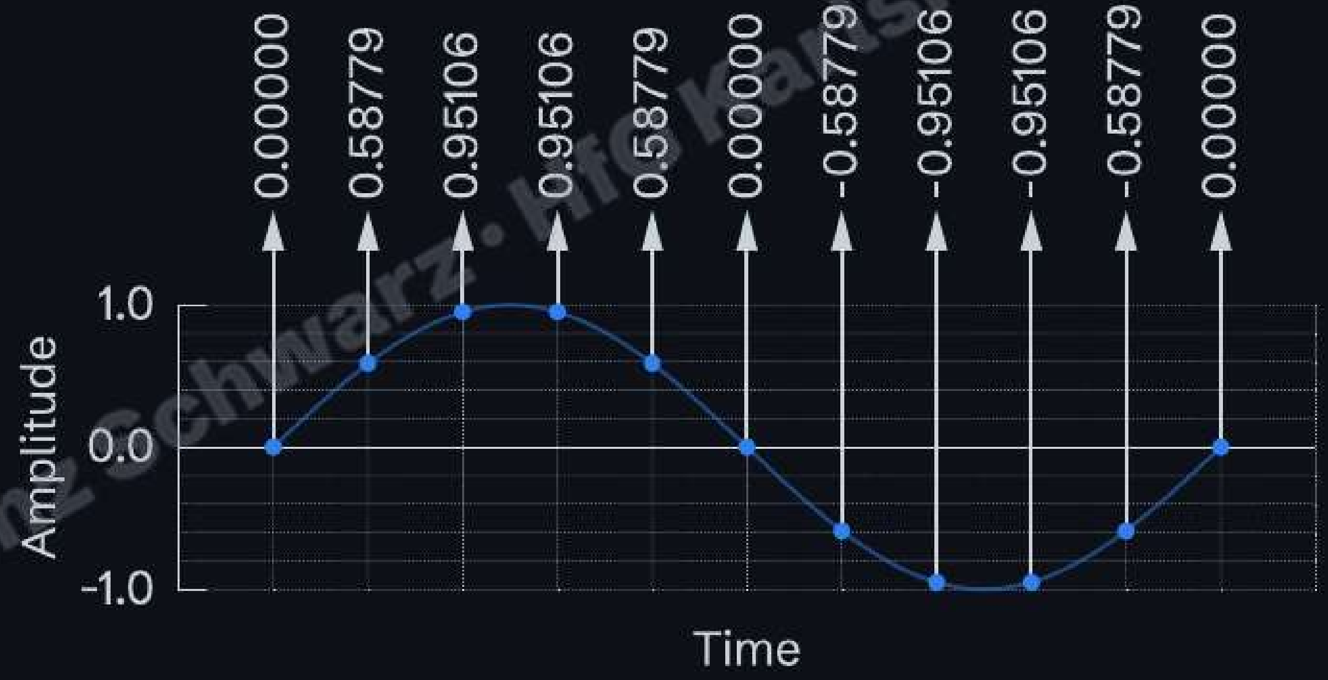
An analog signal is a continuous-time electrical signal whose voltage variations correspond directly to sound pressure variations.

- Continuous in time and amplitude
- Proportional to the original acoustic waveform



Digital representation of sound

Digital audio systems convert the analog audio signal into a stream of numbers.



Advantages of digital audio

- Cost-efficient: storage, duplication, and processing are inexpensive
- Compact and scalable: minimal cabling, no complex analog signal paths
- Low noise floor: no tape hiss or cumulative analog noise
- Deterministic processing: sample-accurate timing and repeatability
- Look-ahead processing: enabled by buffering and latency
- Visual feedback: waveforms, meters, and spectral displays
- Non-destructive editing: undo/redo, versioning, and recall
- Automation: precise recording and playback of parameter changes

Digital representation of sound

In digital audio, the acoustic wave is converted into numerical values representing its amplitude at discrete points in time.

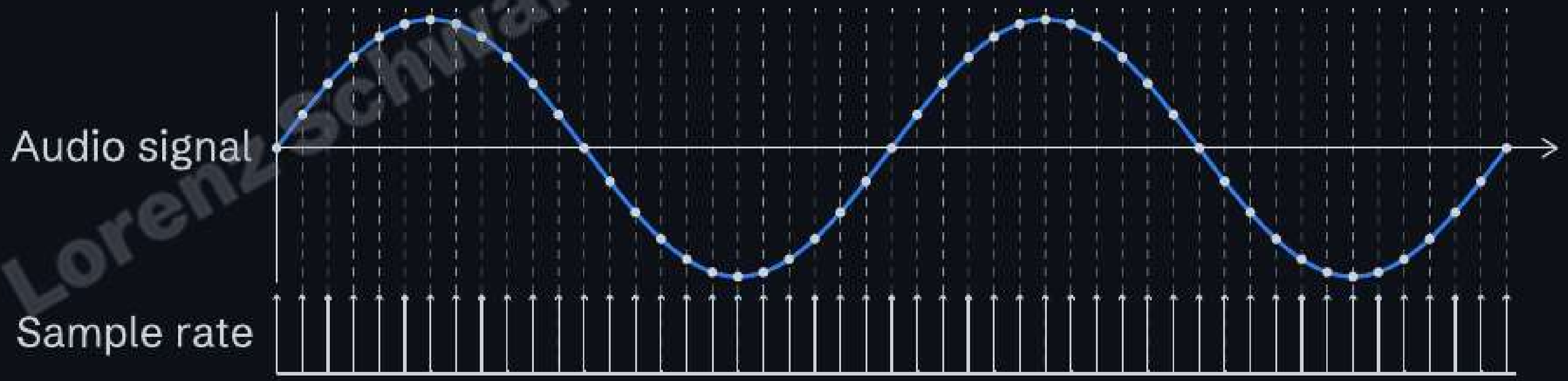
Digital systems can store and represent only:

- Discrete-time values (sampling in time)
- Discrete-amplitude values (quantization in amplitude)

Sample Rate

The sample rate defines how often per time period (s) the discrete voltage levels are measured and stored.

- Defines maximum recordable frequency (Nyquist = ½ rate)
- Common sample rates: 44.1 kHz (CD standard), 48 kHz (professional/video standard), 96 kHz+ (high-resolution audio)



Calculation of the sample interval

$$T_{\text{sample interval}} = \frac{1}{f_{\text{sample rate}}}$$

- $T_{\text{sample interval}}$ is the time between consecutive samples (e.g., 22.67 µs for 44.1 kHz).
- $f_{\text{sample rate}}$ is the sampling rate (e.g., 44.1 kHz for CD audio).

→ *Practical sampling rate must exceed $2 \cdot f_{max}$ by margin due to anti aliasing filter roll-off.*

Nyquist-Shannon Sampling Theorem

To accurately capture all frequencies in a signal, the sampling rate must exceed twice the highest frequency:

$$f_{\text{sampling}} > 2 \cdot f_{\text{max}}$$

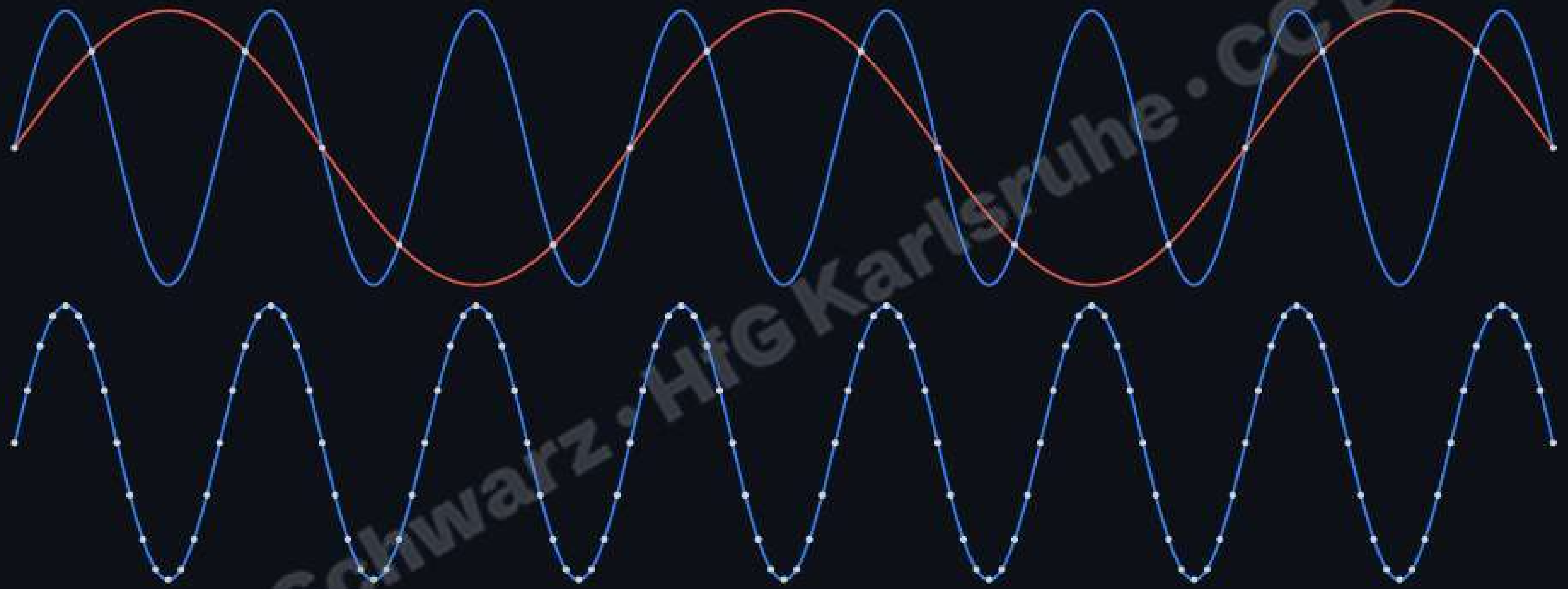
where f_{max} is the highest frequency component in the signal.

Maximum capturable frequency (Nyquist frequency):

$$f_{\text{Nyquist}} = \frac{\text{Sample rate}}{2}$$

→ *Signal frequencies exceeding half the sample rate cause aliasing.*

Aliasing



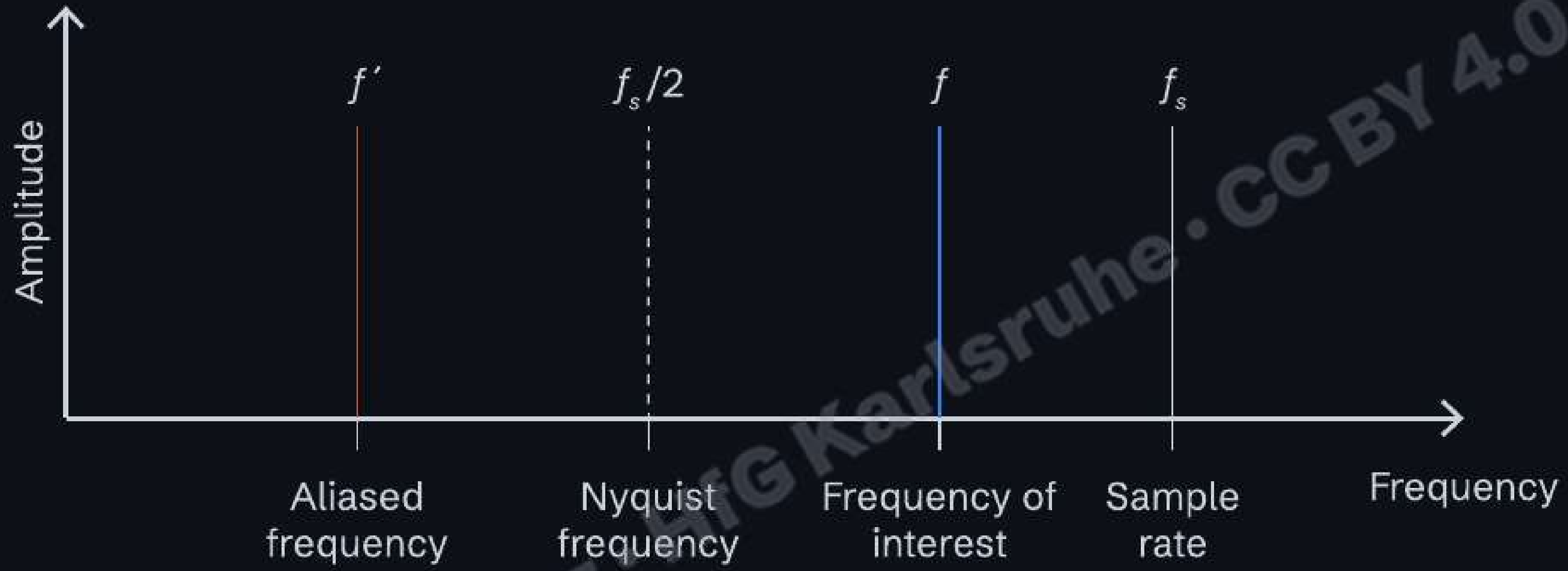
Undersampling (top) vs. correct sampling (bottom) of a sine wave

Constraints of sampling: aliasing

When sampling a signal, frequencies above the Nyquist frequency are reflected back into the audible range, creating unwanted artifacts.

Example:
30 kHz tone sampled at 44.1 kHz (Nyquist = 22.05 kHz) appears as 14.1 kHz, which is the difference between the frequency being sampled and the Nyquist frequency.

The tone is mirrored to the Nyquist frequency and folded back into the useful spectrum.

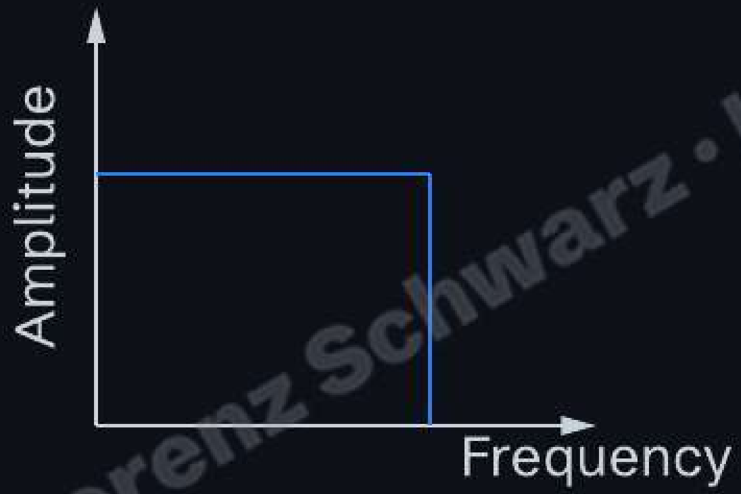


Sine sweep exceeding Nyquist frequency is mirrored back into the audible range.

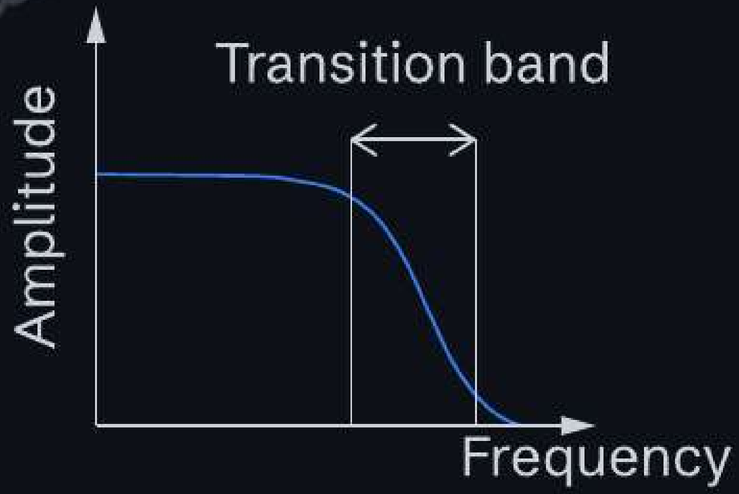
▶ Play aliased sine sweep

Minimizing aliasing: anti-aliasing filter

Low-pass filter is located before the ADC, in the analog domain that attenuates frequencies above Nyquist to prevent them from folding back into the desired signal band.



Ideal filter



Practical filter

Ideal vs. Real Anti-Aliasing Filters

Ideal brick-wall filter:

- Infinite slope at Nyquist frequency
- Perfect pass/stop separation
- Impossible to build in analog domain

Real analog filters:

- Gradual roll-off (typically 12-18 dB/octave)
- Require transition band between passband and stopband
- Introduce phase shifts

Practical implications of anti-aliasing filters

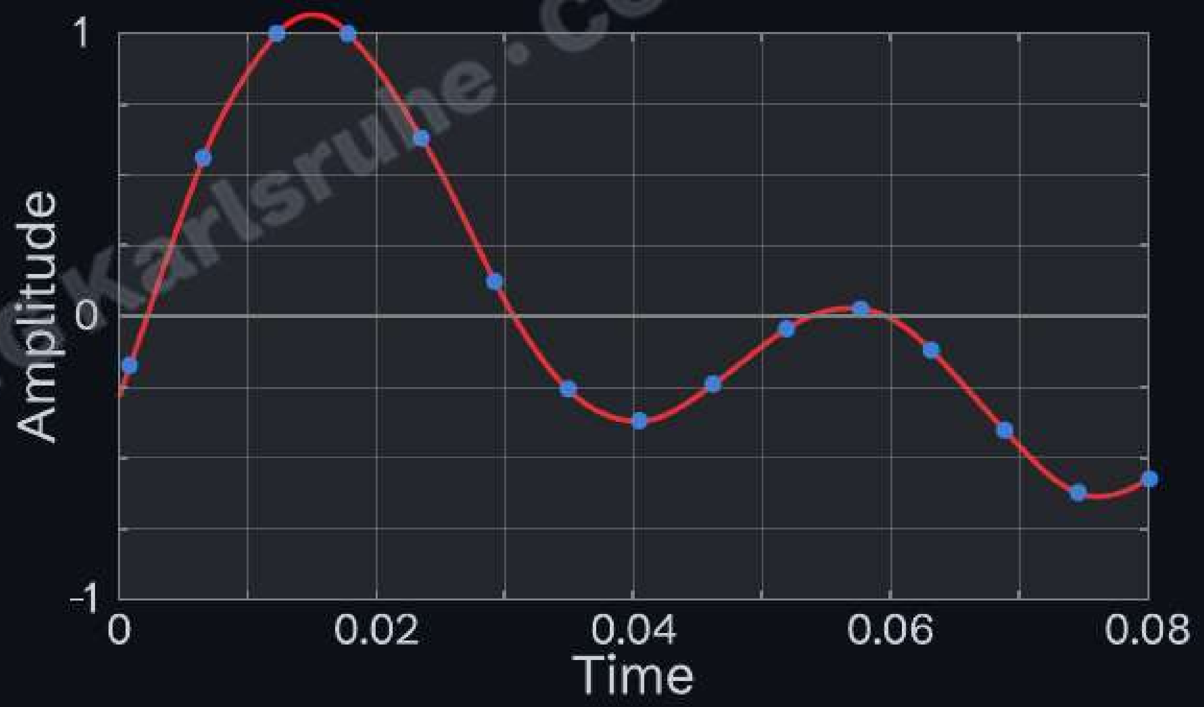
Sample rate must exceed the Nyquist frequency (2× the highest frequency) by sufficient margin to accommodate filter slope.

- 44.1 kHz allows ~20 kHz audio with practical filter design
- Filter begins attenuating around 20 kHz
- Full attenuation reached above 22.05 kHz (Nyquist)

Inter-sample peaks

Sample values might be at 0 dBFS, but the reconstructed waveform between (inter-sample) them can exceed this, potentially causing clipping or distortion.

→ *Provide a headroom buffer of 1 to 2 dBFS during mastering/export*



Inter-sample peak

Oversampling

Oversampling processes audio at a higher internal sample rate than the project rate.

- Easier anti-aliasing filter design (gentler slope, less phase shift)
- Prevents aliasing when plugins generate new harmonics (distortion, saturation)

Quantization

Besides the time-domain sampling, the second important step to digitize a signal is amplitude-domain quantization:

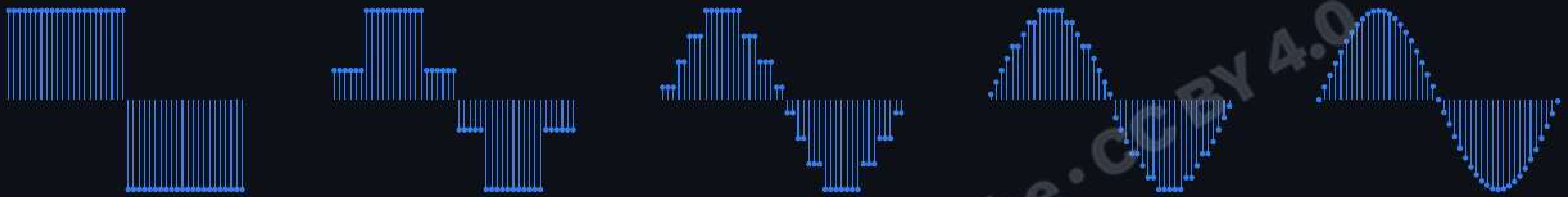
Each sample is rounded to the nearest amplitude value set by the bit depth, introducing small quantization errors. (Quantization and bit depth)

Quantization

Digital systems can only represent numbers with finite (limited) precision.

Sampling requires mapping each sample to the nearest value within a finite set of amplitude levels.

- Bit depth defines resolution, rounding/truncation error, and dynamic range
- Quantization introduces small errors (noise) in the signal
- Common Bit Depths: 16, 24, 32-bit float



Examples (from left to right):

1-bit quantization (2 levels) ▶

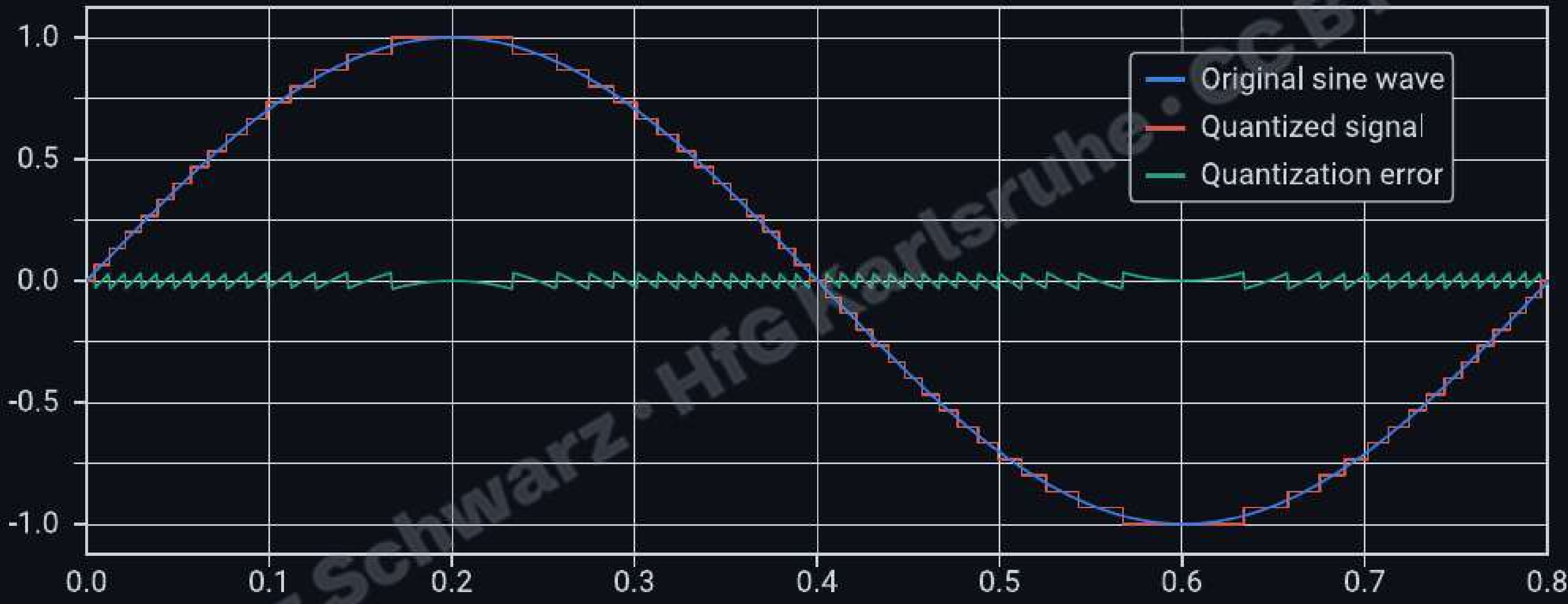
2-bit quantization (4 levels) ▶

3-bit quantization (8 levels) ▶

4-bit quantization (16 levels) ▶

8-bit quantization (256 levels) ▶

Quantization error



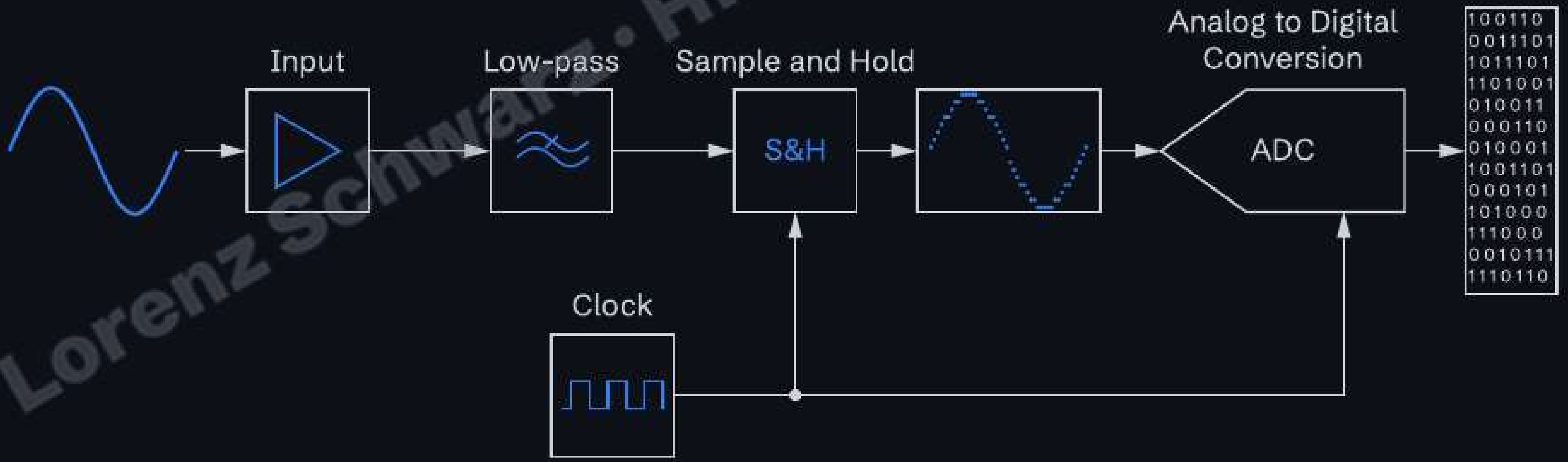
The difference between the actual amplitude (blue) and the quantized value (stepped red line) is the quantization error (green).

Summary: Sampling and quantization

- **Horizontal resolution (time):** determined by sample rate
- **Vertical resolution (amplitude):** determined by bit depth

Analog-to-Digital Conversion (ADC)

1. **Anti-aliasing filter (analog):** remove frequencies > Nyquist
2. **Sampling:** capture instantaneous voltage
3. **Quantization:** round to nearest digital value
4. **Encoding:** convert to binary data



Digital-to-Analog Conversion (DAC)

A **DAC** transforms digital signals (binary data) back into continuous analog waveforms.

1. **Decode:** convert binary to amplitude values
2. **Digital-to-analog conversion:** create stepped analog signal
3. **Reconstruction filter (analog):** smooth steps into continuous wave
4. **Amplification:** boost to line level

Dither

Quantization creates systematic rounding errors that produce audible distortion at low signal levels.

Dither adds very low-level noise before quantization to randomize these errors, transforming harsh distortion into low background noise.

→ *Always dither when exporting to lower bit depth to preserve low-level detail.*

Dynamic range

Dynamic range (DR): theoretical maximum determined by bit depth calculation

$$DR \approx 6.02N + 1.76\ \mathrm{dB}$$

- **N = bit depth**
- 6.02 dB per bit + offset (1.76 dB)

→ 24-bit audio enables a 146 dB dynamic range, corresponding to the span from whisper (minimum) to jet engine at close range (maximum).

Dynamic range of various bit rates

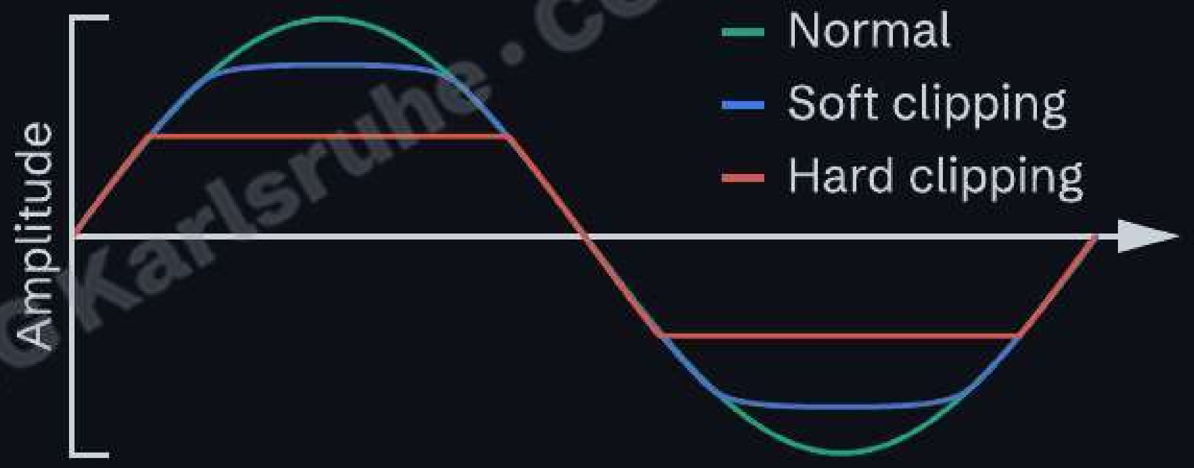
## bits	SNR (Audio)	Minimum amplitude step (dB)	possible values per sample
8	49.93 dB	0.1948 dB	256
16	98.09 dB	0.00598 dB	65,536
24	146.26 dB	0.00000871 dB	16,777,216
32	194.42 dB	0.0000000452 dB	4,294,967,296

→ *Dynamic range of humans: threshold of hearing to threshold of pain ≈ 120 dB*

Clipping

Clipping is a change of the waveform due to electronic or digital limitations.

- Introduces new frequencies (distortion)
- Digital clipping: abrupt flattening (hard clipping)
- Analog clipping: gradual saturation (soft clipping)



Hard clipping vs. soft clipping

▶ Clipping of a sine wave

Floating Point vs. Fixed Point

- **Fixed Point (16/24-bit):**
  - Fixed range, limited dynamic range
  - Used for recording and final delivery
- **Floating Point (32/64-bit):**
  - Audio range: -1.0 to +1.0 represents 0 dBFS at output
  - Internal processing can exceed 1.0 (e.g., value 2.0 ≈ +6 dBFS above 0 dBFS)
  - Prevents clipping during summing (e.g., 0.8 + 0.9 = 1.7, no clip yet)
  - Must be brought back ≤ 1.0 before D/A conversion or file export
  - Used for internal DAW processing

→ *DAWs usually process at 32-bit float.*

Calculating bit rate and file size

Bit rate (amount of data per second):

$$\text{Bit rate} = \text{Sample Rate} \times \text{Bit Depth} \times \text{Channels}$$

File size (total data for duration):

$$\text{File Size (bytes)} = \frac{\text{Bit Rate} \times \text{Duration (s)}}{8}$$

Example: 1 minute stereo, 48 kHz, 24-bit
= (48,000 × 24 × 2 × 60) / 8 = 8,640,000 bytes ≈ **8.64 MB**

Example sizes for WAV file

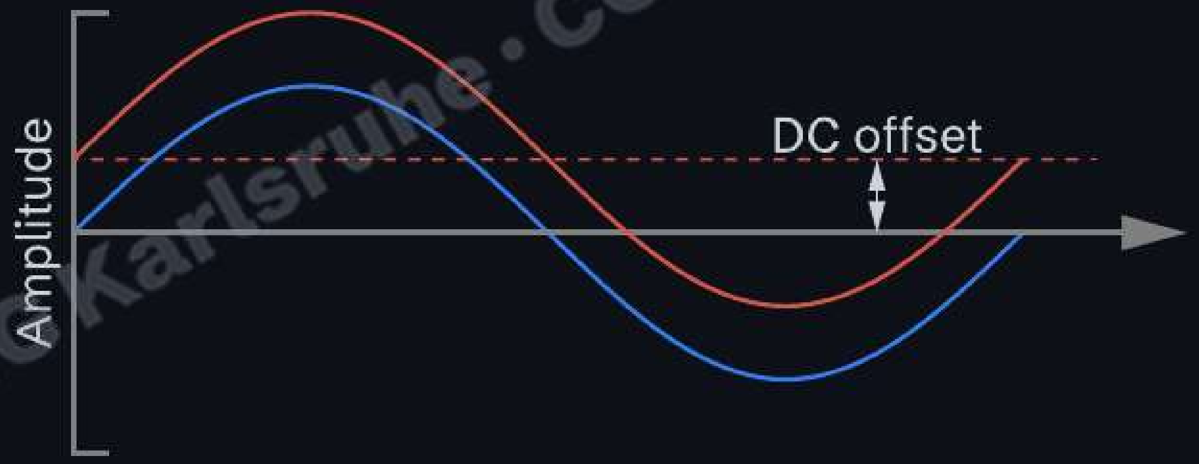
for a one minute long file:

Channels	Sample Rate (kHz)	Bit Depth	File Size (MB)
1	44.1	16	5.29 MB
1	44.1	24	7.94 MB
1	48	24	8.64 MB
1	48	32 float	11.52 MB
1	96	24	17.28 MB
1	96	32 float	23.04 MB
2	**48**	**24**	**17.28 MB**

DC offset

DC offset occurs when a waveform has a non-zero average value, shifting the entire signal away from the zero line.

- Reduces available dynamic range
- Can cause clipping during processing
- Creates clicks/pops when starting/stopping playback
- May damage speakers



Waveform with DC offset

→ *Apply DC offset removal / high-pass filter (e.g., 20 Hz).*

Practical DAW settings

Every project defines two key parameters that determine audio quality and file size:

- **Sample rate (kHz):** how often the sound is measured (e.g. 44.1, 48, or 96 kHz).
- **Bit depth (bits):** how precisely each sample is stored (e.g. 16, 24, or 32-bit float).

→ *Recommended setting: 48 kHz / 24-bit*

Digital processing: latency and buffers

Latency is the time between an audio signal entering and leaving the system.

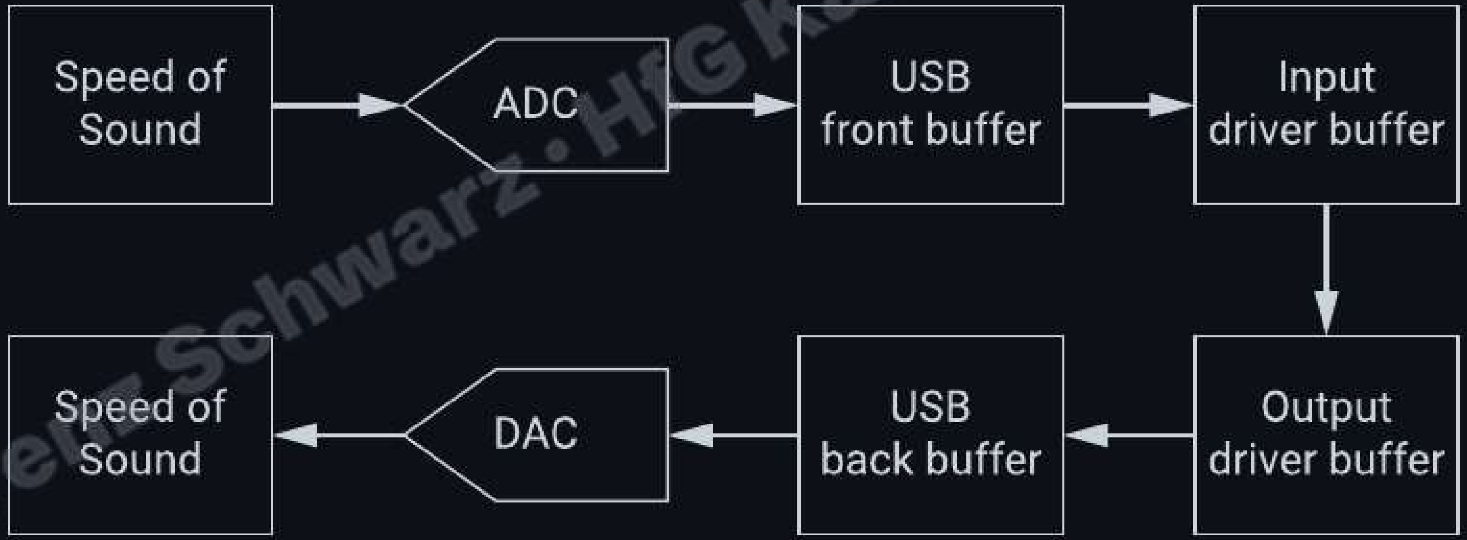
Main causes:

- Buffering in the computer or audio interface
- A/D and D/A conversion
- Digital processing (plugins, effects)

Some latency is unavoidable (conversion, processing) while buffer latency is adjustable.

Buffers

Digital audio systems use buffers (small sections of temporary memory) to process audio in blocks. The buffer size affects both latency (delay) and system stability.



Buffer size and latency

Buffer size defines how many audio samples the system processes at once and directly affects latency and stability.

Small buffers (64–128 samples): Low latency for recording and live monitoring, higher CPU load, risk of dropouts

Large buffers (512–1024 samples): Higher latency, lower CPU load, stable playback for mixing

→ *Use the smallest buffer size that avoids dropouts for the given task.*

Buffer Size & Delay (Latency)

$$\text{Buffer (samples)} = f_s\,[\text{Hz}] \times t\,[\text{s}]$$

$$\text{Delay (ms)} = \frac{\text{buffer}_{\text{samples}}}{f_s} \times 1000$$

Example:

Buffer = 128 samples at 48 kHz:
Delay = (128 / 48000) × 1000 = 2.67 ms

Human perception of latency

humans can detect a silent gap between two sounds of about 2–3 ms.

If sounds are less similar, or in noise / lower intensity, or onsets with less pronounced attack phase, threshold increases (≥ 4-5 ms).

→ *Buffer settings around 128 samples (≈3ms at 48kHz) feel immediate to most musicians during recording*

Buffer size and delay

Buffer Size in samples	Delay in ms for 44.1kHz	Delay in ms for 48kHz
32	0.72	0.66
64	1.45	1.33
128	2.9	2.6
256	5.8	5.3
512	11.6	10.6
1024	23.2	21.3
2048	45.9	42.1

Physical vs. Digital Latency

Well-optimized digital systems introduce less latency than the physical distance between performers and their monitoring systems (speed of sound ≈343 m/s in air).

Acoustic propagation delay:

- 1 meter: ≈3 ms
- 3 meters: ≈9 ms (typical distance to studio monitors)
- Distance between band members on stage: 3–6 meters (≈9–18 ms)

→ *Digital audio latency (3–10 ms) is comparable to or shorter than acoustic delays musicians naturally encounter.*

File formats and storage

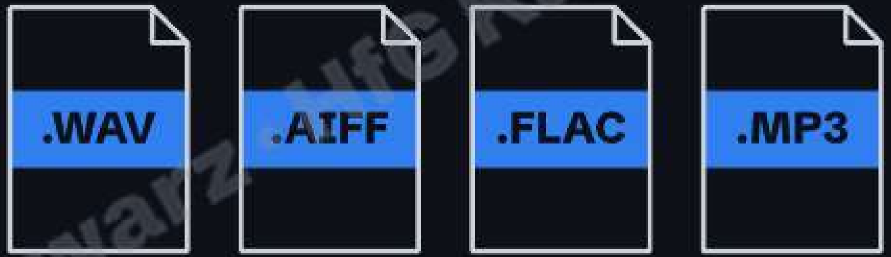
Pulse-Code Modulation (PCM):

Analog signal amplitude is **sampled at uniform intervals** and each sample is **quantized to the nearest digital step.**

→ *PCM = sampling + quantization.*

Audio File Formats

- **WAV / AIFF (PCM):** standard uncompressed, full quality.
- **FLAC / ALAC:** lossless compression.
- **MP3 / AAC / OGG:** lossy, smaller size but reduced fidelity.



Audio file formats and compression

- Uncompressed (PCM)
  - WAV / AIFF: full quality, standard for recording and production
- Lossless compression
  - FLAC / ALAC: identical audio quality, reduced file size
  - → suitable for storage and archiving
- Lossy compression
  - MP3 / AAC / OGG: smaller files with irreversible quality loss
  - → suitable only for final delivery/distribution

Wordclock

Word clock is used when multiple digital devices (interface and converters) are connected:

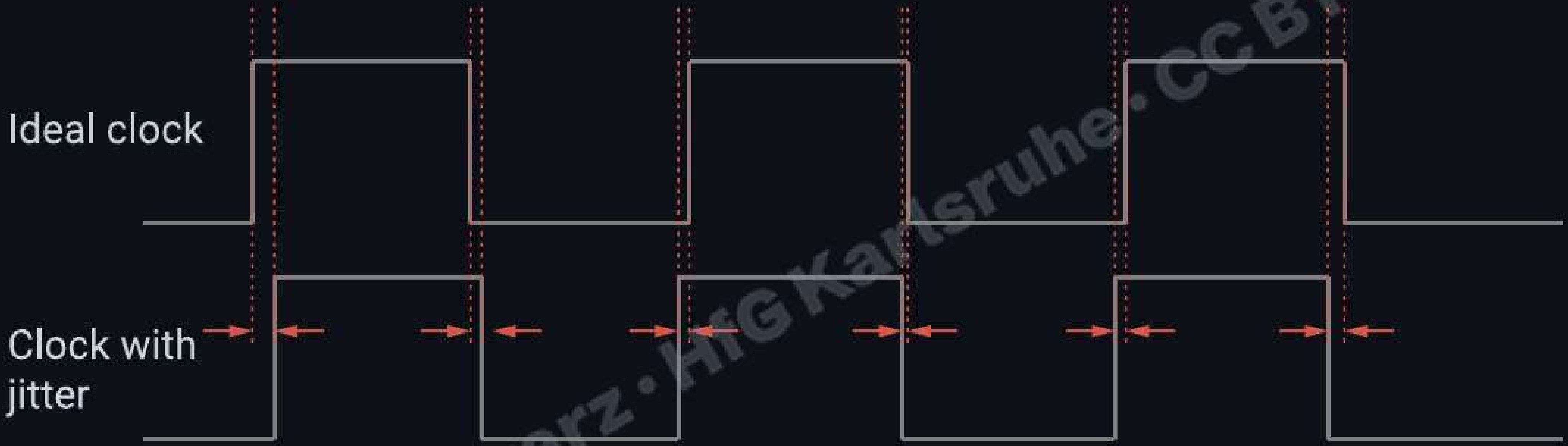
- One device as "master clock," others as "slave"
- A clock signal that synchronizes sampling across digital audio devices
- Ensures all devices sample at the same time
- Prevents clicks, jitter, and drift

Jitter and synchronization

Jitter is unwanted timing variation in the digital audio clock, causing samples to be processed at incorrect times and potentially introducing distortion.

Synchronization aligns multiple devices to a common clock to minimize jitter and ensure stable audio transfer.

Jitter



Copyright and Licensing

Contact: lschwarz@hfg-karlsruhe.de

← Chapters