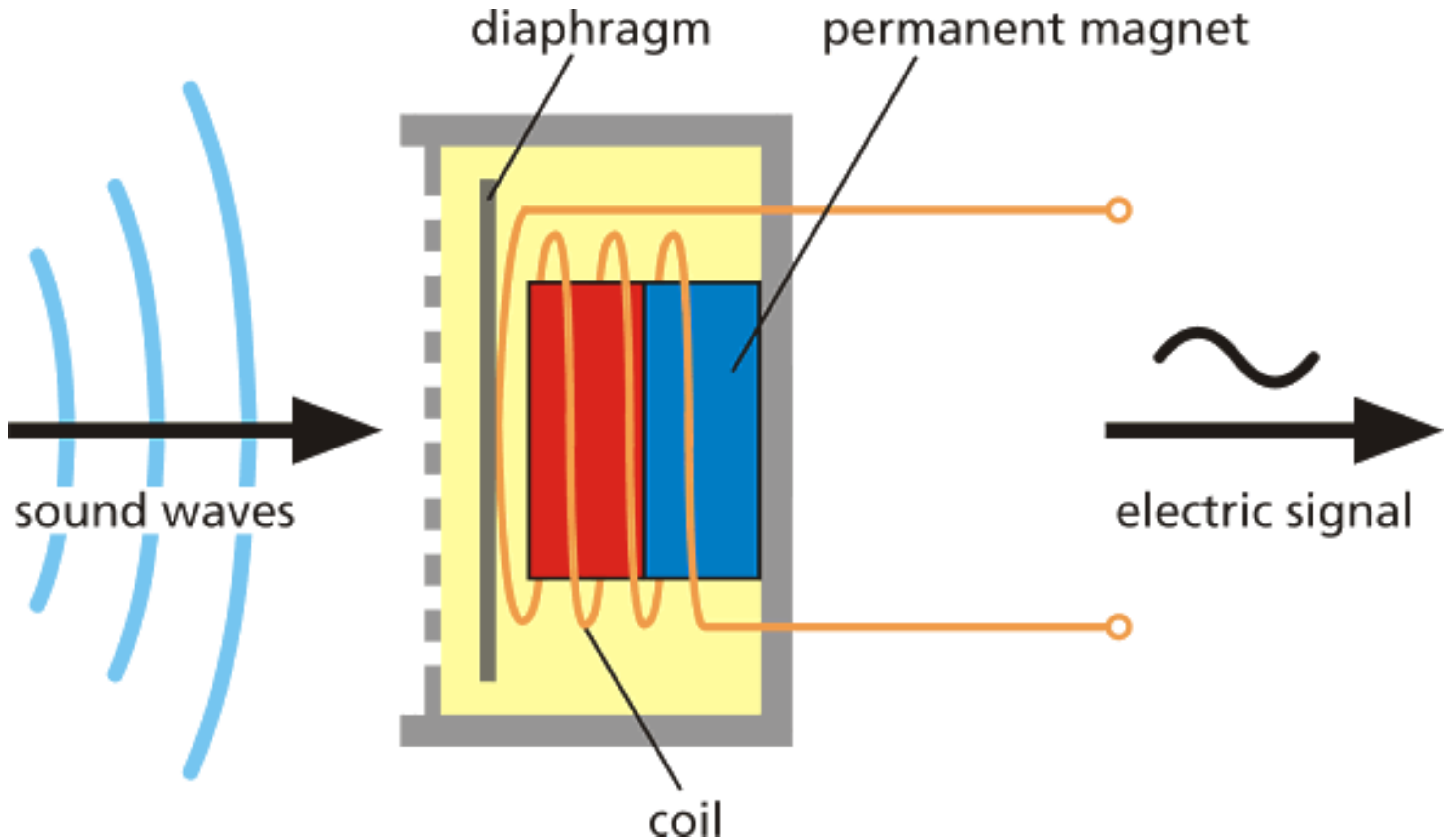
Digitizing Sound

EECS 4462 - Digital Audio

September 29, 2020

Analog audio



- Sound waves travelling through the air hit the diaphragm of a microphone (a transducer) and get converted to an electric signal whose amplitude is **analogous** to that of the sound waves

https://garretthes.wordpress.com/2015/04/07/audio-signal-flow-section-1-microphone-preamp

Analog audio

- The electric signal is usually amplified as it is quite weak
- It can then be recorded onto tape as magnetic waveforms, or a vinyl record as grooves
- The electric signal can be reproduced from the recording since the magnetic waveforms or the groove undulations are analogous to those of the electric signal
- The electric signal can be converted to sound using a speaker (another transducer) whose vibrations are analogous to those of the electric signal

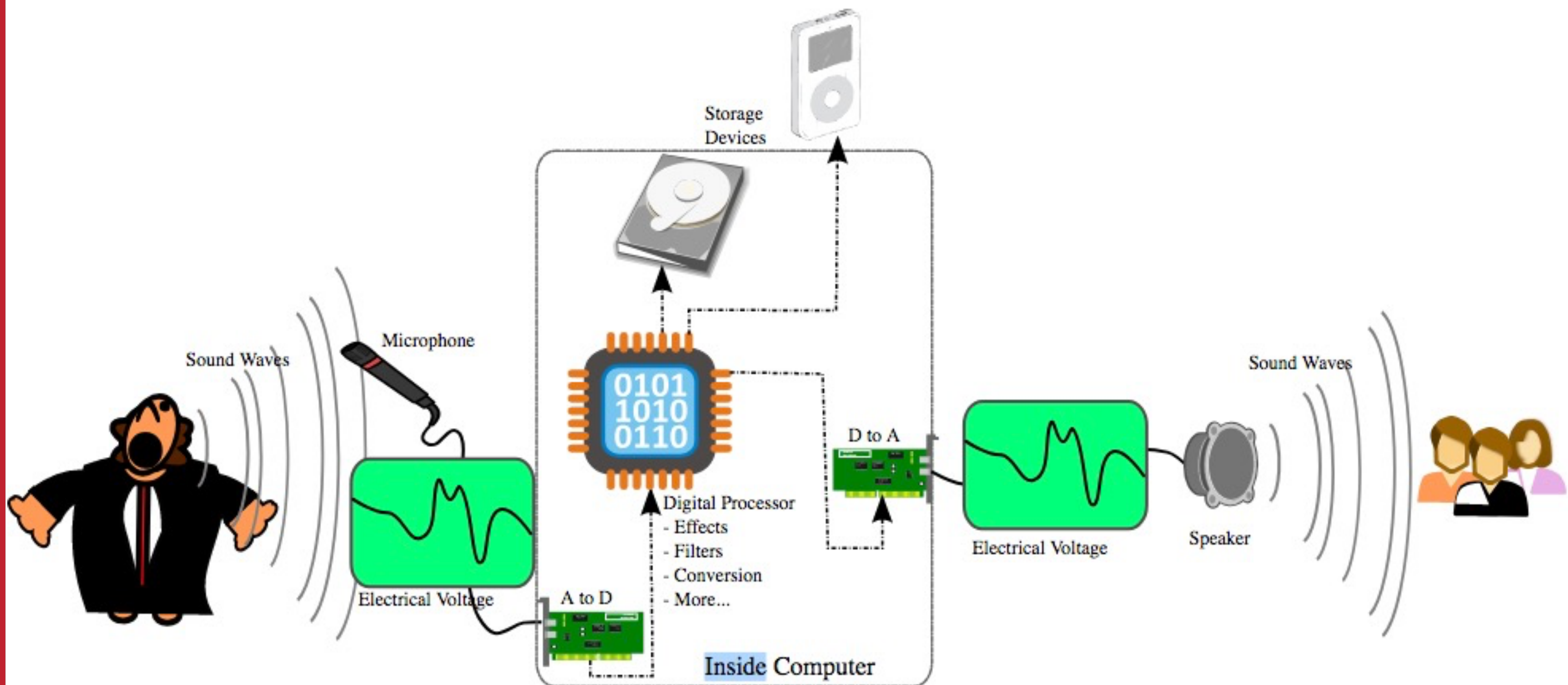
Digital Audio

- An alternative way to store and reproduce an electric signal that represents audio
- Easy/Cheap to store
- Easy to copy
- Easy to transmit
- Easy to process
- Does not degrade
- No interference/noise/distortion

Analog to Digital

- Analog audio is continuous
- Digital audio, as anything in computers, is discrete
- Need an **Analog-to-Digital Converter** (ADC)
- An ADC converts a continuous-time and continuous-amplitude analog signal to a **discrete-time** and **discrete-amplitude** digital signal
- A Digital-to-Analog Converter (DAC) performs the opposite operation

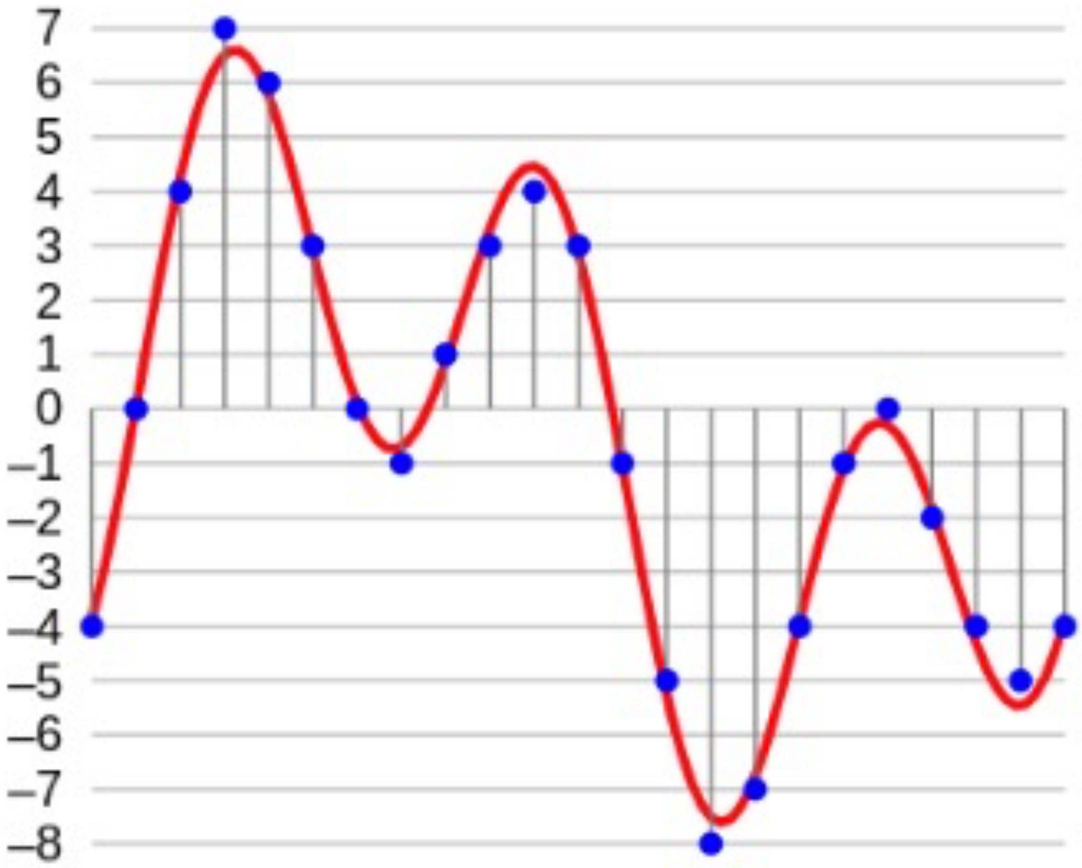
The lifecycle of sound



Discrete Amplitude ➔ Bit Depth

- At a given point in time, the ADC observes the level of the incoming signal and produces a binary value for it
- The accuracy of the digital representation will depend on how many bits are used to represent this binary value
- The number of bits used for a particular conversion are referred to as the **bit depth**
- A bit depth of 16 bits means that the amplitude of the digital signal will have a resolution of 2^{16} different values
  - CDs have a bit depth of 16
  - Most DAWs have a bit depth of 24, but use 32-bit floating point numbers internally

Bit depth example



What is the bit depth for this conversion?

Quantization error

- The process of converting analog level to a binary value is called **quantization**
- Because there is a finite number of possible binary values this process introduces error, referred to as **quantization error**
- Quantization error results to noise when the electric signal is reproduced

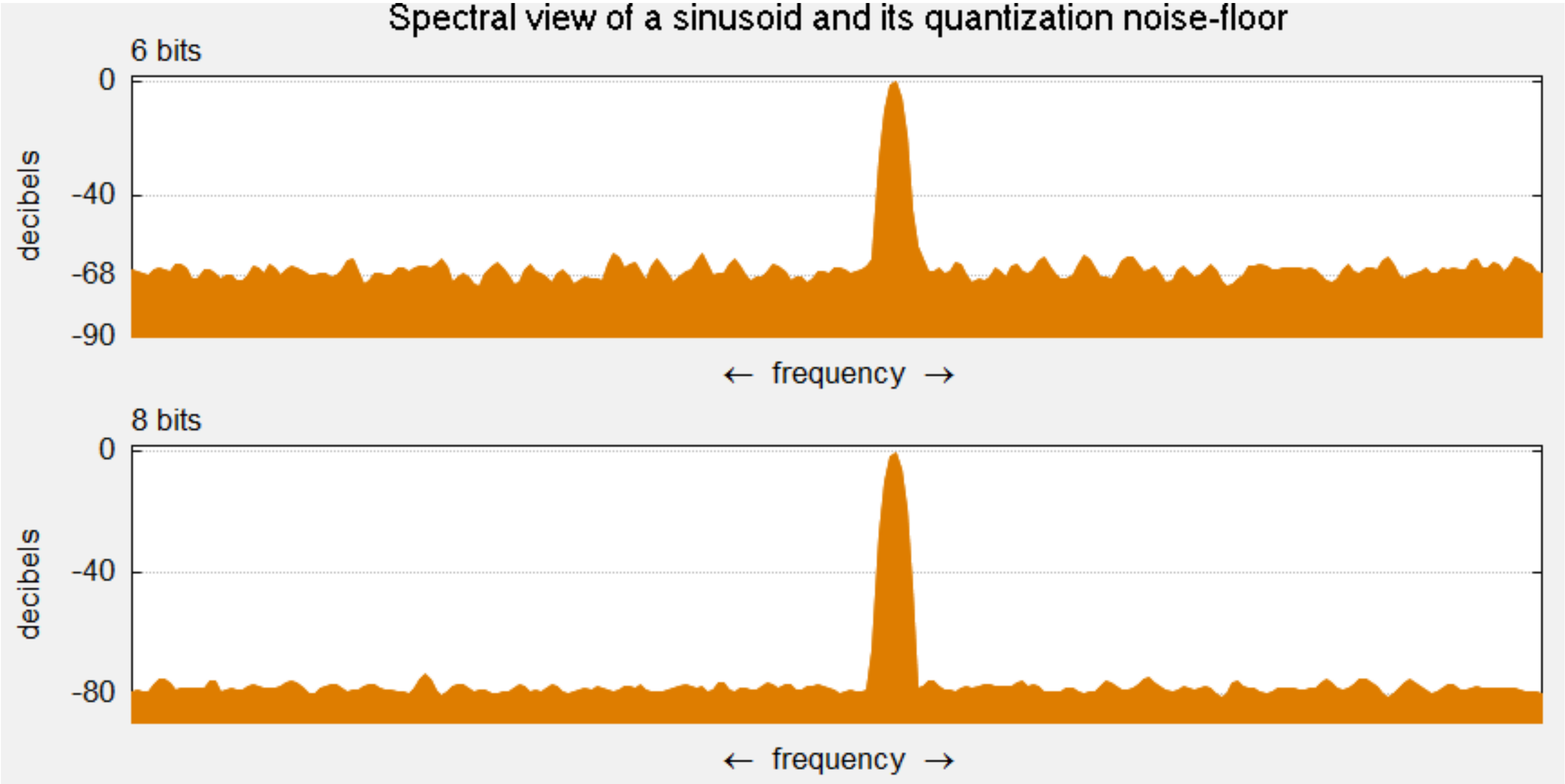
Quantization error

- Signal to quantization noise ratio is calculated as

  $$SQNR = 20 \log_{10}(2^{Q}) \approx 6.02Q \text{ dB}$$

- A 16-bit digital audio signal has SQNR of 96.3 dB
- Therefore audio signals that have a dynamic range of up to 96.3 dB can be represented accurately with 16 bits
  - **Dynamic range**: The difference between the loudest and the softest part of an audio signal
- Music has a dynamic range of at most 40dB, usually 10-20dB

Noise floor



By Bob K - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=31989085

Dithering

- When quantizing audio signals of very low level (very quiet), the resulting binary value may be stuck at 0 even though the input varies
- This can also happen when reducing the bit depth of a recording
- By adding **dither**, a small amount of random noise, to the input, the binary values oscillate between two different levels resulting in a more accurate result at the expense of a bit of added noise

Discrete Time→ Sampling Rate

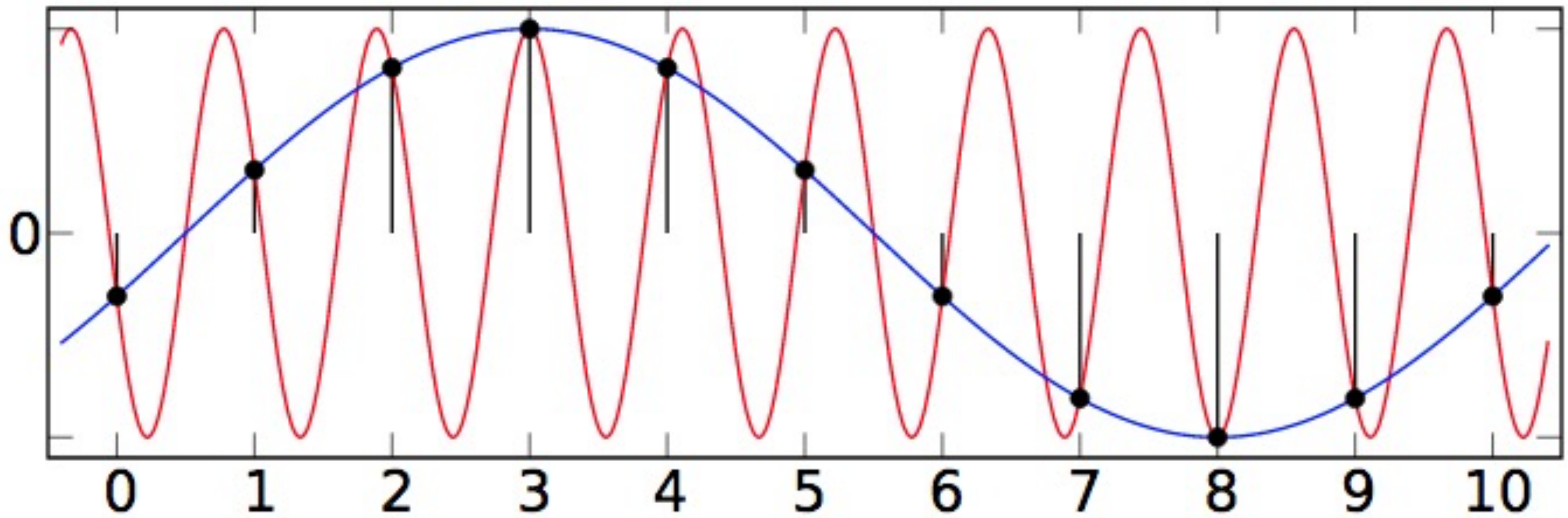
- An ADC takes periodic *samples* (voltages) of the analog signal at fixed intervals
- **Sampling rate** indicates how many times per second we sample the analog signal
  - Measured in Hz
  - Telephone sampling rate: 8kHz
  - CD sampling rate: 44.1kHz
  - DVD sampling rate: 48kHz (or more)
- The sampling rate determines which frequencies in the input can be accurately represented in the digital signal

Sampling theorem

- If the highest frequency in a signal is B, then a sampling rate of 2B will ensure no loss of information in the digital signal, and a perfect reconstruction of the analog signal from a frequency perspective
- 2B is referred to as the Nyquist rate
- B is referred to as the Nyquist frequency
- For perfect reconstruction, it is important that the input signal is *bandlimited,* i.e. the band of frequencies in the signal is limited, e.g. < B, otherwise we have artifacts in the reconstructed signal due to **aliasing**

Aliasing

- If the sampling rate is 2B, sampling a frequency higher than B, will result in a phantom frequency below B to appear in the output



https://en.wikipedia.org/wiki/Aliasing

Avoiding aliasing

- An analog audio signal may contain frequencies higher than 20kHz due to noise, interference etc.
- This can cause aliasing artifacts in the reconstructed signal
- An **anti-aliasing filter** is employed prior to sampling to remove all frequencies above 20kHz
- Since analog filters are not perfect, the CD sampling rate was set at 44.1kHz, to allow for frequencies up to 22.05kHz that may remain in the signal after the anti-aliasing filter

Avoiding aliasing

- Modern ADCs use many other methods to improve the quality of the digital signal reconstruction
- **Oversampling** is very common
- The ADC samples at 256 times the Nyquist rate
  - Very cheap to do with modern circuitry
- Removal of inaudible frequencies happens in the digital world
  - Much cleaner, no noise
- The signal is then downsampled to 44.1kHz or 48kHz for storage and processing

Pulse code modulation (PCM)

- Representing an analog signal as a series of sampled values is referred to as the Pulse Code Modulation (PCM) format
- Each value is stored as a 2's complement signed integer
- Endianness differs with format
  - .wav files are little endian
  - .aiff files are big endian
- If there are more than one audio channels in the signal (as in stereo sound) the samples for each channel are interleaved