CTP 431 Music and Audio Computing

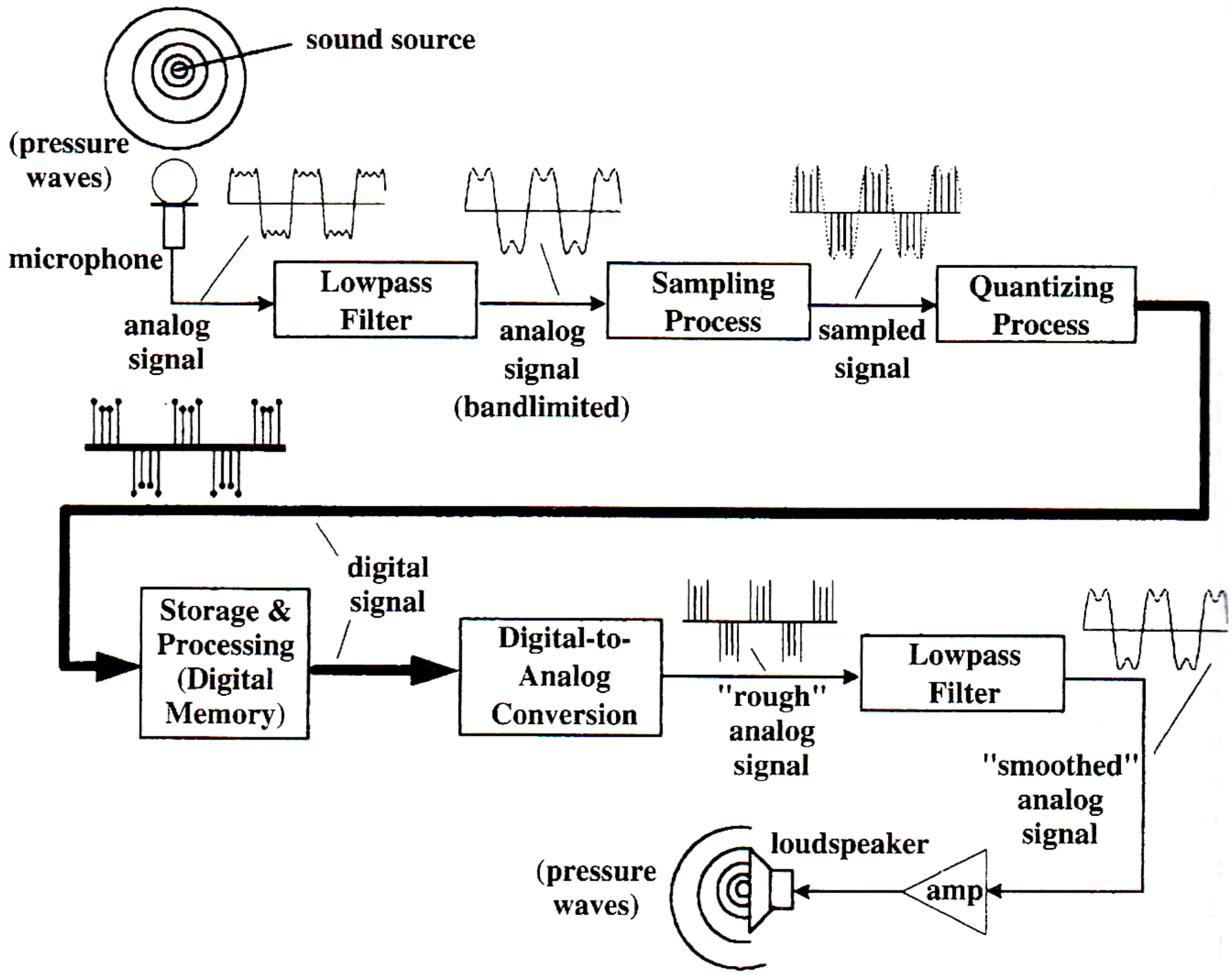
Digital Audio

Graduate School of Culture Technology (GSCT)
Juhan Nam

Outlines

- Introduction
  - Digital audio chain
  - Transducers
- Sampling
  - Sampling theorem
  - Aliasing and reconstruction
- Quantization
  - Quantization error: SNR
  - Dynamic range

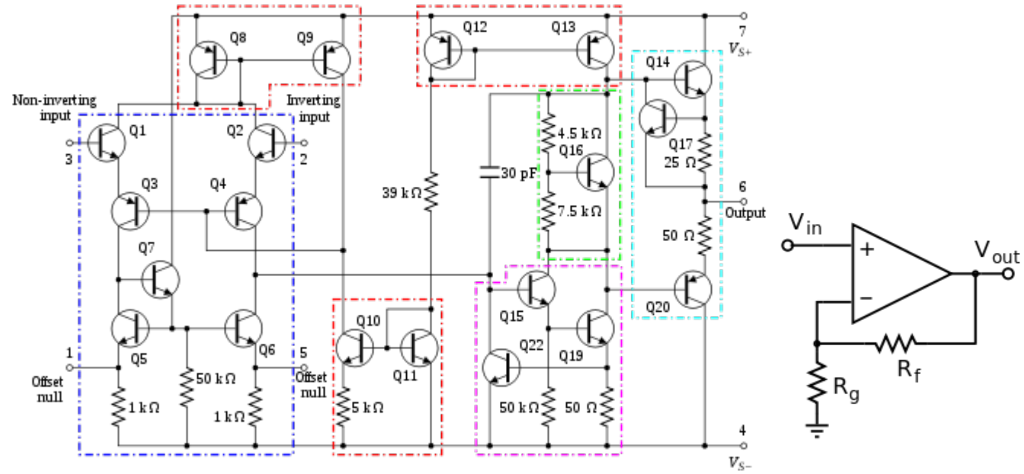
Digital Audio Chain



...0 0 1 0 1 0 ...

Why Digital?

Helmholtz Resonators



Op-amp and amplifier

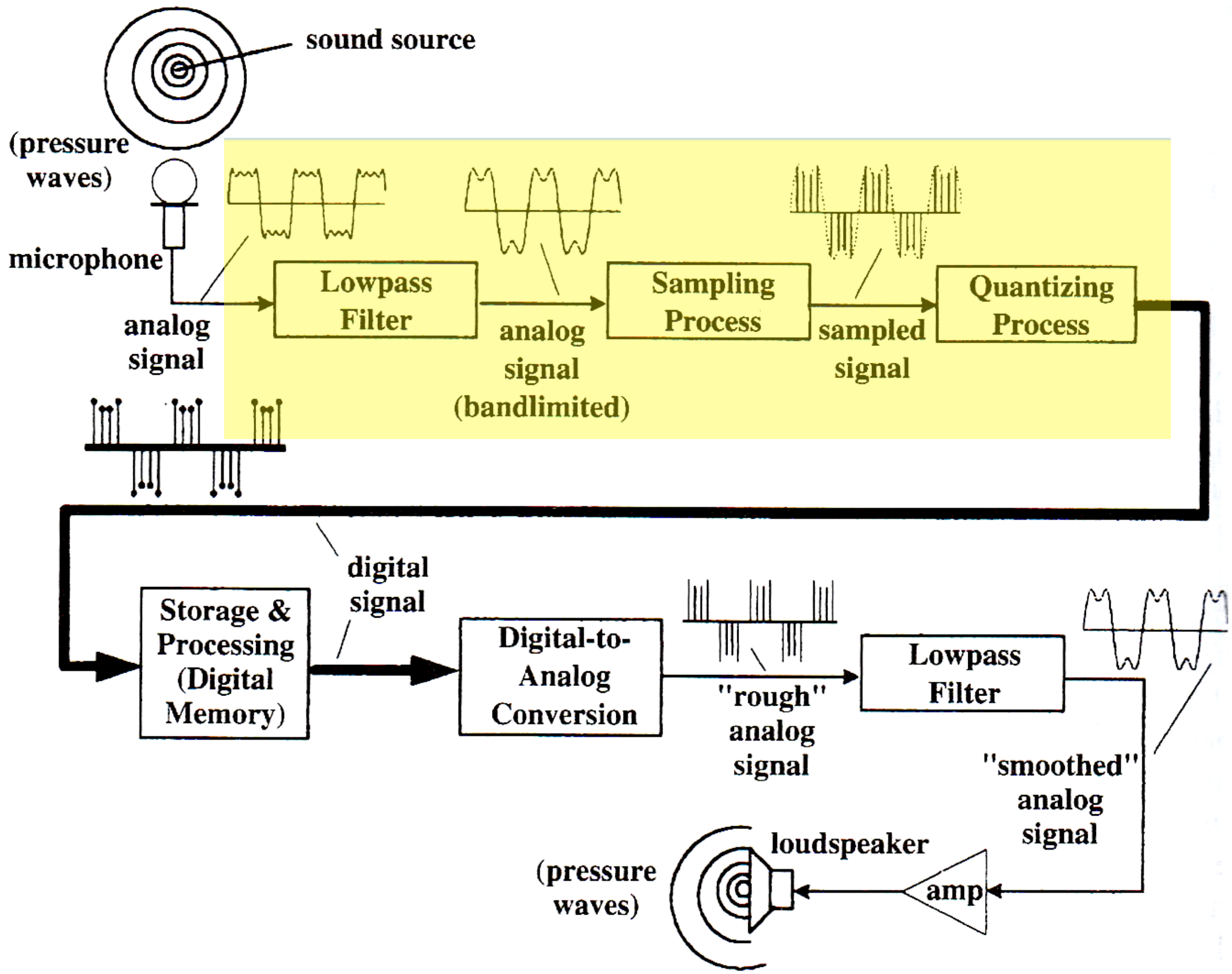
$$y[n] = g * x[n]$$

On computer

Transducers

- Convert one form of energy to another form
  - The forms are different but the information remains (almost) the same
- Microphones
  - Sound wave to electrical signal
  - Dynamic / condenser microphones
- Speakers
  - Electrical signal to sound wave
  - Generate distortion (by diaphragm)
  - Crossover networks: woofer / tweeter

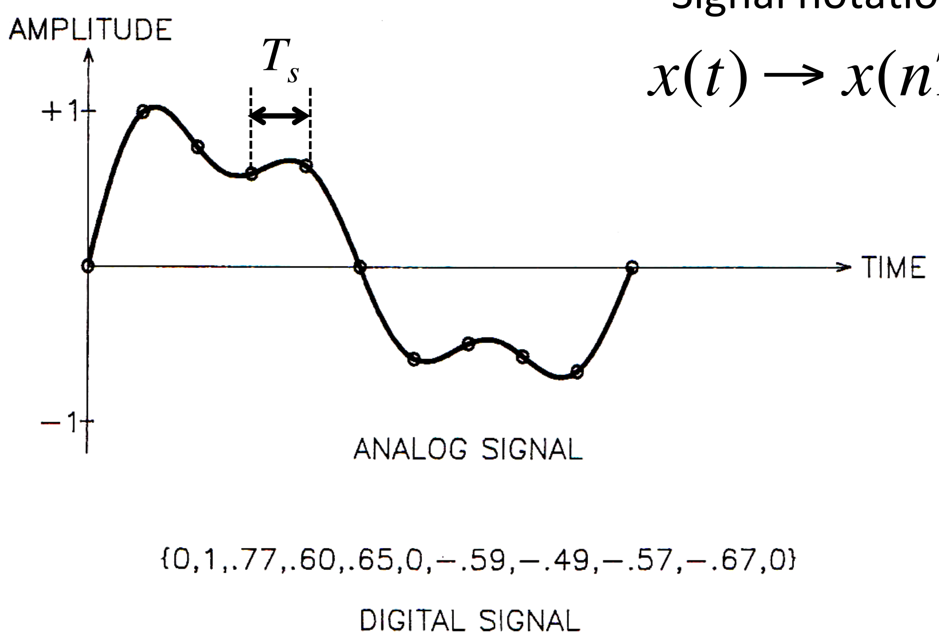
Analog to Digital



...0 0 1 0 1 0 ...

Sampling

- Convert continuous-time signal to discrete-time signal by periodically picking up the instantaneous values
  - Represented as a sequence of numbers; pulse code modulation (PCM)
  - Sampling period (T_s): the amount of time between samples
  - Sampling rate ($f_s = 1/T_s$)



Signal notation

$$x(t) \rightarrow x(nT_s)$$

Sampling Theorem

- What is an appropriate sampling rate?
  - Too high: increase data rate
  - Too low: become hard to reconstruct the original signal

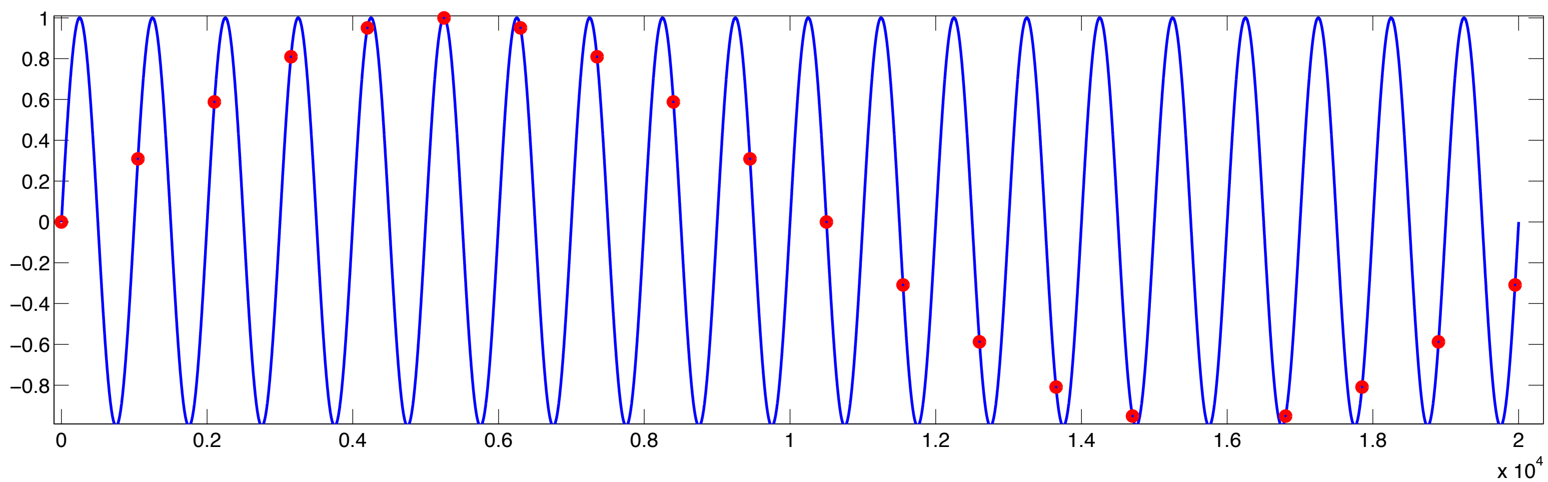
- Sampling Theorem
  - In order for a band-limited signal to be reconstructed fully, the sampling rate must be greater than twice the maximum frequency in the signal

$$\boxed{f_s > 2 \cdot f_m}$$

  - Half the sampling rate is called Nyquist frequency ($\frac{f_s}{2}$)

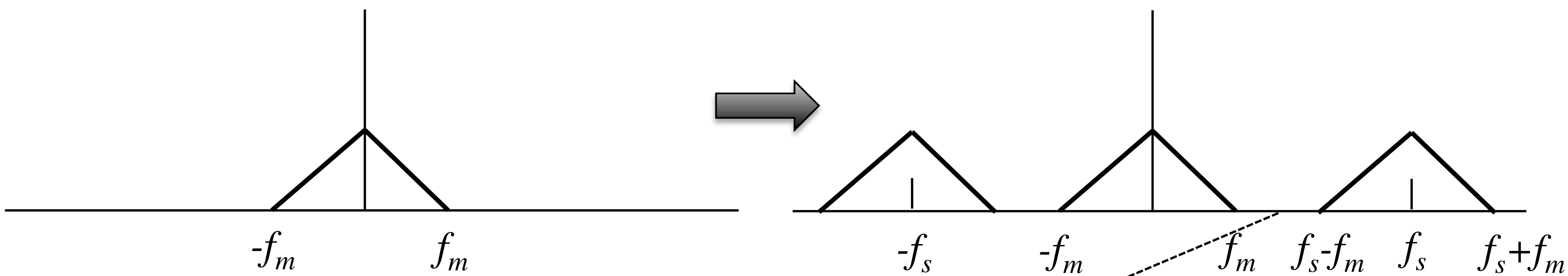
Aliasing

- If the sampling rate is less than twice the maximum frequency, the high-frequency content is folded over to lower frequency range



Sampling in Frequency Domain

- Sampling in time corresponds to replicating the original signal at every f_s frequency



$$\boxed{f_m < f_s - f_m}$$

- Why ?

$$\boxed{f_2 = f_1 \pm m f_s}$$

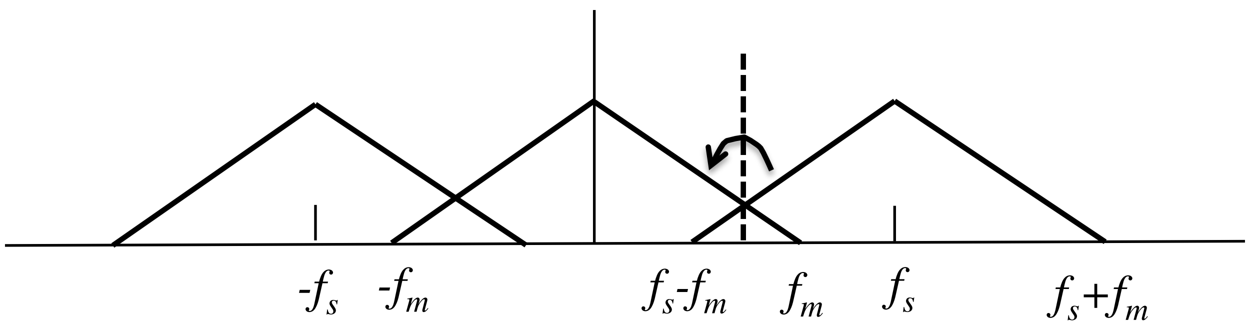
$$x_1(t) = A\sin(\omega_1 t) = A\sin(2\pi f_1 n / f_s)$$

$$x_2(t) = A\sin(\omega_2 t) = A\sin(2\pi f_2 n / f_s) = A\sin(2\pi (f_1 \pm m f_s) n / f_s)$$

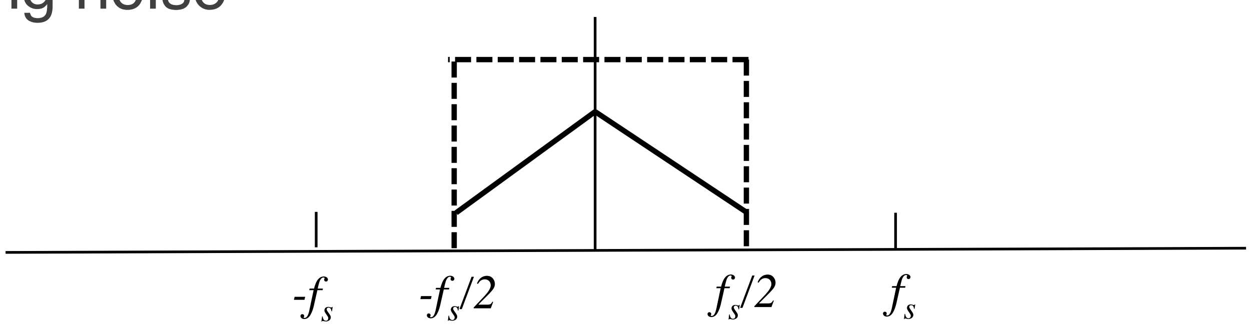
$$= A\sin(2\pi f_1 n / f_s \pm 2\pi m n) = A\sin(2\pi f_1 n / f_s) = x_1(t)$$

Aliasing in Frequency Domain

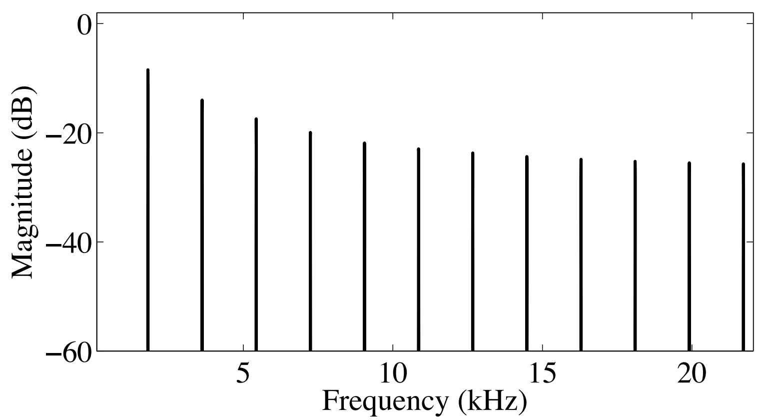
- The high-frequency content is folded over to lower frequency range from the replicated images



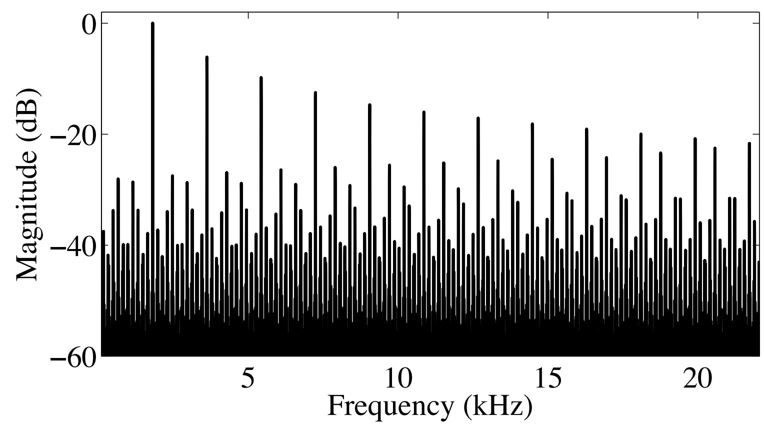
- A low-pass filter is applied before sampling to avoid the aliasing noise



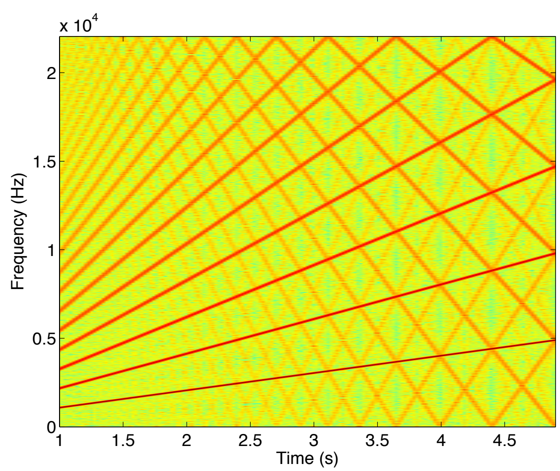
Example of Aliasing



Bandlimited sawtooth wave spectrum



Trivial sawtooth wave spectrum



Frequency sweep of the trivial sawtooth wave

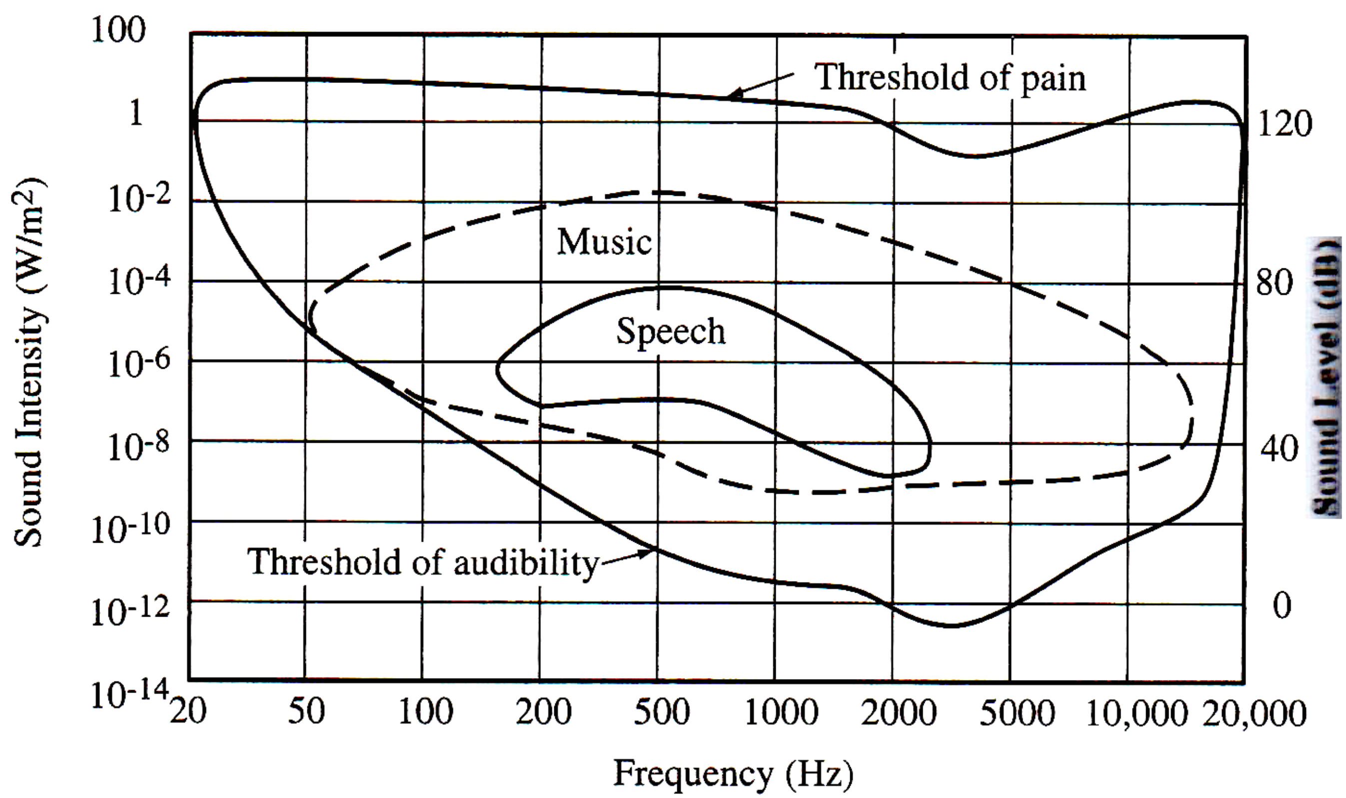
Example of Aliasing

- Aliasing in Video
  - https://www.youtube.com/watch?v=QOqtdl2sJk0
  - https://www.youtube.com/watch?v=jHS9JGkEOmA

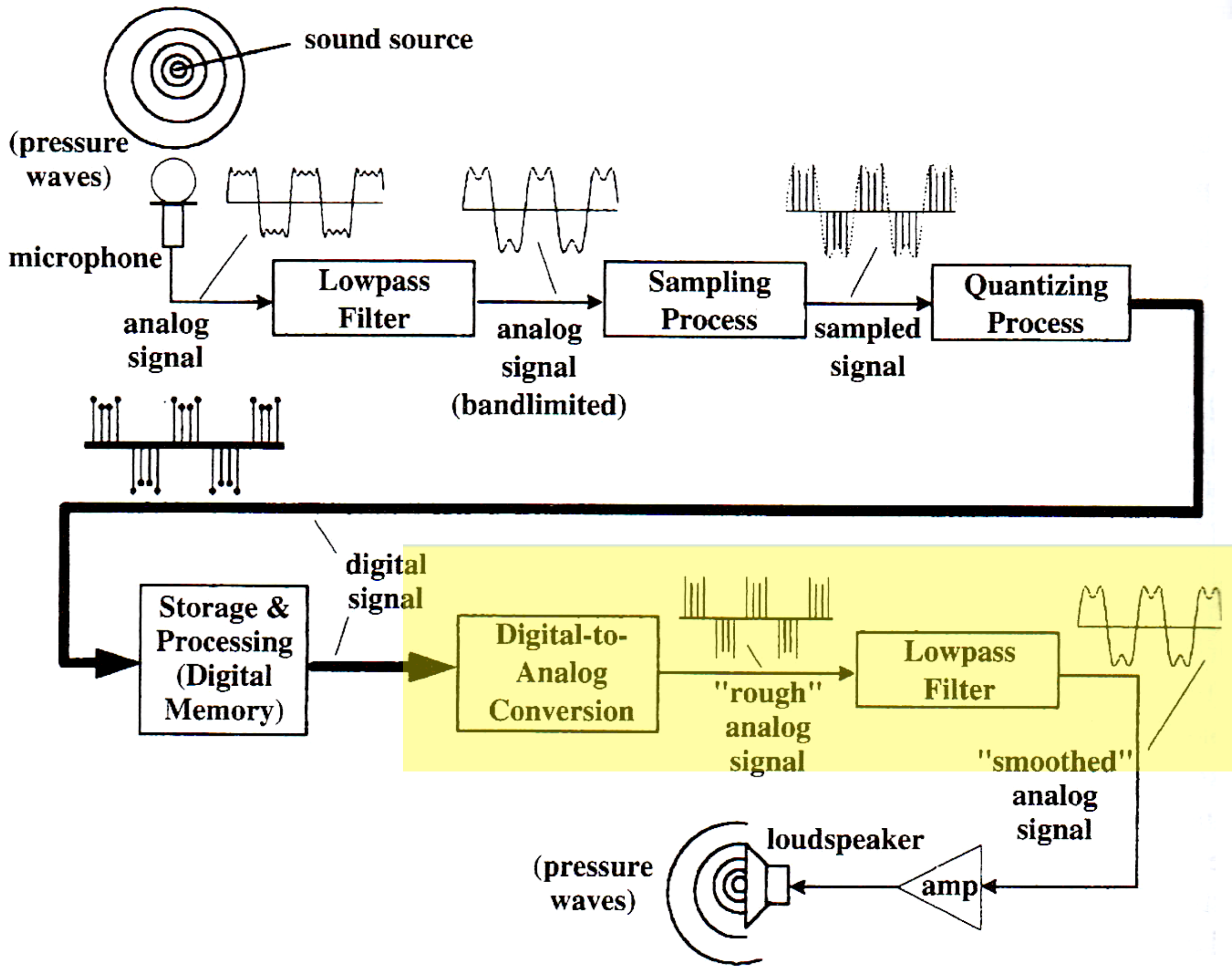
(Note that video frame rate corresponds to the sampling rate)

Sampling Rates

- Determined by the bandwidth of signals or hearing limits
  - Consumer audio product: 44.1 kHz (CD)
  - Professional audio gears: 48/96/192 kHz
  - Speech communication: 8/16 kHz



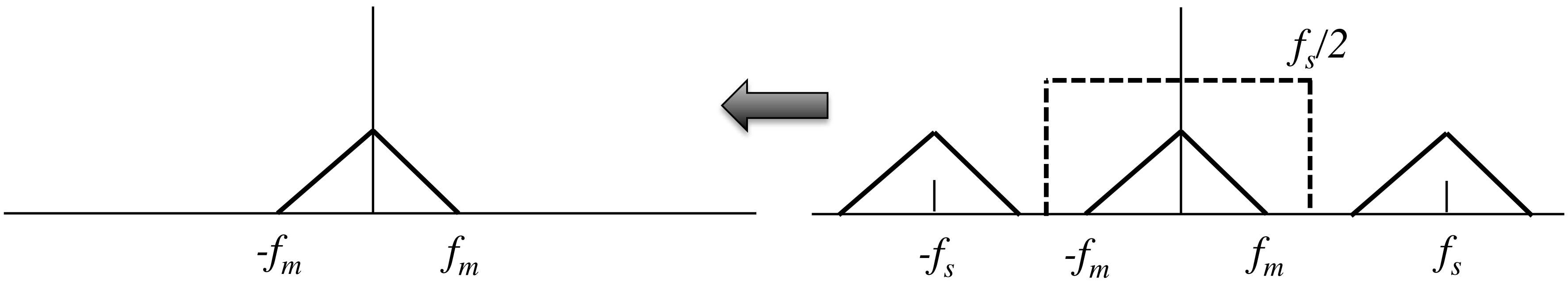
Digital to Analog



…0 0 1 0 1 0 …

Reconstruction in Frequency Domain

- In the view of frequency domain, the signal before sampling (continuous-time) signals can be reconstructed by applying a low-pass filter



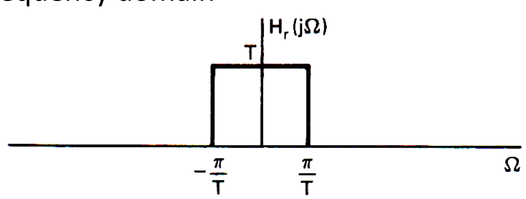
- Conceptually, this is the operation in digital-to-analog converters.
  - In practice, DACs are composed of sample-and-hold and low-pass filtering circuitry

Reconstruction in Time Domain

- In time domain, the reconstruction corresponds to interpolation with the sinc function
  - The ideal low-pass corresponds to sinc function
  - The interpolation is actually convolution with the sinc function

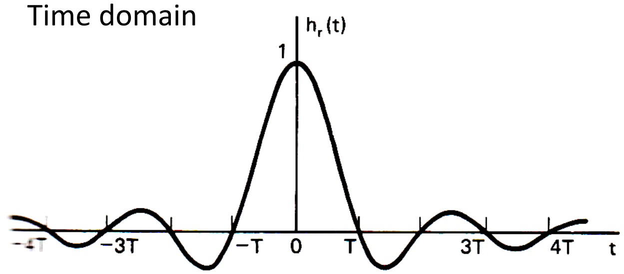
$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

Frequency domain



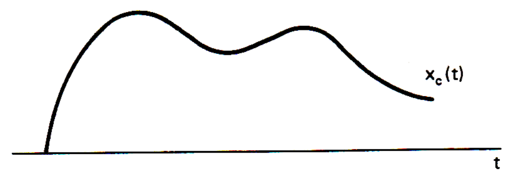
(b)

Time domain



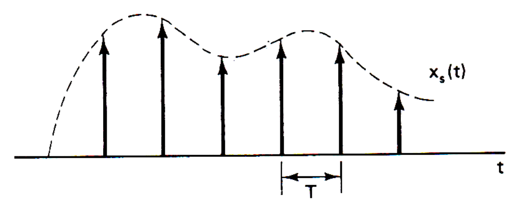
(c)

sinc functions!



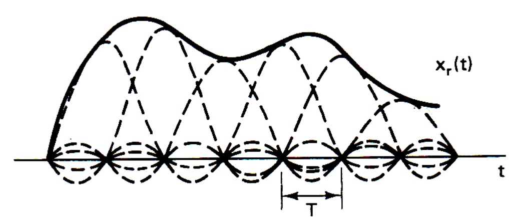
(a)

Before sampling



(b)

After sampling

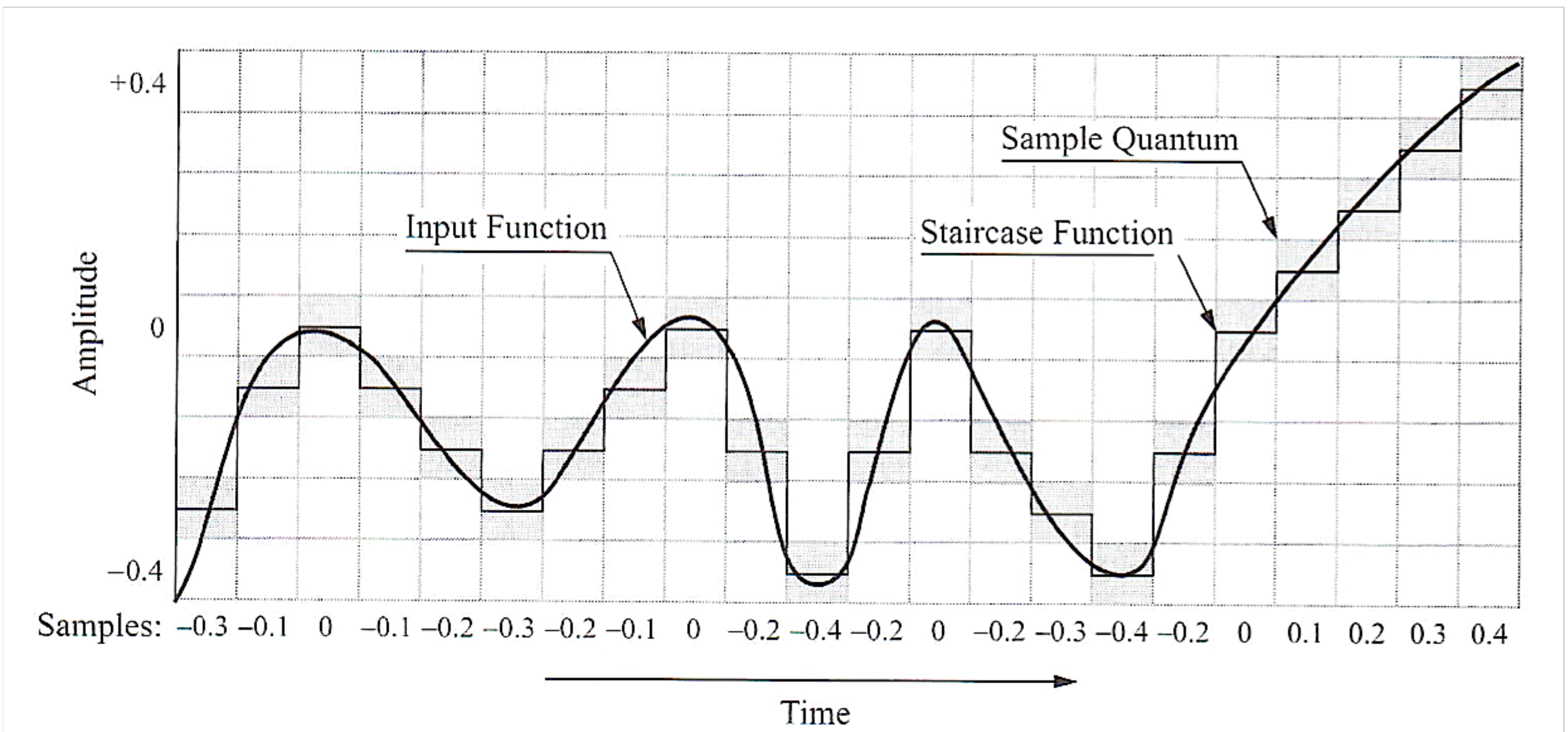


(c)

Reconstruction

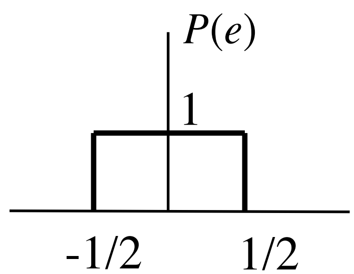
Quantization

- Discretizing the amplitude of real-valued signals
  - Round the amplitude to the nearest discrete steps
  - The discrete steps are determined by the number of bit bits
    - Audio CD: 16 bits ($-2^{15} \sim 2^{15}-1$)



Quantization Error

- Quantization causes noise
  - Average power of quantization noise: obtained from the probability density function (PDF) of the error



Root mean square (RMS) of noise

$$\sqrt{\int_{-1/2}^{1/2} x^2 p(e)\,dx} = \sqrt{\tfrac{1}{12}}$$

- Signal to Noise Ratio (SNR)
  - Based on average power

RMS of full-scale sine wave

$$20\log_{10}\frac{S_{\text{rms}}}{N_{\text{rms}}} = 20\log_{10}\frac{2^{B-1}/\sqrt{2}}{\sqrt{\tfrac{1}{12}}} = 6.02B + 1.76\ \text{dB}$$

(With 16bits, SNR = 98.08dB)

  - Based on the max levels

$$20\log_{10}\frac{S_{\max}}{N_{\max}} = 20\log_{10}\frac{2^{B-1}}{\tfrac{1}{2}} = 6.02B\ \text{dB}$$

(With 16bits, SNR = 96.32 dB)

Dynamic Range, Clipping and Headroom

- Dynamic range
  - The ratio between the loudest and softest levels

Again, RMS of full-scale sine wave for both loudest and softest

$$20\log_{10}\frac{S_{\mathrm{rms,max}}}{S_{\mathrm{rms,min}}} = 20\log_{10}\frac{2^{B-1}/\sqrt{2}}{1/\sqrt{2}} = 6.02B-6$$

(With 16bits, DR = 90.31 dB)

- Clipping
  - Non-linear distortion that occurs when a signal is above the max level

- Headroom
  - Different between the signal level and the max level

In digital audio, 0dB is regarded as the maximum level (dBFS)