# Digital Sound Capstone
# DXARTS 460

## Lecture 2:
## Digital audio fundamentals

Spring 2011

Instructor: Stelios Manousakis

# Analog versus Digital

- Waves in nature, including sound waves (amplitude as a function of time), are *continuous* . The mathematical definition of this term means that between any two points in time, even two points representing a virtually instantaneous interval, an infinite number of amplitude points may be identified.

- ***Analog*** audio (vinyl, tape, analog synthesizers, etc.) involves the creation or imitation of a continuous wave through a continuous fluctuation of electricity voltage. But analog waves are incompatible with computers. Computers can only deal with ***discrete*** values and do not have the capacity to represent continuity (or infinity). ***Digital*** technology is based on converting continuous values to discrete values.

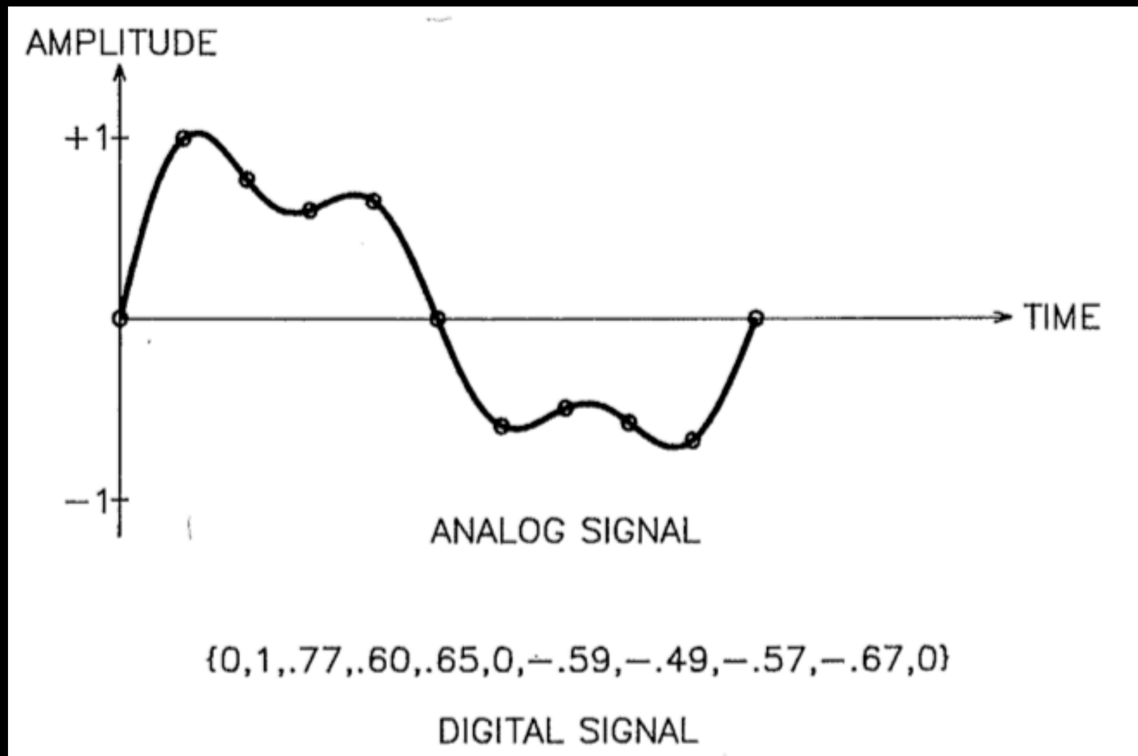
# Sampling

- To enable computer processing of the sound, the analog signal is converted to a sequence of numbers by applying it to an analog-to-digital (A/D) converter (ADC). The conversion process relies on the principle that, at any point in time, an analog electrical signal can be assigned an instantaneous value by measuring its voltage.

- The analog voltage that corresponds to an acoustic signal changes continuously, so that at each instant in time it has a different value. It is not possible for the computer to receive the value of the voltage for every possible instant, because of the physical limitations of both the computer and the data converter.

- Instead, the analog voltage is measured (***sampled***) at intervals of equal duration. The output of the sampling process is a discrete or digital signal: a sequence of numbers corresponding to the voltage at each successive sample time.

# Sampling

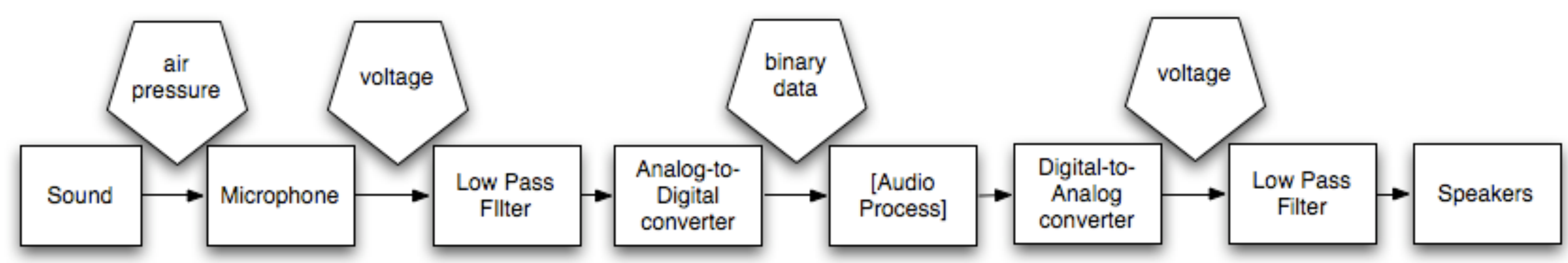



- The digital signal is not continuous because it consists of a sequence of specific values sampled at discrete times. In the literature of engineering, this method of representing an analog signal by a sequence of numbers is known as ***Pulse Code Modulation*** (PCM).

# Acoustic to Digital to Acoustic

- Here is a sketch of the path of a signal:

From the analog to the digital and back to the analog world



# Sampling Rate

- The amount of time between samples is known as the sampling interval or sampling period.

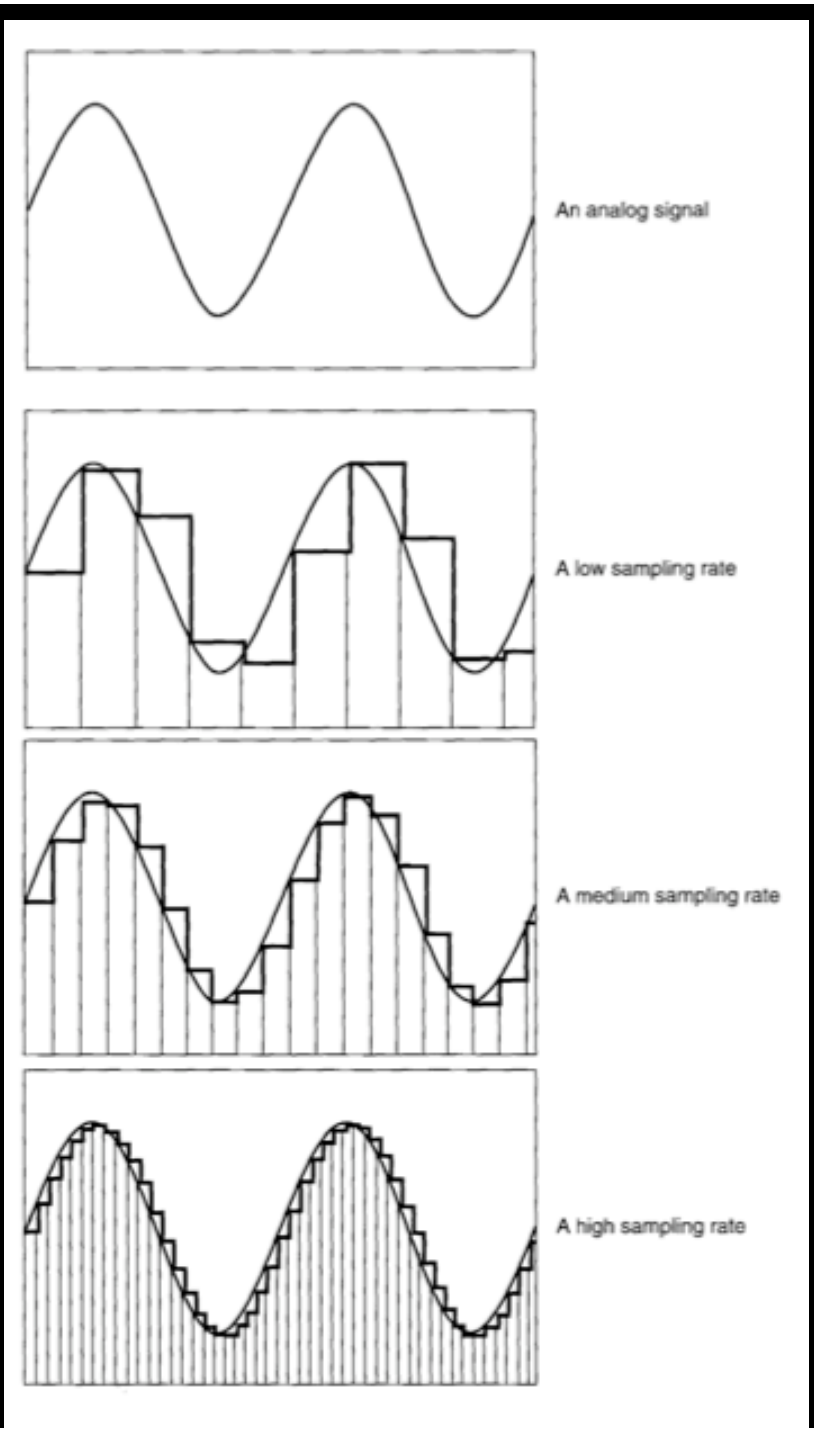
- Its inverse, the number of times the signal is sampled in each segment, is called the ***sampling rate*** or sampling frequency (fs) and is measured in hertz (samples per second).

- The sampling rate describes how often an audio signal is sampled - the number of samples per second. The more often an audio signal is sampled, the better it is represented in discrete form (see Figure) . The question regarding optimal sampling rates is: "How often is often enough?"

- Some standard sampling rates:
  32kHz (used for satellite transmissions)
  44.1kHz (CD),
  48kHz (film),
  96kHz (high quality studio recording),
  192kHz (best available quality currently).

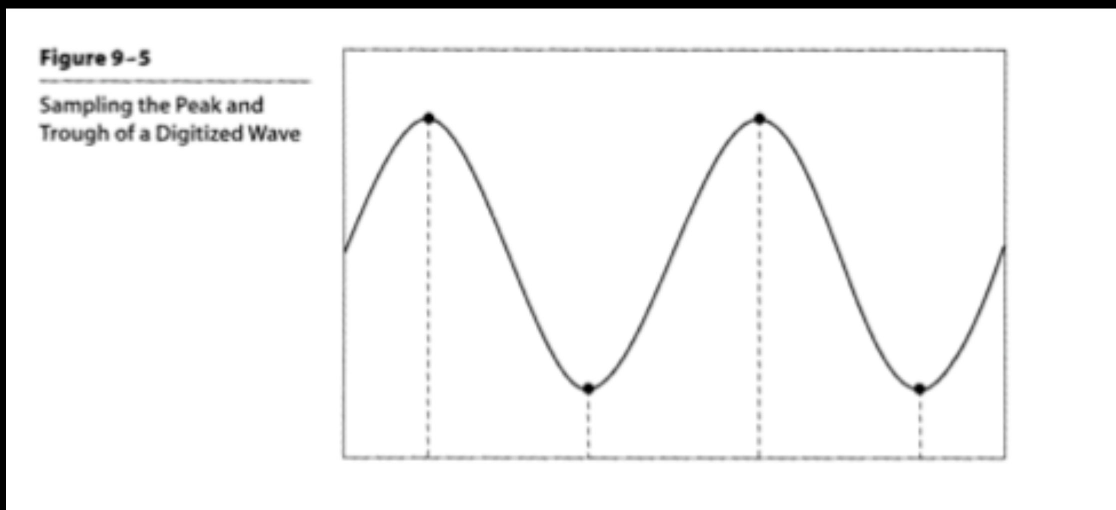
# Sampling Rate

- An analog signal, represented with different sampling rates



# Sampling theorem

- The ***Nyquist*** theorem (or Sampling Theorem):
  *To represent digitally a signal containing frequency components up to X Hz, it is necessary to use a sampling rate of at least 2X Hz.*
- In other words, it is sufficient to sample both its peak and its trough.




- Conversely:
  *The maximum frequency contained in a signal sampled at a rate of SR Hz is SR/2 Hz.*
  The frequency SR/2 Hz is also termed the ***Nyquist frequency***.

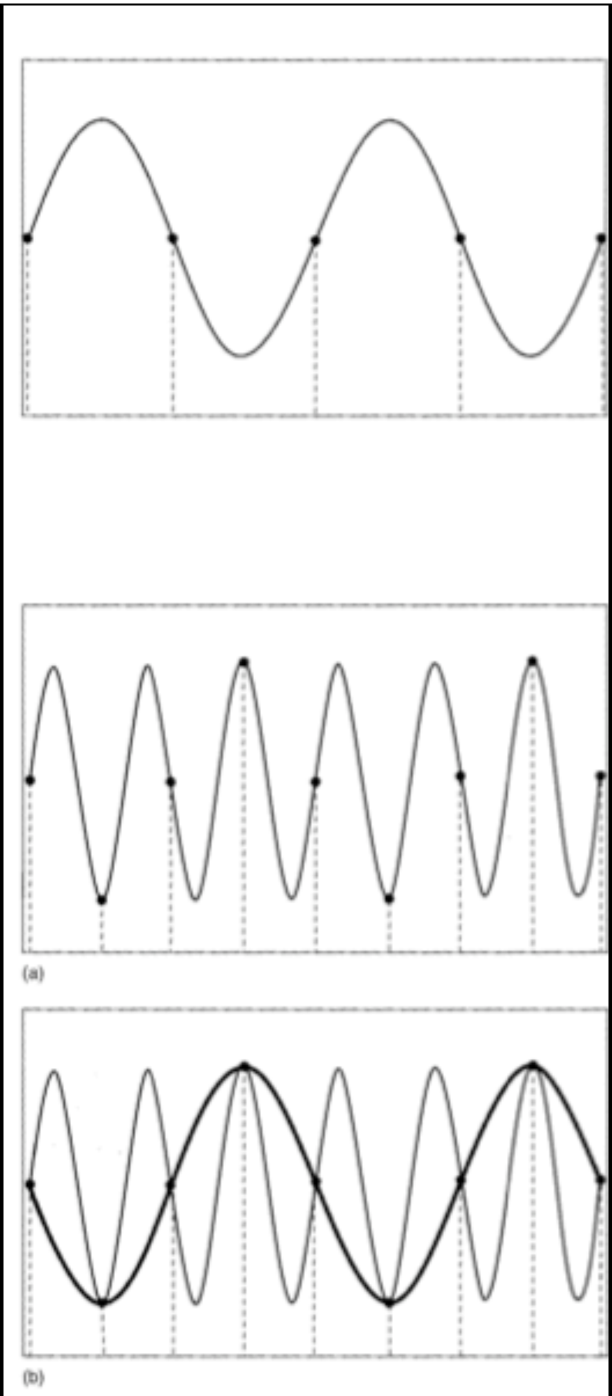
# Aliasing

- Misrepresented frequencies are termed ***aliases****.* The alteration caused by the sampling process of frequencies higher than the Nyquist frequency is known as ***aliasing*** or ***foldover***.

- In general, if a frequency F sampled at a sampling rate of SR exceeds the Nyquist frequency, that frequency will alias to a frequency of ***- ( SR - F)*** where the minus sign indicates that the frequency phase inverted.

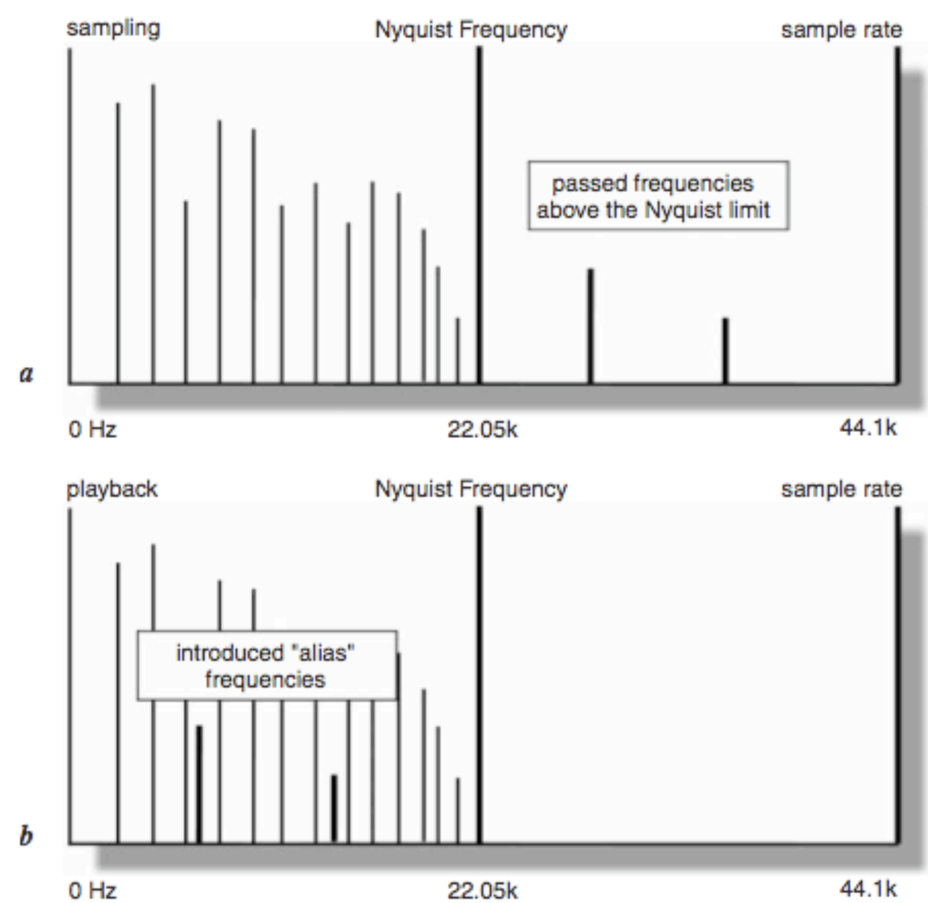
- Figure:
  a) An Audio signal of 30 kHz, sampled at 40 kHz.
  b) The frequency is misrepresented at 10k Hz , at reverse phase.

- To prevent aliasing ADCs and DACs have embedded Low-pass filters



# Aliasing

- A frequency domain representation:



# Sampling:
## Over- | Under- | Critical

- Therefore, in theory, since the maximum audible frequency is 20kHz, a sampling rate of 40kHz would be sufficient to recreate a signal containing all audible frequencies

- Most audible frequencies are ***oversampled,*** meaning that the audio frequency is below the Nyquist frequency

- A signal sampled at precisely the Nyquist frequency is said to be ***critically sampled*** . Sampling at exactly the Nyquist frequency runs the risk of missing peaks and troughs and only sampling the zero crossings

- A far more significant problem is that of ***undersampling***, recording a frequency greater than the Nyquist frequency (see next Figure)

# Binary numeral system

- Represents numerical values using two symbols, commonly 0 and 1

- Can be used by any mechanism capable of two different states.
  Is the basis of digital data storage, processing and transmission, as it is easy to implement electronically (ON/OFF).

- One of many numerical systems (e.g, decimal, octal, hexadecimal). Exists for about 2-2.5 thousand years.

- Each digit represents an increasing power of 2 (in the same way as in decimal each digit represents an increasing power of 10):
  The rightmost is (digit) * $2^0$, the next (digit) * $2^1$, then (digit) * $2^2$ etc

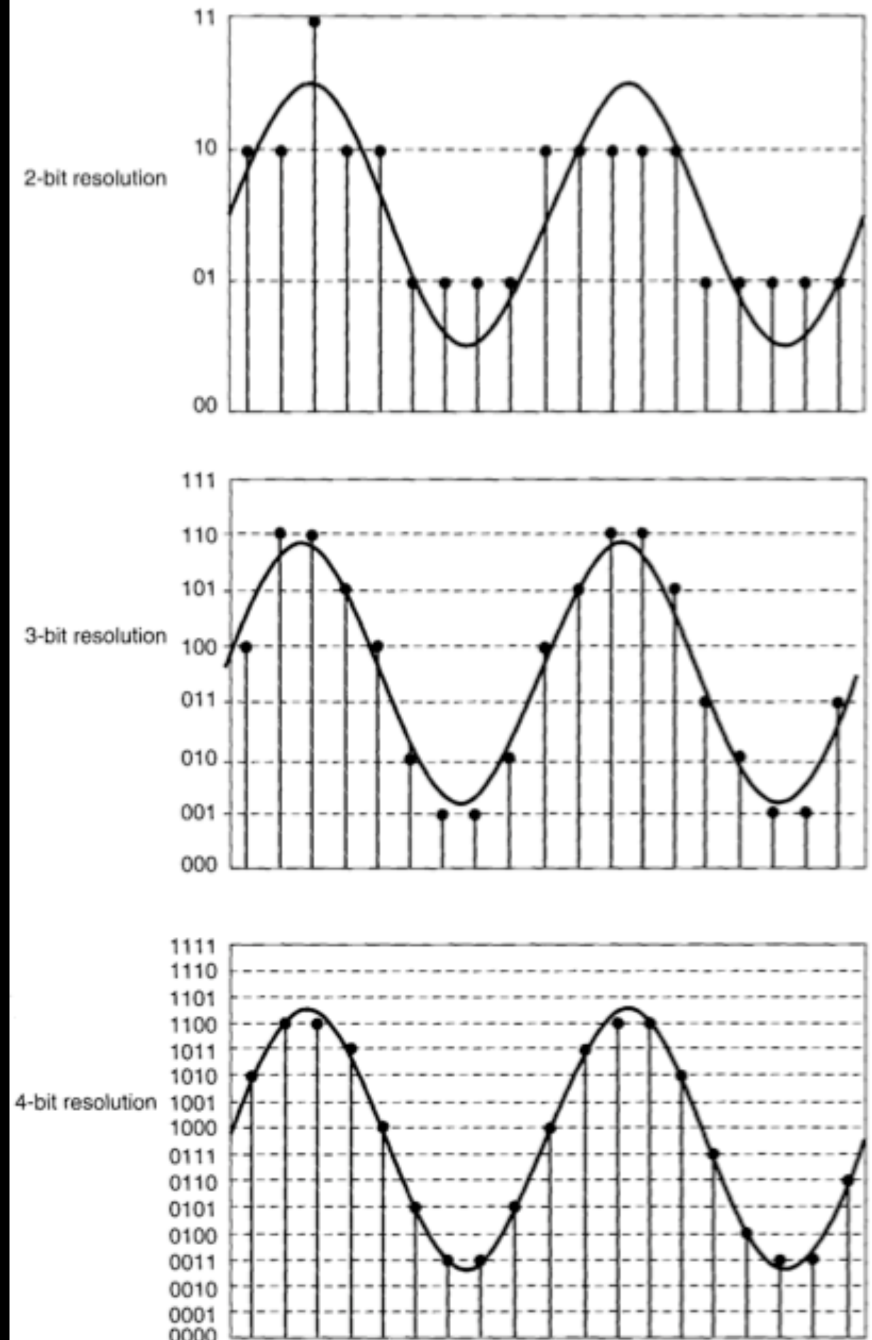
- E.g: The 4-bit number 0101 = (0 * $2^{3)}$ + (1 * $2^2$) + (0 * $2^1$) + (1 * $2^0$) = 5

# Bit Depth | resolution

- Sampling rate considers only how *often* the amplitude of the wave was measured. But it says nothing of how *accurate* these measurements actually are.

- The accuracy of the measurement depends on how many bits are available to represent these values. Clearly, with more bits, the **resolution** becomes finer.
  The ***bit-depth*** or ***bit-rate*** corresponds essentially to the number of values that the A/D converter uses to describe a signal. E.g:
  8-bit = $2^8$ = 256 values;
  12-bit = $2^{12}$ = 4096 values;
  16-bit = $2^{16}$ = 65536 values (−32768 to 32767)
  24-bit = $2^{24}$ = 16,777,216 values
  32-bit = $2^{32}$ = 4,294,967,296

- The  group of bits uses to describe a single sample value is called a ***word***; thus, the bit-depth also referred to as ***wordlength***



# Dynamic Range

- The ***bit depth*** determines the *dynamic range* of a signal.
- Dynamic range is the ratio of the strongest to the weakest sound that can exist in a signal
- Dynamic ratio is measured in decibels, which is a logarithmic scale (more on that in a future lecture)
- In an electronic system, it is measured as the ratio between the background noise of the electronic components and the loudest sound that can be represented without distortion.
- Here are the theoretical ranges of some standard bit-depths:

  8-bit word = 49:8 dB
  16-bit word = 97:8 dB
  20-bit word = 121:8 dB
  24-bit word = 145:8 dB
  32-bit word = 193.8 dB

# Signal-to-Noise Ratio

- A characteristic associated with *dynamic range* is the ***signal-to-noise ratio*** (SNR), which compares the level of a given signal with that of the *noise* in the system. This is sometimes called **Signal-to-Error** in the digital domain

- The ***SNR*** for a given signal varies depending on the nature of the audio contents. This error ratio may be acceptable for high signal levels. However, low-level signals do not use all available bits, and therefore the signal-to-noise ratio is greater.

- Unlike the constant hissing noise of analog recordings, ***quantization error*** is correlated with the digital signal, varying with the signal level, and is thus a type of *distortion*, rather than *noise.*

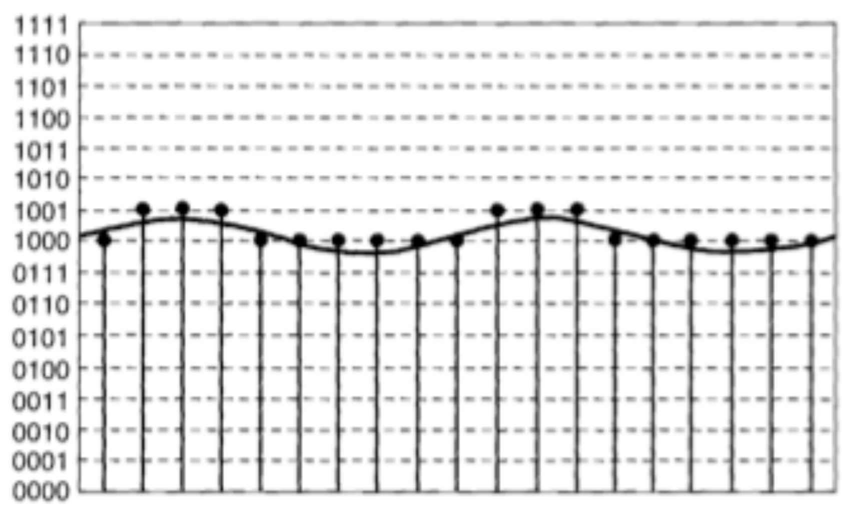



**Figure 9–12**

**Quantization Error**

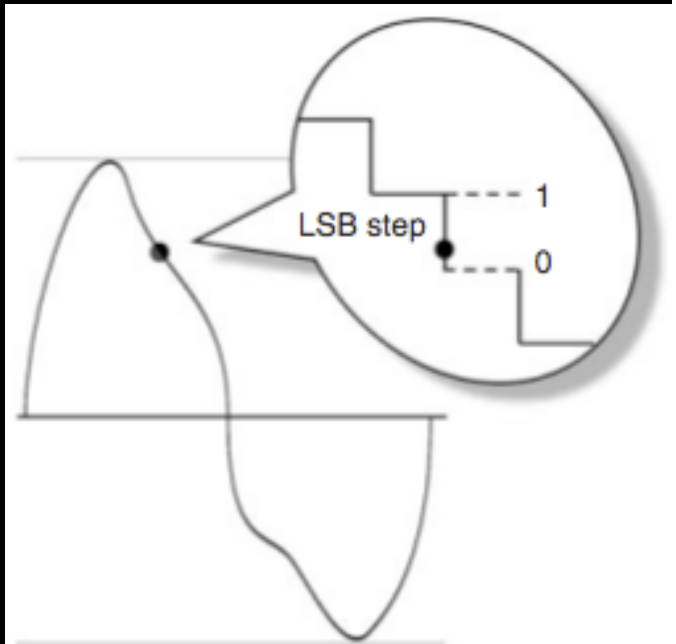
Worst case: a sine wave fluctuating within one quantization increment is stored as a square wave.

# Dither

- As opposed to analog audio, digital audio is discrete both in time (sampling rate) and dynamics (bit-depth). These steps are quite small, and to most people inaudible. However, the do add a tiny square-wave component on the Least Significant Bit (LSB), which in its turn adds low-level harmonic distortion to the original signal.

- To counteract this square wave, we can add a small amount of *perceptually encoded noise* (or a similarly functioning periodic signal) to *statistically improve the resolution below the LSB*. Meaning, resolution can be improved below the  LSB level!

- Dither should be used whenever a high-resolution signal is reduced in resolution, as a quantization error will be introduced into it.
  BUT: This conversion should only happen once, as a final stage when rendering.



# Conversions

- Avoid unnecessary sampling-rate and bit-depth conversions, as each time you do that, there is a bit of distortion introduced.

- In the case of upsampling or increasing the bit depth, no information is added by the conversion. These processes simply increase the size of a file.

- ESPECIALLY avoid downsampling or lowering the bit-depth, as these are *destructive* processes – meaning information will be lost permanently. Only do these conversions as a last step when rendering.

# File formats

- Digital audio consists of binary data, which have to be organized (encoded) in specific ways (protocols).
- Many different audio formats exist. Some are ***uncompressed*** (or, to be technically accurate, they use *lossless compression*), and others are **compressed** (*lossy compression*)
- When working on a project, you should ALWAYS use uncompressed formats (i.e. no mp3s). If you need to compress a file (e.g. for internet broadcast), do this on your final mixdown uncompressed file
- The two most widespread uncompressed audio file formats are:

  \- **wave** (.wav or .WAV): Microsoft Windows format

  \- **aiff** (.aif, .aiff, or .snd): Apple format (Audio Interchange File Format)

  They both use PCM (Pulse Code Modulation)

# Summary

- Analog signals are continuous, digital are discrete
- Digital signals are represented with bundles of binary numbers (words)
- The sampling rate (i.e., how many words per second) determines a signal's frequency content.
- Aliasing can be introduced by misrepresentation of frequencies
- The bit-depth determines the accuracy of representation of a signal, its dynamic range and the signal-to-noise ratio
- Always use uncompressed file formats when working with audio

# Bibliography


- Ballora M. (2002). *Essentials of Music Technology*. Prentice Hall, 2002.
- Dodge, C. and Jerse, T. (1997). *Computer Music Synthesis, Composition, and Performance (2nd ed.).* New York: Schirmer Books
- Huber, D. M., & Runstein, R. E. (2005). *Modern recording techniques.* Boston: Focal Press/Elsevier
- Hugill, A. (2008). *The Digital Musician.* New York, NY: Routledge
- Katz B. (2007). *Mastering Audio: The Art and the Science* (2nd ed). Oxford: Focal Press
- Moore, F. R. (1990). *Elements of Computer Music*. Englewood Cliffs, NJ: Prentice Hall


*NOTE: All the images were taken from these books and the internet.*