The engineering aspects of an audio system

Gary Tuttle

gtuttle@iastate.edu

http://tuttle.merc.iastate.edu/eClubs/homepage.htm

What are the parts of a basic audio system?

- Stored source material
- Digital to Analog conversion
- Transmission to an amplifier
- Amplification
- Electrical to acoustical transducer (a speaker)
- Transmission through the air to the listener's ear

Design considerations

- Power / efficiency
- Distortion
- Frequency response & signal processing

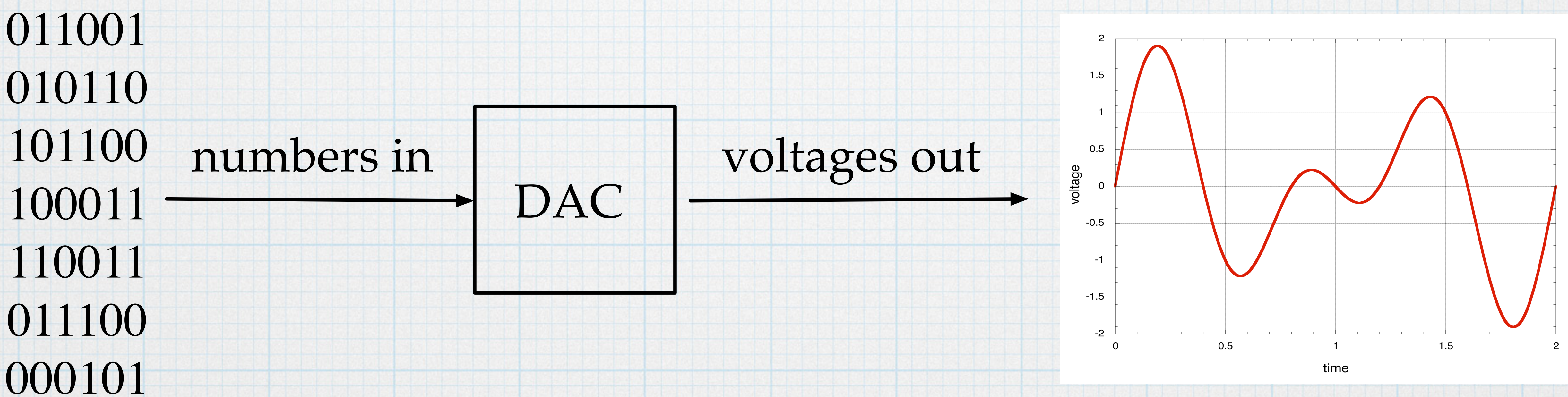
Some important facts

- Our ears are analog — they respond to a continuous series of pressure vibrations arriving from the surrounding air.
- Most modern data storage is digital — numbers stored in some form of electronic memory element.
- Humans can hear vibrations that are oscillating at frequencies between 20 Hz and 20,000 Hz. Frequencies below are “infra-sonic” (elephants) and those above are “ultra-sonic” (dogs). All aspects of an audio system should work properly for all frequencies in this range. (Note though the human range of hearing changes with age.)
- Because we have two ears, we are able to “image” sound and determine the direction that it is coming from. (At least for some frequencies.)
- The range of “loudness” (acoustic power) that we can detect and tolerate is huge. If we assign the smallest possible sound that a typical human can hear as 1, then the loudest that we can handle is 10^{12} (one trillion). (Above this, our ears may be damaged.) Using technical jargon, the range is 120 dB.

Audio storage

Digital-to-analog conversion

- Necessary to go from modern data systems to ancient analog ears.
- Can happen at different points along the audio system path, but must happen somewhere.
- Requires an electronic circuit — digit-to-analog converter (DAC). (EE 230)
- Converting in an effective manner requires some signal processing. (EE 224 and up).



- Also, often the digital data is stored in a “lossy" format (mp3, aac, or similar) and must be converted back to a native format before using the DAC. (Digital signal processing, EE 424).

Transmission to the amplifier

- It used to be that this was always a wire. Still true for many audio systems.
- However, an alternative method has become available in recent years — wireless transmission using Bluetooth, Wifi, or some other proprietary communication scheme (eg Sonos). With modern wireless systems, the data is transmitted in digital form, so the D-to-A conversion takes place *after* transmission.
- Requires all the elements of a standard communication system — transmitter and receiver circuits and antennas at both ends.
- There are analog transmission systems still in use — AM and FM radio.
- There are also optical fiber transmission systems which send the audio info via light pulses. These also require transmitters and receivers.

The amplifier

- The signals coming from the source are too weak to drive typical speakers. Making a “normal” desktop or living-room speaker give out some sound requires at least a few tenths of a watt of power. The audio signal from a DAC might have only a few milliwatts of power available. The amplifier adds power to the signal. (Remember: $P = V \cdot I$, so both voltage and current must be boosted.)
- The amp takes energy from a DC power supply and adds it to the audio signal. The DC power can come from a battery or from an AC-to-DC converter that is plugged into a wall outlet.
- Consideration: total power available. More power requires bigger power supplies and bigger components. This adds to cost and size.
- Consideration: efficiency. If running from a battery, efficiency is crucial. Inefficient amps may require significant heat sinking to protect from getting too hot. This also adds to cost and size. (class A vs. class B vs. class D.)
- Consideration: distortion. Amplifiers are inherently non-linear, meaning that the signal might be changed as it passes through the amplifier. Usually, this is bad. It is expressed in terms of total harmonic distortion. (THD)
- Consideration: frequency response. The amp should amplify all frequencies the same amount.

The speaker (acoustic transducer)

- Once the audio signal has enough power (voltage and current), it can used to drive a speaker.
- The speaker is essentially a liner motor, converting electrical energy into mechanical energy and then into acoustic energy. https://www.youtube.com/watch?v=AP2Nu4MZJRs
- Speaker design requires consideration of several inter-related topics: filtering of the audio signal, magnetics, mechanical vibration, and interference of sound waves. For something that is so common and outwardly simple, the detailed operation is quite complex.
- Consideration: total power. The coil, magnet, and speaker cone must sized to match the desired acoustic power output — more power = bigger (and more expensive) components.
- Consideration: distortion. Like with the amplifier, the speaker should not change the signal as it is changing from one form to another.
- Consideration: multiple drivers. Because a single driver cannot produce all frequencies (due to the limits of mechanical vibrations), a speaker may have multiple drivers of different sizes. Each size will produce a particular frequency range. In the case, the electrical signal has to be modified so that only the appropriate frequencies are passed to each driver.

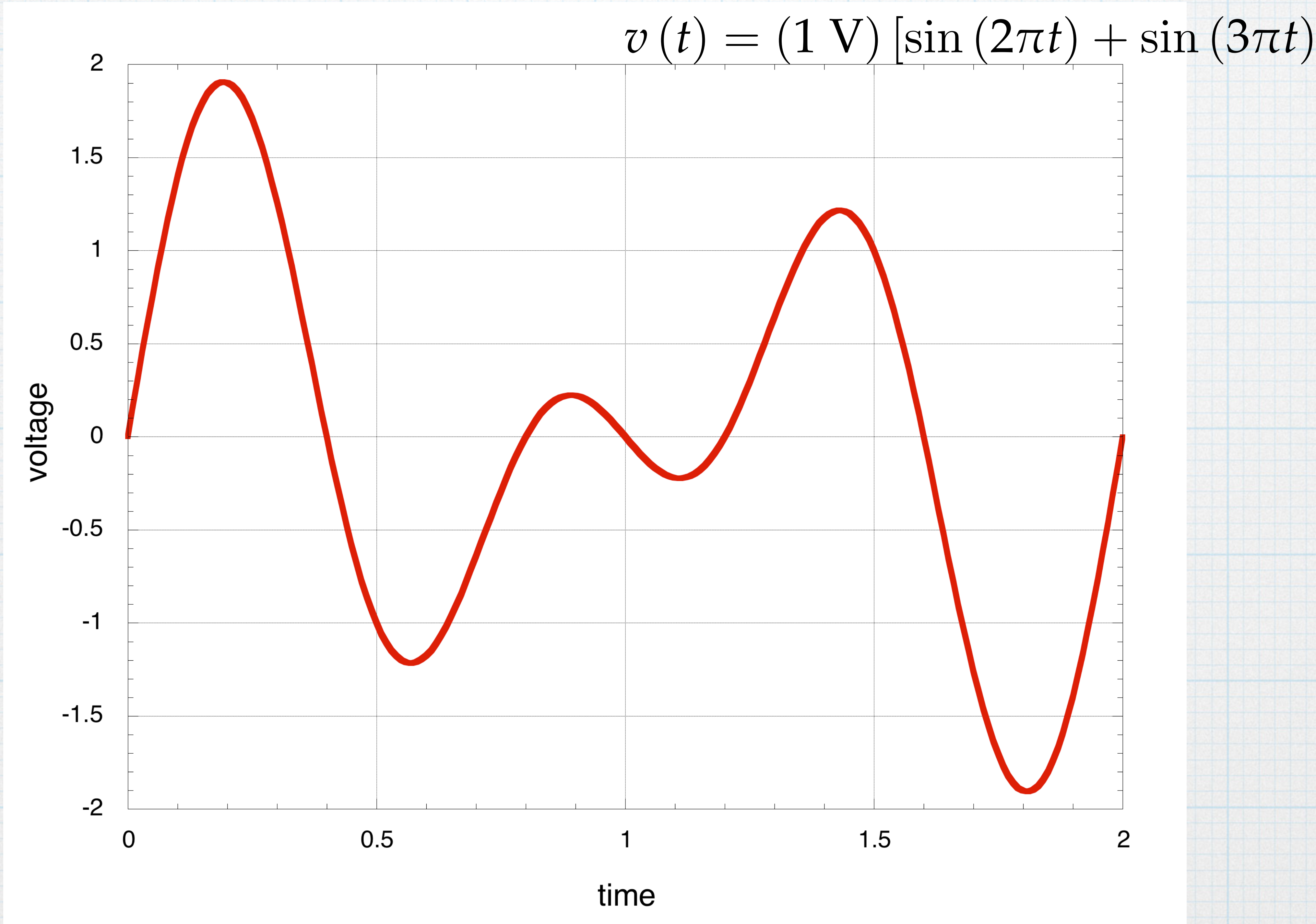
Sound waves

- The energy travels through the air as sounds waves — moving pressure variations.
- All waves have a characteristic wavelength that is related to the frequency and speed that wave at which the wave travels, $\lambda = v / f$. (v = 343 m/s = 767 mph). At f = 20 Hz, λ = 17 m = 56 ft. At f = 20 kHz, λ = 1.7 cm = 0.67 inches.
- The three-orders-of-magnitude difference in the wavelengths has important implications for how we perceive the sounds. The short wavelengths of the high-frequency sounds mean that they can be viewed as traveling almost like rays. (We often treat light waves as being ray like.) This means that we can discern the direction the high frequency sounds comes from. The long wavelengths of low frequencies make determining directions impossible.
- Speakers are most effective when their physical size is commensurate with the wavelengths they are producing. While a 56-ft diameter speaker is not practical — although it would be awesome! — the general trends hold. Small diameter speakers are used to produce higher sounds and bigger speakers produce lower sounds. A three-way system is common, a small "tweeter", a medium-sized "mid-range", and a big "woofer" or sub-woofer.
- The wave nature of sound also means that reflections will lead to constructive and destructive interference. So the shape of a room — and hence the shape of the reflections — can have big impact on what we hear.

Stereophonic sound

- The fact that we have two ears separated by the width of our head means that sounds reaches the two “sensors”at slightly different times. We can use the time difference to determine the direction that the sounds comes from. At least for higher frequencies.
- When listening to a live performance, we can discern where the drums are, where the strings are located, and where the horn section is seated on the stage.
- We can use two speakers (stereo) when playing recorded music to “fake” the effect of listening to a live performance. It is easy to separate the various sounds from a live performance and record those into two separate “channels” — some sounds go into the “right” channel and some into the “left”. If the two channels are played back through two speakers that are spaced apart in front of us, we will hear slightly different sounds coming from each speaker. Our brain will combine the sounds in such a way that we perceive a spread out orchestra in front. It is an illusion, but a very potent one. Almost all music is recorded in “stereo” for this reason. (If the sounds are not separated, then the audio is said to “mono”.
- The effect can be extended to more speakers — a quadraphonic system or surround sound that comes with video. But stereo is the most common for straight audio applications.

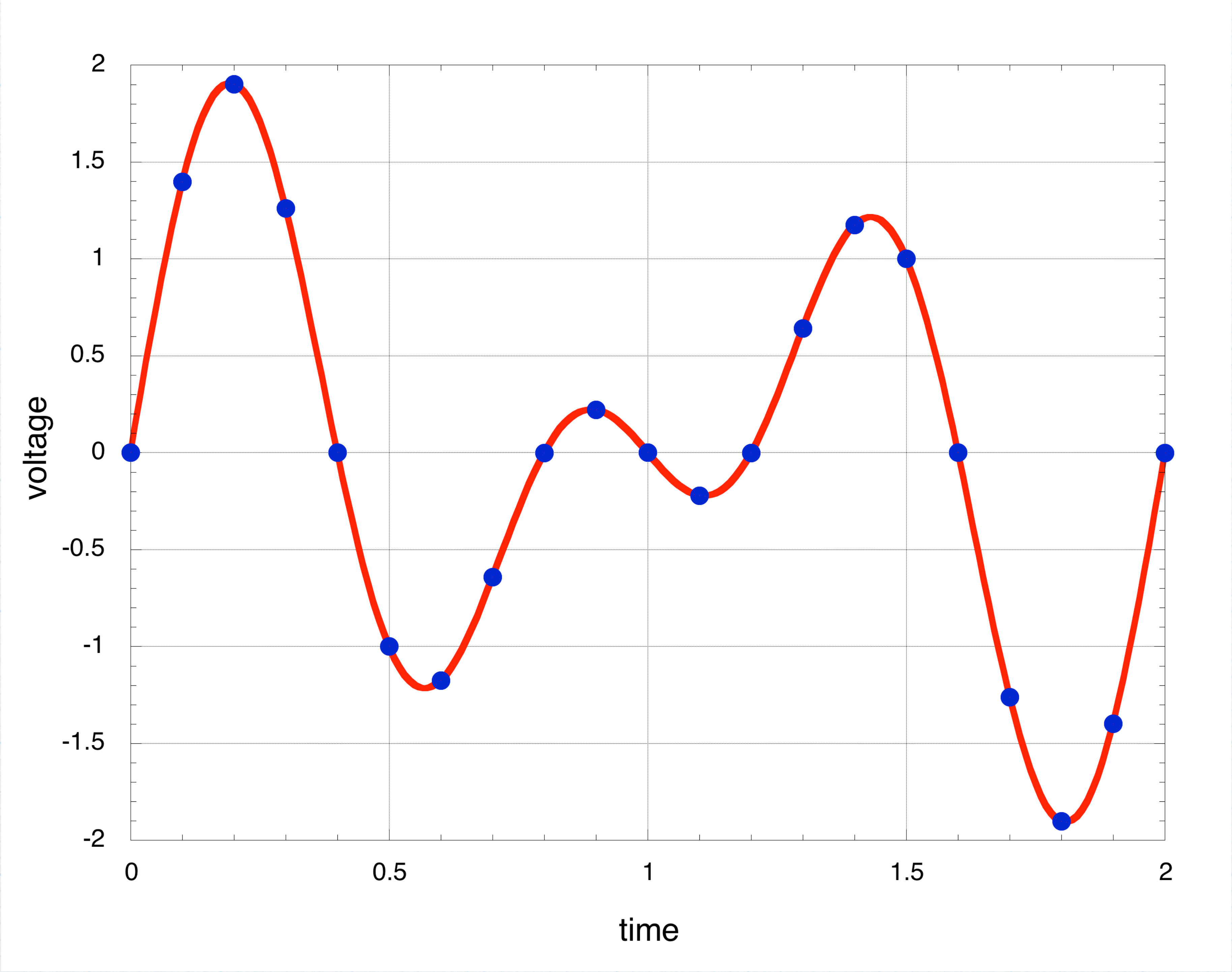
An analog signal



Defined for all times. (Continuous time.)

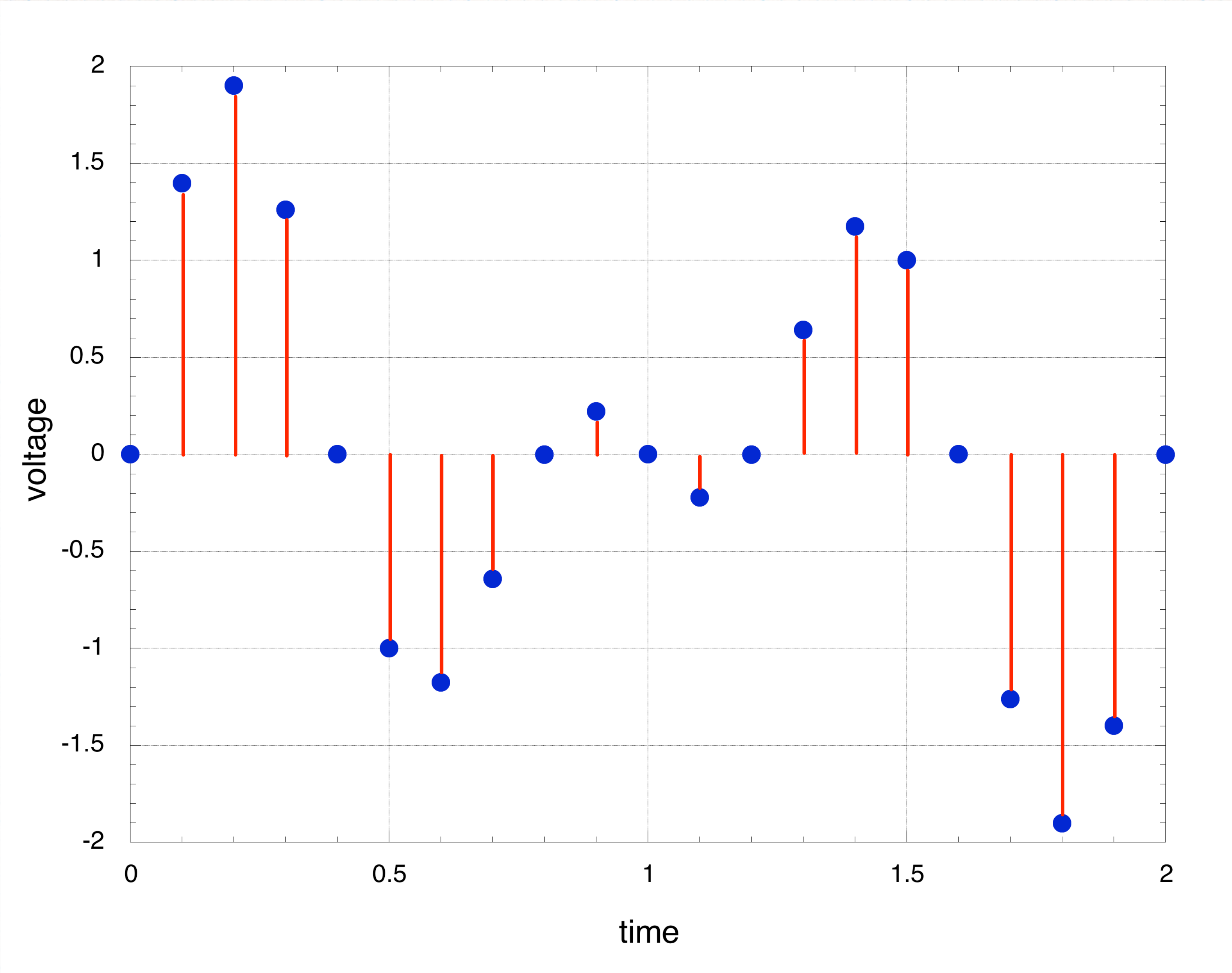
Any voltage is allowed.

Sampling



Measure the voltage every Δt. In this case, $\Delta t = 0.1$ s. (Sampling rate of 10 samples/sec.) All voltages are still possible, but we are looking only at specific discrete points in time.

Sampled (discrete) analog signal



Lollipop graph.

Ignore everything between samples. This is a *discrete-time* representation of the analog signal. But it is still a completely accurate representation of the original signal. As long as the the sampling is above the *Nyquist rate*.

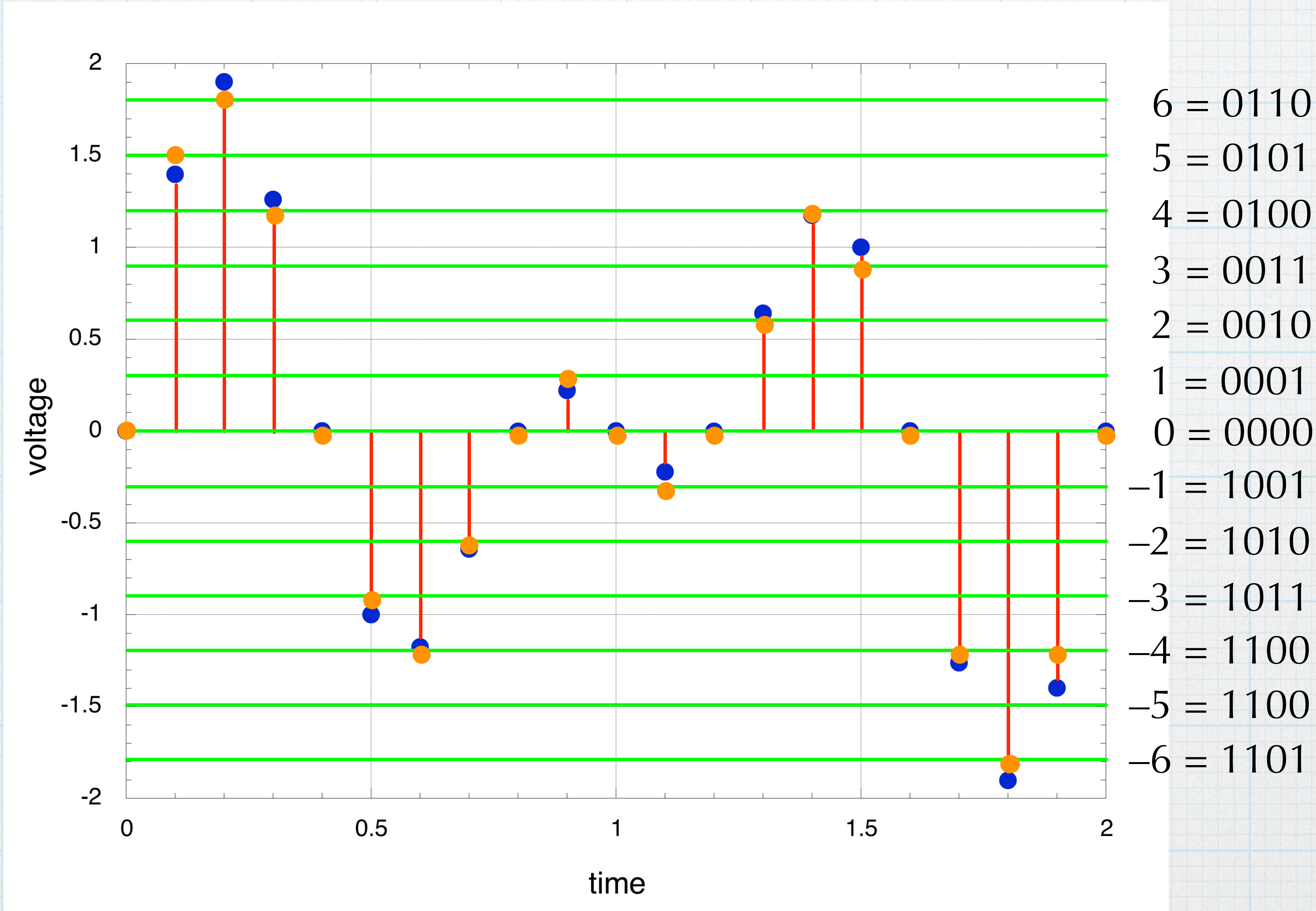
Nyquist rate

The Nyquist rate is defined as twice the maximum frequency of a band-limited signal. As long as the sampling rate is at least twice the highest frequency in a signal, then the complete original analog signal can be uniquely reconstructed from the from the set of sampled data.

Stated another way: If sampling rate is above the Nyquist rate, there will be only one possible continuous function that will fit the sampled data.

In our example, there are only two frequencies in the signal, $f = 1$ Hz and $f = 1.5$ Hz, so the Nyquist rate is 3 Hz. We sampled the data at 10 Hz — 5 samples per second. We are safely above the Nyquist rate.

Finally, digitize the voltage values



Slice the voltage range into equal-sized chunks. Represent each chunk with a binary number. In this case, we used 4-bit signed binary numbers (–7 to +7), with each chunk 0.3 V wide. There may be some *quantization error.*

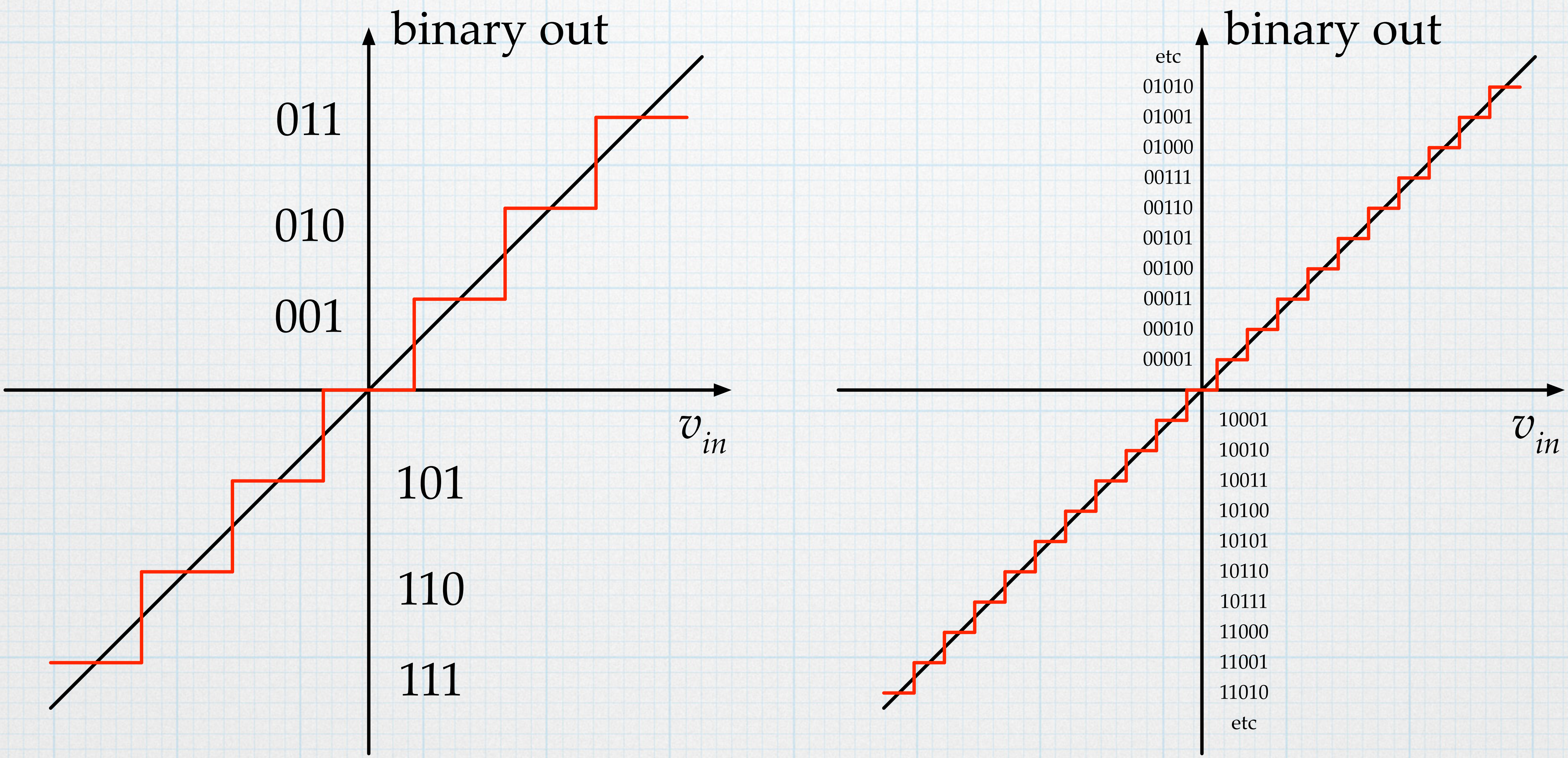
Data converters

While the sampling and digitizing steps are distinct conceptually, the two steps are usually done simultaneously using a circuit known an *analog-to-digital converter* (ADC).

There are a number of methods for doing analog-to-digital conversion. One of the simplest and fastest methods is use a stack of comparators (flash converter). Recall from EE 230 that a single comparator can be viewed as a one-bit ADC. However, higher bit-depth flash converters require a large of number of comparators, which is a disadvantage. At higher bit depths, alternative topologies, like *dual-slope* converters are preferred. (See Prof. Google or take EE 505).

At the other end of the audio system, the digital audio data must be put back into analog form so that our ears can hear it. This requires a *digital-to-analog converter* (DAC). Most DAC circuits are straight-forward implementations of summing-type op amp circuits.

Quantization error can be reduced by increasing the *bit depth*, meaning to use more bits to represent the range. With more bits, the voltage slices are smaller, and the quantization errors will be smaller. So this is good. However, more bits means more digital data to store.



Most consumer-grade digital audio equipment uses a bit depth of 16 bits (-32,768 to +32,768, if using a simple signed binary representation).

Storing digital audio

Increasing the sampling rate will also increase the amount of data that must be stored. However, using a sampling rate above the Nyquist frequency does not improve the reconstruction of the original signal. However, higher sampling rate will help reduce some of the noise that produced when a signal is digitized. So higher sampling rates are useful up to a point.

For audio, we assume that maximum frequency that we need to reproduce is about 20 kHz — the limit of human hearing. So we can choose to *band limit* the signal to 20 kHz and then sample at a frequency somewhere above 40 kHz.

When musics CDs were being developed in the late 1970s, the sampling rate was chosen to be 44.1 kHz.

2 channels x 16 bits/sample x 44,100 samples/sec x 60 sec/min x 4 min/song ≈ 340 Mbits = 42 Mbytes per song. CDs were standardized to hold up of 74 minutes of music (= 783 Mbytes), about 20 songs max.

Lossy storage

Storing all of the samples of an audio signal on CD still required a lot of physical space.

In the 1980s, a number of lossy encoding techniques were developed. The most famous is the MP3 format.

The idea behind a lossy data format is that we don't need to hear every single bit of every single sample in the data file in order to hear it "accurately enough". Much of the stored data is either useless or redundant.

By cleverly removing the unnecessary data, the amount that that is stored can be greatly reduced. This is an excellent example of *digital signal processing* (EE 424). With MP3, the reduction is usually about a factor of 10, meaning that a four-minute song that required around 40 Mbytes on a lossless CD can be reduced about 4 Mbytes. So a CD encoded as MP3 could hold 200 songs! Of course, with the increase in digital transmission speeds brought on by the internet, we don't need physical storage anymore. Even the CD is a quaint notion.

A few of links with background info

The first two links go to nicely-produced videos that present a lot of information in a short period of time. The other links are to Wikipedia articles, which seem to be decent. (Although I have not read every word of those.

https://xiph.org/video/vid1.shtml

https://xiph.org/video/vid2.shtml

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

https://en.wikipedia.org/wiki/Digital-to-analog_converter

https://en.wikipedia.org/wiki/Analog-to-digital_converter