

3.17.1.2: Recording

Vinyl

As mentioned above, earliest sound recordings were made without the use of electronics at all; grooves made by a vibrating needle were made in a cylinder of hot wax and these recorded the oscillations of sound being funneled through a large cone. Once sound was recorded on the cylinder it was cooled to solidify the wax or plastic. To play the recorded sound, the needle was returned to the beginning of the now cool cylinder and the movement of the cylinder was duplicated. Now, however, the fluctuating groove would cause the needle to vibrate which would cause the diaphragm to vibrate, recreating the sound.

Cylinder recordings were popular from the 1880s until the 1920s when they were gradually replaced by disk shaped **vinyl records**. Using basically the same idea as the cylinder recording, a rotating disk made of vinyl recorded the fluctuating groove. Electronic amplification of the needle vibrations eventually replaced the mechanical diaphragm and cone system to reproduce the sound. Motion of the needle is detected by a coil and magnet system, again using Faraday's law. A metal copy of the original disk is made and used as a mold to make multiple copies of the same record. An electron microscope picture of the grooves in a vinyl record is shown below. Vinyl records are analog recordings; the grooves in the vinyl vary the same way the original sound did. Faster vibrations of sound produce grooves with more oscillations and louder sounds produce bigger oscillations.

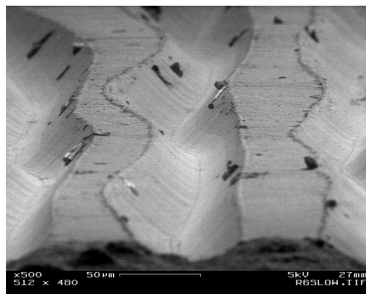


Figure 3.17.1.2.1

For **stereo recordings**, each side of the groove records the fluctuations coming from a different microphone. It is also possible to record four different fluctuations from four microphones in a single groove, a process known as quadrasonic sound recording.

Vinyl records have several disadvantages as a sound recording medium. They are somewhat fragile in that they can easily be scratched, broken or melted. Leaving a vinyl record in your car on a sunny day (or even in a sunny spot in front of a window) generally means you will not be able to play it again due to warping of vinyl in the heat. Stacking many records on top of each other or under books will also warp vinyl records. Scratches on the vinyl will be translated into sounds as hiss and pop which interfere with the recorded music. Records must be kept dust free to avoid having the needle skip over dirt in the grooves. Over time the needle will wear away the vinyl, reducing the accuracy of the recording. The needle also deteriorates over time; even the best diamond needles eventually wear out and have to be replaced.

Video/audio examples:

- Wikipedia on [vinyl records](#).
- Web page with more [electron microscope pictures of grooves in vinyl](#).
- Somewhat long YouTube on [how to fix a scratch on a vinyl record](#).

Sound sample of a scratch vinyl record:

Tape

Magnetic tape recording was developed in Germany before the Second World War but was not available commercially until around 1950. A magnetic tape is a long thin piece of plastic, embedded with iron compounds (ferric oxide; Fe_2O_3) in powdered form. Recordings are made on the tape by passing it over a magnetic **write (or recording) head** that is receiving fluctuating electrical signals from a microphone. The write head is basically an electromagnet that magnetizes the iron compound on the tape in a pattern identical to fluctuating current in the head. The recorded signal is analog; the magnetic field of the iron in the tape varies with amplitude and frequency just like the sound did. In the diagram below a schematic of a write head is shown with a top

view of the tape. The tape is moving from right to left, unwinding from a reel on the right (not shown) and rewinding on a reel on the left (not shown). The left and right stereo channels are recorded side by side.

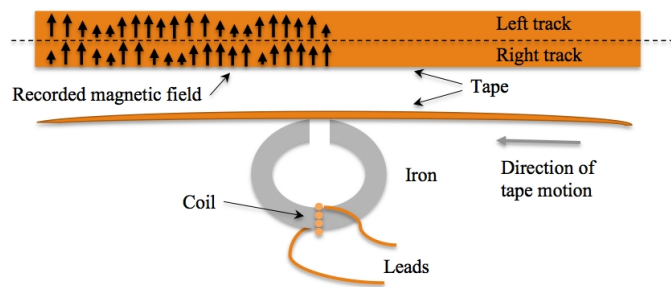


Figure 3.17.1.2.2

The **read head** of a tape player does the reverse of the write head. As the tape with a magnetic signal passes over an iron loop it induces an oscillating magnetic field in the iron. The changing magnetic field in the iron causes current to flow in a coil wrapped around the iron (Faraday's law again). The schematic for this process would look identical to the diagram above (and in fact some tape players have a single read/write head that performs both functions).

Magnetic tape became a recording industry standard because the sound from many microphones could be recorded simultaneously on a wide piece of tape as separate tracks -as many as 16 tracks could be recorded simultaneously. The same technique made it possible to record video on one track and audio on another. Video recorders began to be available in the 1960s. Being able to move the tape at different speeds is also an advantage. Using tape moving past the write head at a faster speed allows higher frequencies to be recorded more accurately. The trade-off is that more tape has to be used for the same amount of recording.

There are several drawbacks with using magnetic tapes as a recording medium. The plastic can stretch, break or melt. The size of the compound metallic grains in the tape means that very rapid fluctuations in magnetic field cannot be recorded. As a result, magnetic tape does not record high frequencies very well except at very high tape speeds which introduce other problems. If the tape is exposed to a magnetic field, the information is changed or lost. Gradually the magnetized iron may lose its field as the magnetic fields of neighboring regions of tape interact with each other. Most magnetic tape is wound onto a reel so that one layer can affect the layers above and below, creating "ghost sounds". As the tape moves past the read head it will pick up randomly oriented magnetic fields of the iron compounds, even on regions of the tape where no sound is recorded. This is heard as **tape hiss** and is especially noticeable in quiet sections of music.

Several clever ways to suppress noise on magnetic tapes have been developed. The most common method is to amplify softer parts of the music when they are recorded and then reduce their volume as they are played back. As shown in the diagram below, the loud parts are not amplified when recorded and are played back at normal volume but sounds with lower amplitude are first amplified before recording.

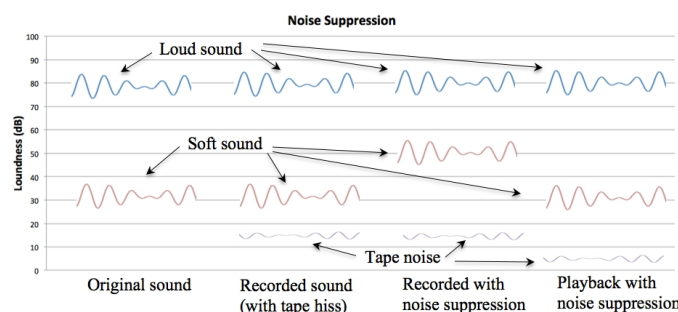


Figure 3.17.1.2.3

Video/audio examples:

- Wikipedia on [reel-to-reel tape](#).
- Wikipedia on [cassette tape](#).
- Wikipedia on [noise reduction](#).

Sound Sample of tape hiss:

Sound sample of tape distortion:

Sound sample of tape distortion:

Digital

Because vinyl records and most magnetic tapes are used to capture the actual amplitudes and vibrational frequencies of the sounds they are recording they are known as **analog recordings**. The grooves in the record or the magnetic fields of the iron compounds on the tape have variations proportional in size and frequency to the music they have recorded. An entirely different way to record sound called **digital recording** was developed beginning in the late 1950s.

Let's look at a sine wave voltage (the blue curve in the figure below). This could represent the signal coming from a microphone which is picking up the sound of a tuning fork. The signal varies from 1000 millivolts (mV) or 1 volt to -1000 mV (-1 volt). Instead of recording the actual shape of the curve, suppose we sample the amplitude (in millivolts) of the curve at many different times. So for example, we could record the voltage every 0.1 millisecond (ms). This would give us the red, stair shaped curve in the figure below. For the first 0.1 ms the recorded voltage is zero, at 0.1 ms the voltage is 250 mV, at 0.2 ms the voltage is 500 mV, at 0.3 ms the voltage is 750 mV, and so on. This list of numbers (0, 250, 500, 750 etc.) with the times they were taken (0.1 ms, 0.2 ms, 0.3 ms etc.) would be a rough representation of the original curve in numerical form.

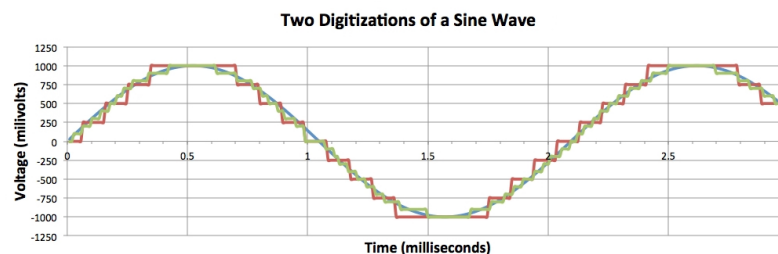


Figure 3.17.1.2.4

How can we get a set of numbers that is closer to the original sine wave? Suppose instead of sampling every 0.1 ms we sample twice as often or every 0.05 ms? This is the green curve in the figure above. So at 0 ms we still have 0 mV but at 0.05 ms we get 100 mV, at 0.1 ms we record 200 mV, at 0.15 ms we record 300 mV, etc. Now we have more numbers and the jumps are smaller (100 mV increases instead of 250 mV increases). What if we want to get even closer to the original curve? In fact we can make as accurate a representation as we want just by taking more points at a shorter **sample rate**. This is the first step in the process called **analog to digital conversion**; we convert an analog signal to a series of numbers.

There are a couple of other details to the process of recording in a digital format. Computers can only work with **binary numbers**; in other words, numbers that are either one or zero. This is because the electronic states inside a computer chip are either on or off. This isn't really a problem because there is a binary number for every ordinary number. Below is a table of binary numbers from one to 15.

Number	Binary Equivalent
0	000000
1	000001
2	000010
3	000011
4	000100
5	000101
6	000110
7	000111

Number	Binary Equivalent
8	001000
9	001001
10	001010
11	001011
12	001100
13	001101
14	001110
15	001111

Table 3.17.1.2.1

Another limitation is the number voltage steps available for dividing the amplitude of the signal. Sampling more often doesn't do any good if the voltage steps cannot be made small enough. In early analog to digital converters the voltage step size was limited by the number of ones and zeros (bits) that could fit into a memory slot and this was called the bit depth. A larger number of bits (larger bit depth) means you can divide the voltage of any given sample into smaller steps and thus have a more accurate picture of the sound wave. Most voltages are now divided into the number of steps represented by the largest number that can be stored using 16 bits (in other words the largest binary number with 16 digits which turns out to be the number 65535). The **bit rate**, usually measured in kbps (thousand bits per second) is the number of bits per sample (the bit depth) times the sample rate. For sound sampled at 44.1 kHz with a 16 bit A to D converter produces a bit rate of $44.1 \text{ kHz} \times 16 \text{ bits} = 705.6 \text{ kbps}$ (for two channel stereo the bit rate would be twice this).

Once we have a string of binary numbers recorded, how do we get the sound back? To play back the digital recorded wave we feed the list of numbers to a device that produces a voltage equal to the number it reads. The changing voltages are amplified and fed to a speaker to reproduce the sound. This is called **digital to analog conversion**. Notice that this means the reproduced sine wave will not be exactly the same as the original. Instead it will now be one of the stair step waves shown above. However, if the reproduced wave is close enough to the original our ear-brain system is fooled.

Computer disks, both the outdated floppy disk and current hard drive technology record information using the same method as magnetic tape. A plastic medium is embedded with iron compounds which can be magnetized as they pass underneath a coil. The information is read (Faraday's law again) by a coil held very near the surface of the disk (a computer crash originally meant literally that either the read head or the write head hit the surface of the disk). Instead of recording analog information (variations that are proportional to the sound variations), the data is stored as either on (a magnetic field) or off (a reversed magnetic field). In other words, the data is stored as binary information.

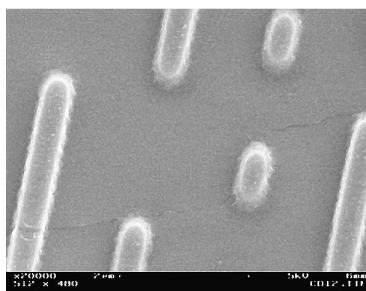


Figure 3.17.1.2.5

A **Compact Disk** or CD records the digital information as a series of divots called pits (shown above in an electron microscope picture) that are burned into the surface of a plastic disk. The flat regions between pits are called lands. A short pit might represent the binary number zero and a longer pit the number one. Three (or more) laser beams, slightly offset from each other are used to record and read the data. In the reading stage the center beam reflects off the disk into a photo detector as the disk turns below it. The reflection is detected as a beam that is alternately broken for a short period of time (a short pit) or a slightly longer period of

time (a long pit). Two beams to either side of the read beam keep the center beam aligned on the row of divots as shown in the diagram below.

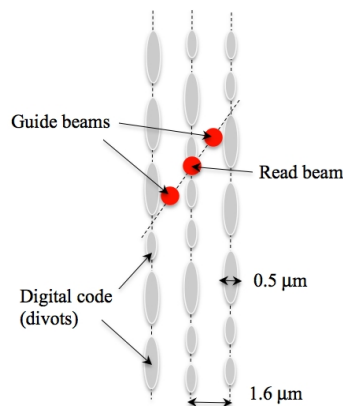


Figure 3.17.1.2.6

CD technology needed the development of the laser in order to work. The first solid state lasers were infrared, followed by red lasers. Lasers in other colors took longer to develop because of technical difficulties. Blu-ray disk technology, which uses a blue laser, wasn't available until the early 2000s after the development of blue lasers. The reason these disks hold more information is the pits are smaller and closer together. Recall from Chapter 7 that waves interact with objects close to the size of their wavelength; laser light doesn't diffract through a doorway because the opening is much larger than the wavelength. The wavelength of red light is too long to read the tiny pits in a Blu-ray disk but the pits can be read by the shorter wavelengths of blue laser light. Regular CDs use light with a wavelength of 780 nm, DVDs use wavelengths of 650 nm and Blu-ray uses 405 nm light.

One obvious problem with digital recording is the trade-off between **sample rate** and **bit rate**. Suppose the sine wave in our example above is oscillating at 60 Hz (60 oscillations per second). If the sample rate is 60 Hz (60 samples per second) each sample will catch the same point on the sine wave so the list of numbers will be constant and the signal is not recorded. In general you have to sample a sine wave at least twice a cycle in order to record the variation (in the sine wave above this would be every 0.5 ms which would record a peak followed by a trough followed by a peak, etc.). And even then the playback voltages would constitute a triangle wave of the same frequency as the original sine wave rather than a sine curve. The minimum sample rate needed to record a given frequency is called the **Nyquist rate**. The highest frequency that can be recorded is one half the Nyquist rate and is called the **Nyquist frequency**.

Humans with perfect hearing can hear up to 20,000 Hz so a sample rate of 40,000 Hz should be sufficient for most recorded music. The recording industry settled on a sample rate 44.1 kHz (44,100 Hz) as the industry standard for CD recordings. However the recording rate used in music studios is usually 48 kHz or higher. Higher sample rates are also used for non-audio signals, for example, DVD and Blu-ray audio sample rates are sometimes 96 kHz or 192 kHz.

Since most people cannot hear frequencies above 15,000 Hz very well, audio sampled at lower sampling rates often does not sound very different. Likewise dividing the sample into 65535 voltage steps isn't always necessary to capture changes in the signal, especially if the signal does not change quickly. So either the sample rate or the bit depth can be lowered without degrading the quality of sound enough to notice for most people. Most software for recording (ripping) a CD to put music onto a MP3 player (for example iTunes) lets the user choose the bit rate so that the size of the files can be adjusted to allow more music to be put onto the storage device. In most cases the sample rate stays fixed but the number of bits used to determine the voltage step size is modified. iTunes, for example, allows the user to select bit rates of 320 kbps down to 64 kbps. A reduction from 320 kbps to 64 kbps will reduce the file size of a typical song recording to one third its initial size since not as many voltage steps are being recorded per sample. For a lot of music the lower sound quality of a lower bit rate is not noticeable.

A second way to record digital music using less computer memory or CD space is by using **compression** software. Although some of software details used commercially are not made public, the general idea behind MP3 (MPEG-3) for audio, JPEG for pictures and movies as well as other compression algorithms is fairly simple. Suppose you digitize an analog signal into a stream of (binary) numbers. As you look at the stream you notice there just happens to be a sequence of ten number 2s in a row. You could simply record the ten numbers onto the CD or computer drive and be done. Or you could record 10×2 to indicate a repeat of the number 2, ten times. This latter way takes up less space because you only have to record two numbers instead of 10. When the recording is

decoded the software produces the stream of 10 number 2s when it reads the code 10×2 so that the correct voltage is played in the speaker. Other strategies for compression include eliminating sounds that are not likely to be audible to a human ear and using pattern recognition to predict the frequencies that will occur rather than accurately recording all of the patterns in the sound sample. Compression is **lossless** if all the original data is recorded. In a **lossy** compression some data that is assumed not to affect sound quality is discarded.

There is one special type of digital signal that is used internally in electronic instruments such as keyboards, drum machines, music sequencers and computers connected to these devices. **MIDI** stands for Musical Instrument Digital Interface and is an industry standard for communicating between electronic music devices. A MIDI recording is a bit like recording the sheet music to a piece of music instead of the actual sound. When a key is pressed on an electronic keyboard, information is collected about how long the key is pressed, how hard and possibly other physical information about the movement of the key. This information is digital in form (binary) and can be recorded by a computer, manipulated by a computer program, or sent to an output device that converts the digital signal into an analog signal that can be amplified and sent to a speaker or headphone. Because the output is computer controlled, the key sequence can be used to control any sound, for example flute sounds or trumpet sounds, etc. MIDI files are usually much smaller than audio files which can be an advantage. A disadvantage is the full range of analog musical frequencies cannot be recorded this way.

Video/audio examples:

- A comparison of [CD, DVD, HD DVD and Blu-ray](#).
- Sound samples of a song from a CD, recorded at different sample rates. Original file was AIFF format, 52.5 MB. All recordings were made at fixed speed. The first number is the bit rate in kilobites per second, the second number is the sample rate, the third number is the file size.
- The Wikipedia [history of the choice of 44.1 kHz](#).
- A more detailed explanation of MIDI.
- A [MIDI file](#). You can download this, open it with Audacity and compare the file with a regular mp3. It will also open with GarageBand and other software.

This page titled [3.17.1.2: Recording](#) is shared under a [CC BY-NC-SA](#) license and was authored, remixed, and/or curated by [Kyle Forinash and Wolfgang Christian](#).