Data Communications

Transmission System: Structure and Function

Nathaniel Kinsey
Hamilton College

The purpose of this discussion is to outline the major components and theories that comprise any data transmission system. Digital signaling and analog signaling are discussed as are the differences between digital and analog signals. One specific example of a data transmission system is discussed. This study is not a complete survey of data/computer networking and areas of further study are suggested. A problem set is given to illustrate the material and test the reader's understanding of the concepts of this paper.

1. Introduction

It is the purpose of this paper to provide an overview of some of the problems and techniques associated with data communications. We will cover most of the physics and theory which explain the transmission of data between points on some form of network. The general communications model consists of a *source, transmitter, transmission system, receiver,* and *destination*.

Source: A device that generates data to be transmitted; examples are telephones and computers. We will assume that we wish to transmit digital data (consisting of 0s and 1s) for the majority of this paper.

Transmitter: A device that encodes the data to be transmitted in such a way as to create electromagnetic signals that can be transmitted across some kind of transmission system.

Transmission System: The physical medium that carries the signals produced by the transmitter. This can be a single wire connecting two telephones or a vast network connecting thousands of computers.

Receiver: A device that accepts a signal from the transmission system and converts it into a usable form for the destination device.

Destination: Accepts the incoming data from the receiver. The destination device can be a telephone, a computer, a network switch, etc. The device usually processes the data in some way so that it is usable by a human user or a program running on the device.

The communications model defines the transmission system as the data carrier between source and destination. We will examine the structure and function of the transmission system in this paper. There are two physical ways to carry data: *guided transmission media* or *unguided transmission media*. Guided transmission occurs over solid wave carrying materials (wire or fiber optic cable). Unguided transmission (TV/radio, space communication) broadcasts signals into air or space and allows them to propagate to the receiver. Each of these will be discussed in greater depth later.

We first examine the terminology of electromagnetic waves and then two specific cases of data transmission. After a general discussion of the transmission system we will discuss a specific example of how transmitter and receiver work in a system used in modern computer networks.

2. Electromagnetic Waves

For now we will concern ourselves with the vocabulary and theory of electromagnetic waves. Regardless of transmission media type and data type, all data is transmitted using an electromagnetic signal. Note that an electromagnetic signal is composed of one or more electromagnetic waves. For the sake of this discussion we will examine a simple signal as opposed to a complex wave consisting of several simple waves.

2.1 Time Domain

Think of a signal as a relationship between time and signal strength. The function sin(x) is a simple alternating wave that varies constantly in "strength" over its domain. *Time Domain* analyses view electromagnetic signals as functions of time. Viewed this way we can analyze signals with respect to their strengths at given points in time. Simple functions like sin(x) are defined to be *periodic*, that is, they repeat themselves over the range. Formally, we define a function to be periodic if

$$s(t+T)=s(t) \qquad -\infty < t < +\infty \tag{1}$$

where T is the period of the signal (T is the smallest value satisfying equation 1).

2.2 Sine Waves

The function sin(x) is an example of a *sine* wave (Figure 1). Understanding sine waves will be useful later when we discuss a different view of signals. We will use simple sine waves to construct more complicated signals. Sine waves are *continuous* periodic signals. We define three features that characterize a sine wave.

Amplitude: The amplitude of a signal is the peak value or maximum strength of that signal over time. For electromagnetic signals, amplitude is usually measured in volts at a specific point in the domain (See Figure 2).

Frequency: The frequency of a signal is the rate at which a signal repeats itself. Frequency is measured in cycles per second (Hertz, denoted Hz). Note the similarity between frequency and period. The *period* (T) of a signal is the length of time for one repetition; thus $T=\frac{1}{f}$ (See Figure 2).

Phase: Phase is the measure of a horizontal shift of a wave. For example, consider the function f=sin(x) . The function can be shifted to the right by some amount. It is measured in degrees where 360 degrees is defined to be one period. Figure 3d shows a phase shift of plus 180 degrees.

Figure 3 shows the affect of changing each of the three parameters on a sine wave. Part (c) shows that two parameters can be changed at once (amplitude and frequency).

Finally we define the *wavelength* (λ) of a signal to be the distance occupied by one cycle of the signal. The wavelength is the shortest distance between two points of equivalent phase on the signal. If we are given that the signal is travelling at some velocity v , we can define

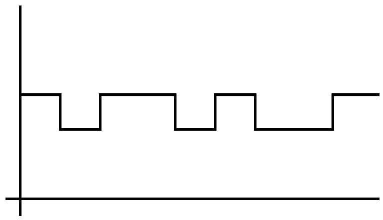
$$\lambda = vT$$

Due to the relationship between T and f we can rewrite the above equation:

$$\lambda = \frac{v}{f}$$

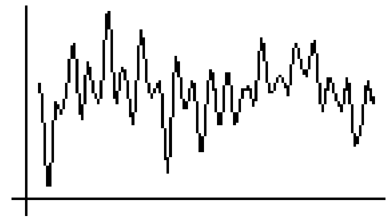
2.3 Continuous/Discrete Signals

Figure 4 (below) shows two special types of signals: a *square wave* (discrete) and a continuous analog signal. The continuous signal could represent a speech recording and the discrete signal could be used to represent binary values.



Discrete

Figure 4



Continuous Analog

3. Frequency Domain

We now shift our discussion from a time domain discussion to a more useful analysis: Frequency Domain. Analyzing signals with respect to the frequencies they contain is valuable because it provides a background for understanding how we measure the capacities of a transmission medium.

3.1 Complex Signal Construction

The individual sine waves that we discussed in the previous section are not used in real world applications because a periodic signal is not useful in real-world data carrying applications because it can not carry any data. In order to make a signal carry data we need to be able to alter it in some observable way. This section discusses the theory of constructing a non-periodic wave from periodic component waves. For example, Figure 5 shows the graphs of two separate frequencies and the effect of adding the two frequencies together. In this case the second frequency (b) is a multiple of the first. The first frequency in this case is called the *fundamental* frequency. Note that the period of the signal in (c) is the same as the period of the fundamental frequency.

Fourier analysis can be used to show that any signal can be constructed using sinusoidal component frequencies. Thus we can construct any given signal from an array of sine waves. If we wish to construct a periodic square wave we would proceed as shown in Figure 6.

The period is defined by the first wave (the *fundamental frequency*) and is refined by the addition of higher and higher frequencies. With each higher frequency we note that the wave moves closer and closer to a representation of the signal that we wish to build. Note that a perfect square wave is not possible because Fourier analysis tells us we need an infinite number of component frequencies. Each step makes the wave a little "squarer" but we cannot create a perfect square wave.

3.2 Bandwidth

The *spectrum* of a signal is the range of frequencies that it contains. Many signals will contain an arbitrary range of frequencies due to the process described in the previous section. If a signal has many sharp peaks and hard edges, we will need many high frequencies to approximate those peaks and edges. Remember that we can create an arbitrarily large sequence of signal approximations, each better than the previous one. In creating that sequence we use a corresponding sequence of higher and higher frequencies. Most of the signal strength is contained in a range of frequencies in the middle of the signal. High frequencies used to construct signals cannot be of higher amplitude than the fundamental frequency. The sum of a set of increasing amplitude signals does not yield a square wave. So, as the frequency goes up, the amplitude goes down and thus each higher frequency is weaker than the previous frequency. Receivers are limited in the range of signal strengths they can detect so most of the detectable energy in the signal is in a narrow band of frequencies. Though theoretically a signal uses a near infinite range of frequencies, equipment can detect only a small portion of it. The bandwidth of any signal used in communications is very wide in theory, but since the transmission medium limits us to a narrower band we define this usable area of the signal to be the *bandwidth* of the signal. Similarly, any transmission medium has a range of frequencies in which it is effective. Copper wire can carry waves varying in frequency from 0 Hz to about 3 MHz. The upper end is not a hard edge but the majority of the power is contained in the given range. Of course, copper wire cannot transmit light waves, so there is an upper bound on the range of frequencies that a medium can transmit. The range of frequencies a medium can transmit effectively is defined as the bandwidth of that medium.

4. Frequency Addition

Just as we can add two simple sine waves together, we can add two (or more) *complex* analog signals together. A complex signal is one that is constructed from several constituent frequencies and may or may not carry data. The result is an even more complex signal that does not appear to preserve any of its constituent frequencies. Look back at Figure 6 and note that it is not immediately obvious that (d) contains a high frequency, let alone what they are. When we add two complex signals together the result is similar to Figure 6 but much more complicated. If we add two complex signals that each use the entire 10-100Hz range of frequencies, the result will be unusable (provided our task is to separate the signals) because we have no way of knowing which signal each component belongs to. If we were to use entirely separate bands of frequencies to construct the two signals, we could add them and have a usable product. The reason that one sum is worthless while the other is usable is explained by how receivers work. A receiver is tuned to listen for a specific band of frequencies. Thus if a receiver is tuned to a range from 50Hz to 100Hz it will ignore frequencies in the 0-49 Hz range. If two signals are summed (one in 50-100Hz, the other in 0-49Hz) and transmitted we can extract both signals by using two differently tuned receivers.

4.1 Frequency Division Multiplexing

The important part of Frequency Addition is that more than one signal can be transmitted at once on a single transmission medium as in radio or TV. We define *Frequency Division Multiplexing* (FDM) as the process where signals are distributed into separate ranges of frequencies in the carrier bandwidth. For example, a voice grade line can carry frequencies in the 300 Hz to 3400 Hz range. Suppose we split this bandwidth at 1700 Hz and center the two carrier frequencies one at 1170 Hz and one at 2125 Hz. Then each carrier frequency can be *modulated* by 100 Hz on either side of the carrier frequency with no overlap in the middle. Modulation is the process of varying the characteristics of a signal in such a way as to represent some set of data. FDM allows many signals to be transmitted at once, greatly improving channel utilization. We can also transmit signals in two directions at once in a guided medium (in separate ranges of the bandwidth). Computer networks rarely use FDM (except for ISDN) but it is valuable for phone companies and unguided computer communications.

5. Data Transmission

Up to now we have limited our discussion to the theory of analog signaling. We will now focus our discussion on communication between two connected points on a network. The process of sending data between points A and B is composed of two steps. The first step is to create a signal that is suitable for the medium connecting A and B. The second step is to transmit the signal to the receiving station. We first deal with signal generation. We assume that transmission media can transmit either digital or analog signals. In particular, if we wish to transmit analog data, but A and B are connected by an optical fiber line, we must first create a digital signal from the data because optical fiber only transmits digital data. Once a signal is created, it is transmitted through the medium connecting A and B. No medium is capable of perfect transmission over an arbitrary distance so the signals may need to be regenerated in some manner.

If the medium carries an analog signal, amplifiers are used to maintain signal strength. An amplifier increases the amplitude of the signal by adding a constant value to it everywhere in the domain. Amplification simply increases the strength of the signal that arrives at an amplifier so any line noise is incorporated into the new signal. Analog signals can withstand more line noise than digital signals before the signal is not useable. Amplified line noise eventually appears as bits in a digital signal (errors) so we must use another method of retransmission. We could recover the digital signal and then retransmit the bit string in a brand new signal. A repeater does just that. The retransmission creates a brand new, perfect digital signal. Either of these techniques allow transmission of signals over vast distances with little concern for data loss.

5.1 Digital Data, Digital Signals

This section will first deal with the aspects of a digital signal and then with transmission of that signal. Previously we discussed bandwidth in terms of frequencies and the possibility of sending more than one signal at one time. The model we use to explain digital signaling is a light switch. The switch is used to create a voltage in the wire to which it is connected. If we turn it on and off quickly, we create pulses of current. Suppose for a simple case that we wish to send a stream of 1s and 0s. We could let a pulse of current indicate a 1, and no current indicate a 0. We refer to each pulse as a signal element. *Data rate* is defined to be the rate at which a mechanism can transmit data (in bits/sec). For some data rate R, the duration (in time) of a bit is $1/R$.

5.1.1 Bandwidth/Signal Rate

What is the relationship between bandwidth and digital signaling rate (data rate)? If we are given some channel with a given bandwidth and wish to send digital signals on it, we signal at close to the highest frequency of the band. Remember that Fourier analysis shows that a square wave can be constructed from a set of sine waves. The construction involves selection of a fundamental frequency that determines the period of the resulting wave. If we wish to construct a simple wave and have a frequency range from 0 Hz to 3 MHz, the highest frequency we can use for the fundamental frequency is the one that alternates at 3 MHz. Note that the upper limit on frequency is not a hard edge; the medium is capable of supporting frequencies higher than the given frequency, but when using high frequencies the transmission is much more susceptible to error because noise and interference become important factors, as does receiver sensitivity. Errors are directly related to line length (and environment) so in some cases it is possible to increase the data rate given shorter distances and more sensitive receivers. Most ethernets signal at 10 Mbps using a wire similar to a phone wire, but the length of a segment is limited to short runs. As a general rule of thumb we will use the maximum bandwidth as the data rate. Thus, since a twisted pair connection can support frequencies from 0 to about 3 MHz, the data rate will be about 3 Mbps.

5.1.2 Baseband

We define the above method of transmitting data as *baseband*. Digital signals are dropped on the wire as voltage pulses and consume the entire bandwidth of the medium. Remember that if we want to create a close approximation of a square wave we need to use lots of high frequencies to approximate the square edges in the wave. A digital signal's voltage pulses are essentially a square wave and Fourier theory tells us that we have to use the whole spectrum of the medium to create the square wave. Since the entire spectrum of the transmission medium is used, FDM is not possible. In order to send data from several sources at once we employ *Time Division Multiplexing*, as discussed in the next section.

5.1.3 Time Division Multiplexing

Unlike analog signaling and transmission, digital signals cannot be combined and transmitted at once. If several data sources are present and all wish to send at once, digital systems must string the data together into one continuous stream. This method of communication is called *Time-Division Multiplexing* (TDM). As the name suggests, the transmission system divides up time into slots and reserves one of the slots for each source. Thus a source can only have its data transmitted during its time slots. A prioritizing system can be implemented so heavily loaded sources have more time slots dedicated to them than to less important sources. Note that if a source has no data to transmit during its time slot that time slot is wasted since the system does not check to see which sources actually have data to send. Due to the nature of digital signaling, multiplexing of signals must occur in the preparation of data before it is sent. Reconstruction of the separate data sets requires advance communication. At the receiving end there are a number of receivers, one for each source. Once the data stream is broken apart into the separate data streams each receiver accepts the data at the same rate it was transmitted.

5.1.4 Full/Half Duplex

Transmitting data in both directions at once is called *Full-Duplex* transmission. Full-duplex transmission can be achieved either by using two separate transmission channels or, in some cases, a single channel. *Half-duplex* systems allow data to flow in both directions but only one station can transmit data at one time. Half-duplex systems can use one or two channels, depending on transmitter/receiver arrangements.

Digital signals cannot be sent in a full-duplex system with a single channel. If two signals are sent at once, the result will be unreadable because the voltages will add or cancel, resulting in a garbled string of bits. Full-duplex transmission of digital signals can be accomplished with two channels, one for each direction. Each station has its receiving apparatus listening on one channel, and its sending capability on the other. One channel can support half-duplex with the understanding that only one station can send at one time. In order to support half-duplex each station must possess both receiving and transmitting equipment. Thus each station can transmit and receive. Note that collisions and corrupted data occur if both stations attempt to transmit at once. We leave the theory of detecting and preventing simultaneous transmission in half-duplex to discussions of data flow control.

5.1.5 Signal Regeneration

As mentioned before, no transmission medium is capable of carrying a signal for arbitrary distances. Electrical resistance and outside interference degrade signals and if the transmission distance is great enough the signal will be reduced to a worthless stream of errors. In order to protect the data stream, some method of regenerating the signal is needed. A device called a repeater is used to maintain the signal. A repeater receives the incoming signals and recovers the digital data and then retransmits a brand new clean signal. Using this technique, noise and interference caused by the medium are eliminated at each repeater. Thus, using a series of repeaters, a signal can be transmitted an arbitrary distance with little concern for data loss.

5.2 Digital Data, Analog Signals

In the previous section, we discussed the concepts of digital signaling and digital signal transmission. This section will discuss the problem of transmitting digital data using analog signals. Computer users are most familiar with this problem when communicating over the public telephone system. The telephone system is designed to switch and transmit analog signals in a range of frequencies that are suitable for voice communication. Digital switching and transmitting hardware is not currently installed, so end users must use a device to translate digital data into some analog signal. Such a device is called a *modem* (from Modulator-Demodulator).

5.2.1 Encoding

Since an analog signal is a continuous wave, we must change that signal in some way to encode digital data. When we explored sine waves, we discussed three characteristics of a signal: Amplitude, Frequency, and Phase. We can modulate (change) a signal with respect to one of those characteristics to encode digital data. We will discuss modulating the amplitude and frequency of a signal. Each of these cases follows the general rule for analog signals: occupation of a given bandwidth centered at a given frequency (carrier frequency).

Amplitude Shift (AS)

In AS, the binary values are represented by two different amplitudes of the carrier frequency. A common method of AS looks a lot like a digital signal because it uses pulses of the carrier frequency to denote 1, and absence of a signal for 0. As one might imagine, this system is very prone to error. Any outside interference is certain to affect the amplitude of the signal so that almost every data point can be in error in a poor transmission environment.

AS is used very successfully to transmit data over optical fiber. A pulse of light is really just an analog signal for a short time. Since light waves have such a high frequency, the pulses can be very short (remember: $T = 1/f$). Thus for very large f, the period, T, is very small. Any pulse of a signal in AS must be at least T long otherwise it doesn't appear as a signal at all! Optical fiber signals are transmitted in pulses, a pulse corresponding to a 1, no signal or a very low background signal corresponding to 0.

Frequency Shift (FS)

In FS the binary values are represented by two different frequencies near the carrier frequency. The two frequencies are usually on opposite sides of the carrier frequency. FS has the advantage that it works under full duplex operation. By splitting the available bandwidth in half and centering the two carrier frequencies in the middle of each half, two channels are created. Then each station uses one of the channels and listens on the other. Bandwidth is reduced by a factor of the number of stations but for some applications the reduced bandwidth is acceptable. FS is less susceptible to error than AS and it can be applied to situations with much higher frequencies than voice grade lines. The carrier frequency can be centered in the 3 MHz to 30 MHz range for radio transmission and can be used for coaxial local area networks at even higher frequencies.

5.2.2 Full/Half Duplex

We showed in the previous section that full-duplex transmission is possible in a single channel. Because analog signal transmission is not baseband, we can use the techniques of FDM to send data in both directions at once. Half-duplex transmission is of course possible provided an arrangement exists to prevent simultaneous transmission at the same frequency.

5.2.3 Signal Regeneration

Just as with digital signals and digital transmission, analog signals degrade over distance for the same reasons. Outside interference and the quality of the transmission medium cause signals to weaken and error to be introduced. Recall from before that most analog signals contain a very wide range of frequencies. Complicated analog signals (voice communications) contain a very large set of frequencies. A repeater will not recreate an analog signal because it would have to detect every constituent frequency and retransmit each. The spectrum of voice communications is infinite so no repeater will ever succeed at detecting and transmitting (separately) every frequency in the signal. Without access to the source of the signal, we cannot reconstruct an exact replica of the original signal due to the accumulated error and there is no way to determine the difference between error and data without access to the source. Instead we use an *amplifier*, a device that increases the strength (amplitude) of a signal by adding a constant value to it everywhere in the domain. Note that amplifiers not only strengthen the signal, they also retransmit whatever noise picked up along the way. We hope that most errors will cause small changes in the overall signal and thus will not harm much of the data it contains. Signal degradation is acceptable to a relatively high degree when transmitting voice communications. The human receiver is quite good at understanding the data stream even if there is a lot of line noise and corruption.

6. Transmission Media

Having discussed methods for encoding and transmitting digital data we must examine several of the transmission media available to carry our signals. We will examine three transmission media that are very common in guided computer and telephone networks. Twisted pair and Coaxial cable are two kinds of electrical cable that can transmit analog or digital data. Optical Fiber is an optical medium that transmits digital signals at extremely high rates. In addition to guided transmission media we will discuss broadcast radio, microwave and infrared communications.

6.1 Guided Transmission Media

6.1.1 Twisted Pair

A twisted pair is two insulated copper wires twisted together. One twisted pair acts as a single communication path. Sets of twisted pairs are bundled together in a larger line for runs of longer distances. The large bundles are shielded to reduce interference from outside electrical sources and from neighboring bundles of twisted pair.

The most common application of twisted pairs is the telephone service to homes and businesses. Each house is connected to a phone company office by a twisted pair as discussed earlier. The system was originally designed for analog voice communications but there are ways to create an analog signal that carries digital data. Twisted pairs can and are used to send digital information in the form of digital signals. However, telephone companies require analog signals since phone company switching offices are equipped to handle analog signals, not digital signals (the switching and retransmission requirements are different).

The capacity of twisted pair is fairly low. The medium was designed for only a few voice channels and its 250 kHz bandwidth is sufficient for that purpose. Because the bandwidth is fairly low, digital data rates of only a few Mbps are possible for long distances. However, 100 Mbps is possible with high quality transmitters and receivers and short distances.

Twisted pair cable comes in several flavors. The simplest is unshielded twisted pair (UTP): regular telephone wire. Most office buildings are pre-wired with lots of UTP. Since it is readily available, cheap, and easy to work with, UTP is frequently used for local area networks. UTP can be improved by adding braided wire shielding. Shielded twisted pair (STP) is more expensive and harder to work with. A new standard was created in 1995 to specify several types of both UTP and STP. There are three categories of UTP and we will discuss the two most important and widely used.

Category 3

Category 3 wire and hardware are rated up to about 16 MHz. This corresponds to a newer voice grade cable and connection system. A data network using Category 3 wire is capable of data rates of close to 16 Mbps. Modern office buildings are wired with Category 3 cable and hardware is readily available for local area networks.

Category 5

Category 5 cable is designed for data transmission and can support much higher frequencies (up to 100 MHz). The physical design is different; the wire is twisted much more tightly. Category 5 cable is more expensive to produce, but the increased data rates it can support are worth the expense for high traffic/high speed networks.

6.1.2 Coaxial Cable

Coaxial cable can be used to transmit analog or digital signals. Coaxial cable consists of 2 conductors like twisted pair, but is constructed with a hollow outer cylindrical conductor separated from a solid inner conductor by insulating material. Think of a pipe with a wire suspended in the middle of it. Due to its construction, coaxial cable is less susceptible to interference than twisted pair. Coaxial cable is capable of supporting a much higher frequency range than twisted pair. Its design limits interference so that repeater/amplifier spacing can be much greater. Coaxial cable supports an analog bandwidth of about 400 MHz. For example, a single voice channel requires about 4 kHz and if we divide the full bandwidth of 400 MHz into 4 kHz blocks we end up with 100,000 voice channels! The volume is not quite that high due to multiplexing techniques and frequency band spacing but this illustrates the capacity of coaxial cable. Such a high capacity makes it an important part of telephone company's long distance networks. Another commonly seen application of coaxial cable is for cable television. Its high bandwidth can carry dozens or close to hundreds of TV channels for distances up 10 miles. The higher bandwidth also allows for data rates to approach 500 Mbps (remember the rule of thumb concerning the relationship between maximum frequency and data rate). Coaxial cable is quite useful in a local area network for applications that experience high load or a high number of network devices or both.

6.1.3 Optical Fiber

Optical fiber is the highest capacity guided medium available for any application. The maximum data rate for coaxial cable is in the hundreds of Mbps over a few kilometers. Optical fiber operates with near infrared light sources and so has a much higher bandwidth. In fact, the maximum data rate is only limited by our ability to modulate the light source and detect the modulation. The frequency of light is approximately 10^{15} Hz. If we take a bandwidth of 1% of that total frequency we have 10^{13} Hz but one GHz is just 10^{9} Hz which is just 1/10000 of the available bandwidth. So the upper limit of 2-3 Ghz is imposed by our ability to modulate the light and detect the changes. The practical bandwidth translates to 2-3 Gbps over tens of kilometers. Since the transmission properties of fiber are very good, repeaters can be spread out dramatically (compared to coaxial or twisted pair). Optical fiber is used exclusively for extremely high load applications. It is also present in local area networks serving as a main backbone between switches or areas of a network. Cross country telephone trunk lines use optical fiber as do large data communication routes.

An optical fiber is constructed from a very thin stand of glass or ultra pure plastic that is incased in a cladding of glass or plastic with optical properties different from those of the inner core. An outer jacket encases the cladding and core and guards them against moisture and damage.

Data is transmitted by sending pulses of light in the infrared or visible spectrums down the core of the fiber. As light enters the core it hits the edges of the core at a variety of angles. The core material is designed to reflect light that hits at a relatively low angle. Light that hits the edge at a steep angle goes though and is absorbed by the jacket. We refer to optical signaling as *multi-mode* if more than one angle can reflect. If the core radius is reduced, fewer angles will be reflected. Reduction to the width of about a wavelength allows rays at only a single angle to reflect. If only one angle can reflect we refer to the signaling as *single-mode.* Single-mode transmission is preferable because multi-mode transmission creates several propagation paths so signal elements spread out in time. This spreading out limits the rate at which the data can be received. Single-mode transmission has only a single propagation path so the data elements do not spread out.

Two different light sources are used for optical signaling. The Light Emitting Diode (LED) and the injection laser diode (ILD) are both semiconductor devices that emit light when supplied with a voltage. The LED is cheaper and operates in less forgiving environments and lasts longer. In contrast, the ILD is more expensive and but can support higher data rates. The receiving end is simply a device that resembles a photo-cell. A photo-cell produces a voltage that varies with the intensity of the light that hits its receiver.

6.2 Unguided Transmission Methods

Unguided media accomplish transmission and reception by means of an antenna. To transmit, an antenna is powered so that it radiates electromagnetic energy into the medium (usually air) where it propagates out away from the antenna. For reception, a similar antenna receives electromagnetic waves from the medium. Two types of transmission are used: omnidirectional and directional. Omnidirectional transmission allows the signal to propagate in all directions from the transmitter (radio/TV). Directional transmission uses higher frequencies focused into a narrow beam so that in this case transmitter and receiver must be carefully aligned.

6.2.1 Terrestrial Microwave

Microwave frequencies are in the 2 Ghz to 40 Ghz range. Because the frequencies are so high, the antennae used to create the signals are small (about 10 ft. in diameter). If a dish type antenna is used the energy of the signal is focused in a beam, enabling point to point transmission. The antennae are mounted high above the ground (towers, tops of hills, high-rise buildings) to achieve a clear line of sight to the receiver and extend the range. Due to the relatively high bandwidth of microwave it is a satisfactory alternative to coaxial cable or optical fiber for long-haul telecommunications links. Microwave also requires far fewer amplifiers or repeaters than guided media do so it is valuable as long as line of sight transmission is possible. Microwave can transmit data through areas that are not inhabited or not amenable to installing cable.

6.2.2 Broadcast Radio

In contrast to microwave, radio transmissions are broadcast. Thus, radio does not use dish-shaped antennas and their location is not as critical. Similarly, the receiver does not require line of sight to the transmitter or any specially shaped antenna. The antenna used for receiving can be found on most cars; it resembles a simple length of wire.

Radio transmissions cover the range from 3 kHz to 300 GHz. The radio we are used to is in the 30 MHz to 1 GHz range. We define the term *broadcast radio* to refer to the smaller range. Broadcast radio covers the FM radio dial as well as UHF and VHF television. Broadcast radio transmission is not used for data communications.

6.2.3 Infrared

Communication using infrared light is accomplished by using transmitters and receivers that modulate infrared light. The transmitters and receivers are paired together and referred to as a transceiver. Infrared transmission is only used in a direct line of sight situation for short distances. Security and interference are not big issues because infrared radiation does not penetrate walls. Further, no licensing is required as each infrared unit does not have to use its own frequency.

7. Ethernet: CSMA/CD

Understanding the fundamentals of data communication is valuable, but we have yet to examine a concrete real world example. We know how two stations can communicate and that in some situations that communication can break down in the form of signal collisions. We will now discuss a specific medium access control technique. A medium access control technique is a set of rules guiding how multiple stations may access the communication medium they use and how they deal with errors. One such technique is Carrier Sense Multiple Access with Collision Detection (CSMA/CD).

7.1 CSMA

With CSMA, a station that wishes to send data first listens to the medium to see if any other station is transmitting. If the medium is in use (a signal is sensed) then the station must wait until no signal is present. When the line is idle, any station with data to send may do so. However, if two stations attempt to send data at once (they both sense the line idle at the same time) a collision will occur. Since we are using baseband signals both data streams will be garbled and unreadable. In CSMA, a station just listens a set length of time (which takes into account propagation delay, i.e. the time it takes a signal to travel between two points, and that the receiver must contend for the medium) for a return acknowledgement from the receiver. If the sending station does not receive an acknowledgement it assumes a collision occurred and attempts to retransmit.

Propagation delay changes the effectiveness of this scheme. If the propagation delay is significantly shorter than the time it takes to transmit one chunk of data, then CSMA is effective. Collisions could only occur if two stations attempt to transmit within the length of the propagation delay. After that period of time the whole length of the medium is filled with the data stream and no other stations can transmit.

We need an algorithm to decide what to do when a station senses the medium and finds it in use. The technique used in the Ethernet standard is called the *1-persistent technique*. A station that wishes to transmit listens to the medium as above and obeys the following rules:

1. If the medium is idle, transmit; otherwise, go to step 2.
2. If the medium is busy, keep listening to it until the channel is sensed empty and then transmit immediately.

If two stations wish to send while another is transmitting, of course a collision is guaranteed and the mess will get sorted out after the collision.

When two data sets collide, the medium is unusable for the duration of transmission of both data sets. If the data sets are especially long, this delay can be substantial. We could eliminate this problem by having the stations listen to the medium as they transmit. If they detect garbled data while transmitting they should immediately stop transmission, as they are causing a collision.

7.2 CSMA/CD

CSMA/CD adds collision detection by refining the rule set as follows:

1. If the medium is idle, transmit; otherwise, go to step 2
2. If the medium is busy, keep listening to it until the channel is sensed empty and then transmit immediately.
3. If a collision is detected while transmitting, transmit a signal to notify other stations of the collision and then cease transmission.
4. After transmitting the collision signal, wait a random length of time and attempt to retransmit (step 1).

With collision detection, efficiency is significantly increased. In fact, the amount of wasted time is reduced to the length of time it takes to detect a collision. How long is that? Consider the worst possible case:

> Take two stations that are as far apart as possible and let the first start transmitting. Let the second station start to transmit just before the first stations signal reaches it. The second station will cause an immediate collision and will detect it stop transmitting very quickly. However, the garbled content must travel back to the first station before it realizes what has happened.

Thus we can say that the time to detect a collision is less than or equal to twice the maximum propagation delay. One important rule to maintain efficiency in this system is that data sets must be long enough to ensure collision detection before the end of transmission. If the data sets are too short, CSMA/CD will have the same efficiency of CSMA.

8. Conclusion

The discussion of data communications does not stop here. We have shown how and why the transmission system functions. The study of computer networking includes many more issues and problems. Flow control and routing algorithms provide logic for sending data through a large network in some efficient manner. Error correction/detection allow for flawed data to be detected and corrected either by retransmission or through other means. Network security studies encryption, authentication, and, increasingly, ethics. And one can study and create network applications: computer address databases, Web browsers, and designing systems for cataloging and retrieving information. But don't forget: All of the advanced topics mentioned here use a transmission system!

9. Problem Set

1. Why does a full duplex analog signal transmission have to be broken up into separate frequencies for each station? What happens if the channel is not divided?

2. Why is frequency shift encoding less prone to error than amplitude shift encoding?

3. Some personal tape/radio players use the headphone wire as their antenna for receiving FM radio broadcasts. How is it possible for the wire to be used as an antenna when it is already carrying an amplified signal to the speakers?

4. With CSMA, if two stations wish to transmit while another is transmitting, a collision is guaranteed. After the guaranteed collision why doesn't the system get caught in an endless cycle of collisions caused by those two stations?

5. How fast must a receiver be able to detect changes in a signal in order to receive an alternating (10101…) bit stream sent using a 10 MHz periodic signal? Generalize this result for any signaling rate. We call this result the *sampling rate.*

6. Why are amplifiers not used for retransmission of digital signals?

7. When can we add two complex signals together and be able to retrieve each signal after transmission?

8. Illustrate, using a graphing calculator or computer algebra system, what happens if a series of sine waves of increasing amplitude and increasing frequency are summed.

9. How many simultaneous voice signals can be transmitted over a 15 MHz channel, assuming a 10 Hz space is left between each voice channel?

10. What is the maximum number of data streams that can be sent using TDM on a 100 MHz channel if we do not want each individual stream interrupted by more than .1 seconds at the receiving end? What is the bandwidth for one station? What are the advantages/disadvantages to this particular system?

11. How long will it take to transmit 3 separate 10000 bit data streams using TDM on a 1 MHz channel if we alternate bit by bit between the streams? If we send 100 bit sections?

12. Describe a situation where full-duplex analog transmission is not possible.

13. What is the highest (theoretical) bandwidth possible using optical fiber? How fast would a receiver have to detect signal modulation to receive an entire signal?

14. Radio antennas often work best when the diameter of the antenna is equal to the wavelength of the radio wave. Feasible antennas range in size from 1 cm to 5 meters in diameter. What range of frequencies does this cover? Assume that radio waves propagate at the speed of light, $c = 3\times10^8$ meters/sec, and assume the antenna is a parabolic dish.

15. A station wagon can easily carry 400 magnetic tapes, each capable of storing 8 GB of data. Suppose the station wagon can travel at 50 km/hour for 100 km. What is the data transmission rate (in bps) of this system? What is the moral of this story? What is the obvious downside to transmitting data in this fashion?

16. Why are broadcast radio frequencies strictly licensed?

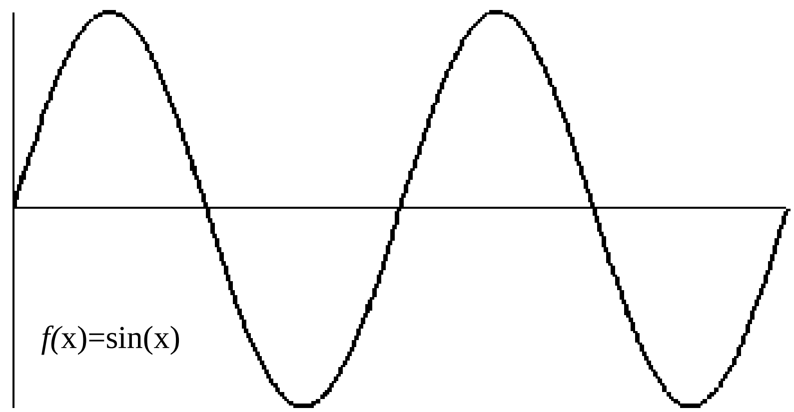


Figure 1

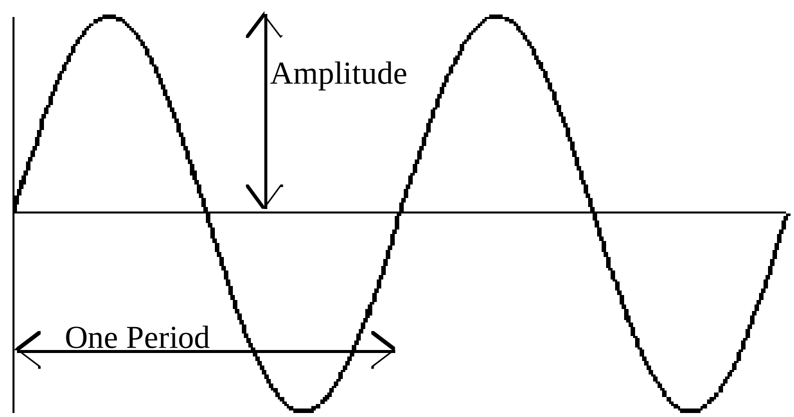
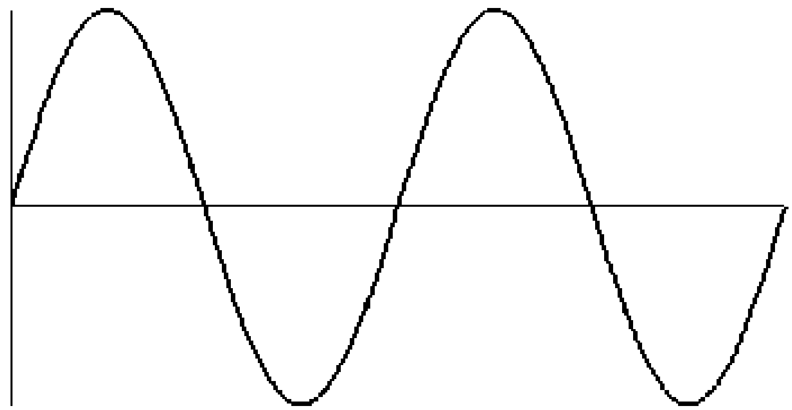
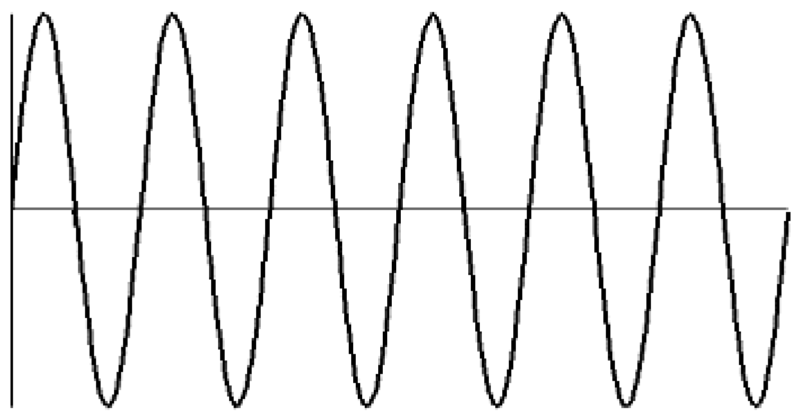


Figure 2

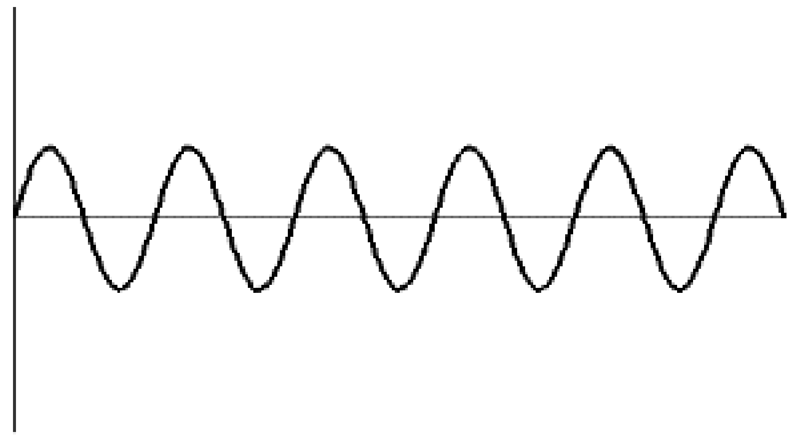
Figure3



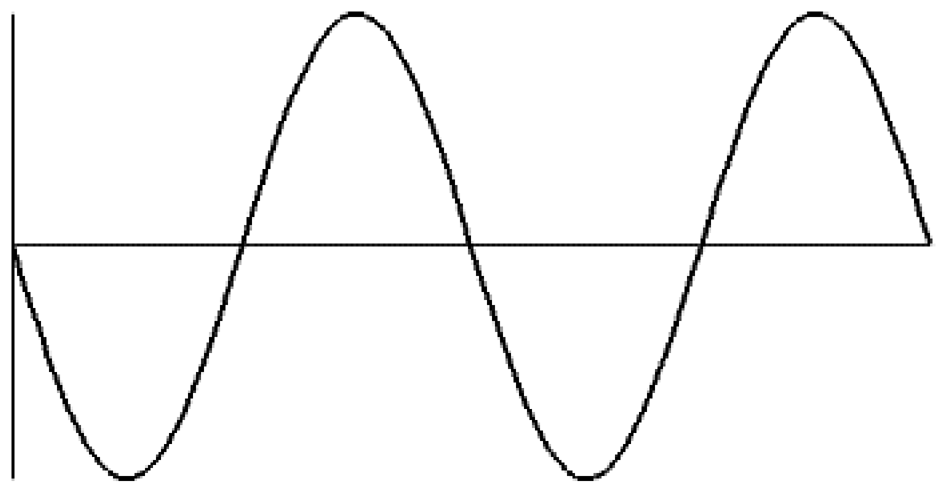
(a)
Fundamental Wave



(b)
Increased Frequency

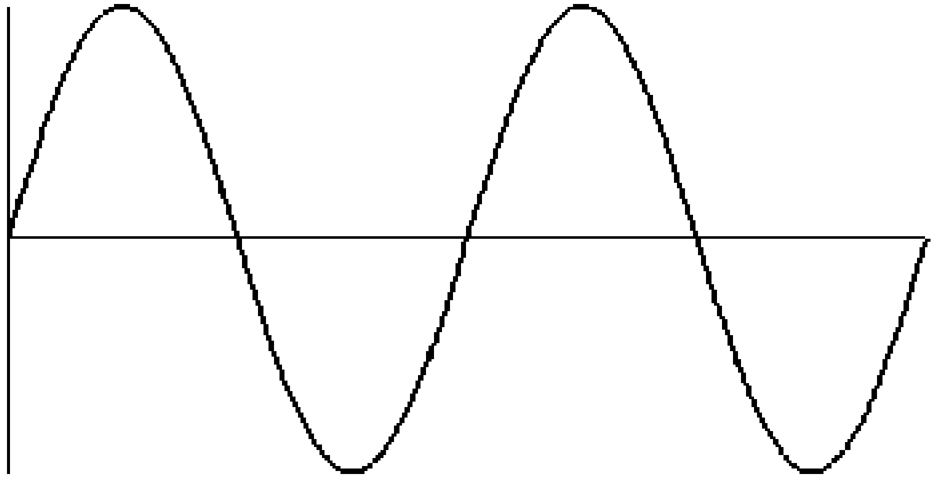


(c)
Low Amplitude, Increased Frequency

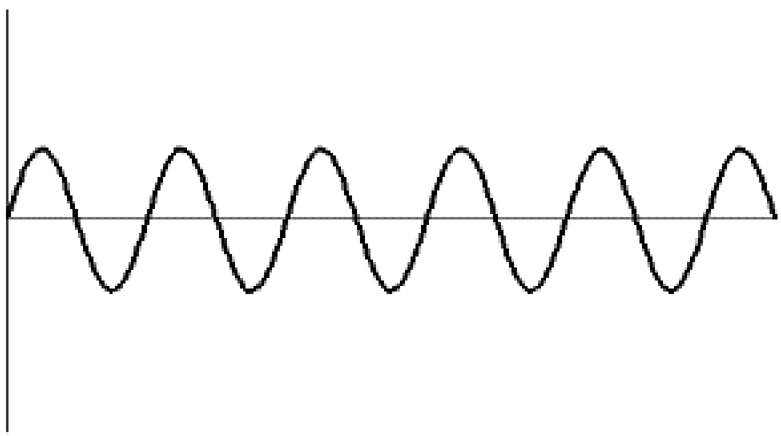


(d)
Phase shift: +180º

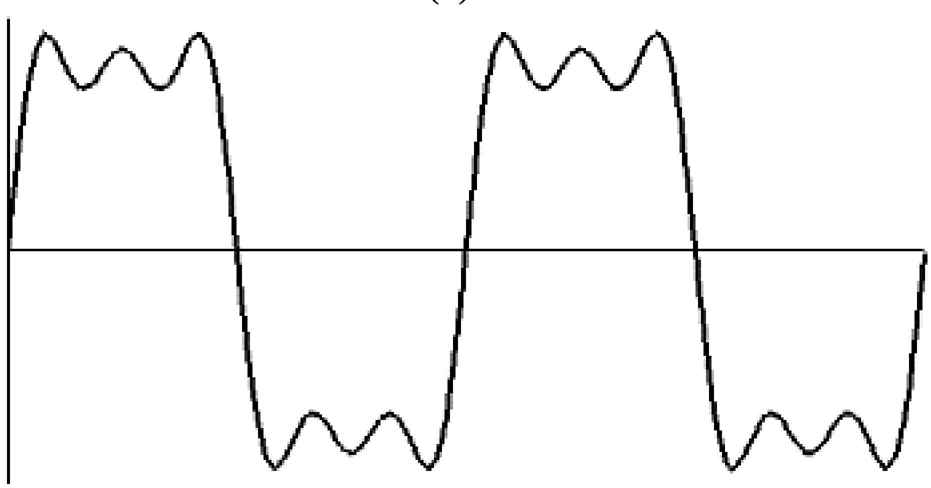
Figure 5



(a)



(b)



(c)

Figure 6