Chapter 2

Wireless Coding and Modulation

Coding and modulation provide a means to map digital information to the underlying signal so that a receiver can retrieve the information from the signal using appropriate decoder and demodulator. As coding and modulation directly affect the achievable capacity and data rate of the communication system, new coding and modulation techniques are constantly proposed and implemented to keep up with the demand for mobile data. This chapter will cover the basic theories and terminologies of coding and modulation in digital wireless communications.

2.1 Frequency, Wavelength, Amplitude, and Phase

Signal waveforms are the fundamental carriers of all types of data that we send over a communication system. It is therefore important to understand the basic properties of such waveforms. Figure 2.1 shows the waves that are created when a rock is thrown into the water. Similar waves are also created in wireless networking for communication purposes, but these are called *electromagnetic waves*, and special electronics circuits are used to generate and receive them.

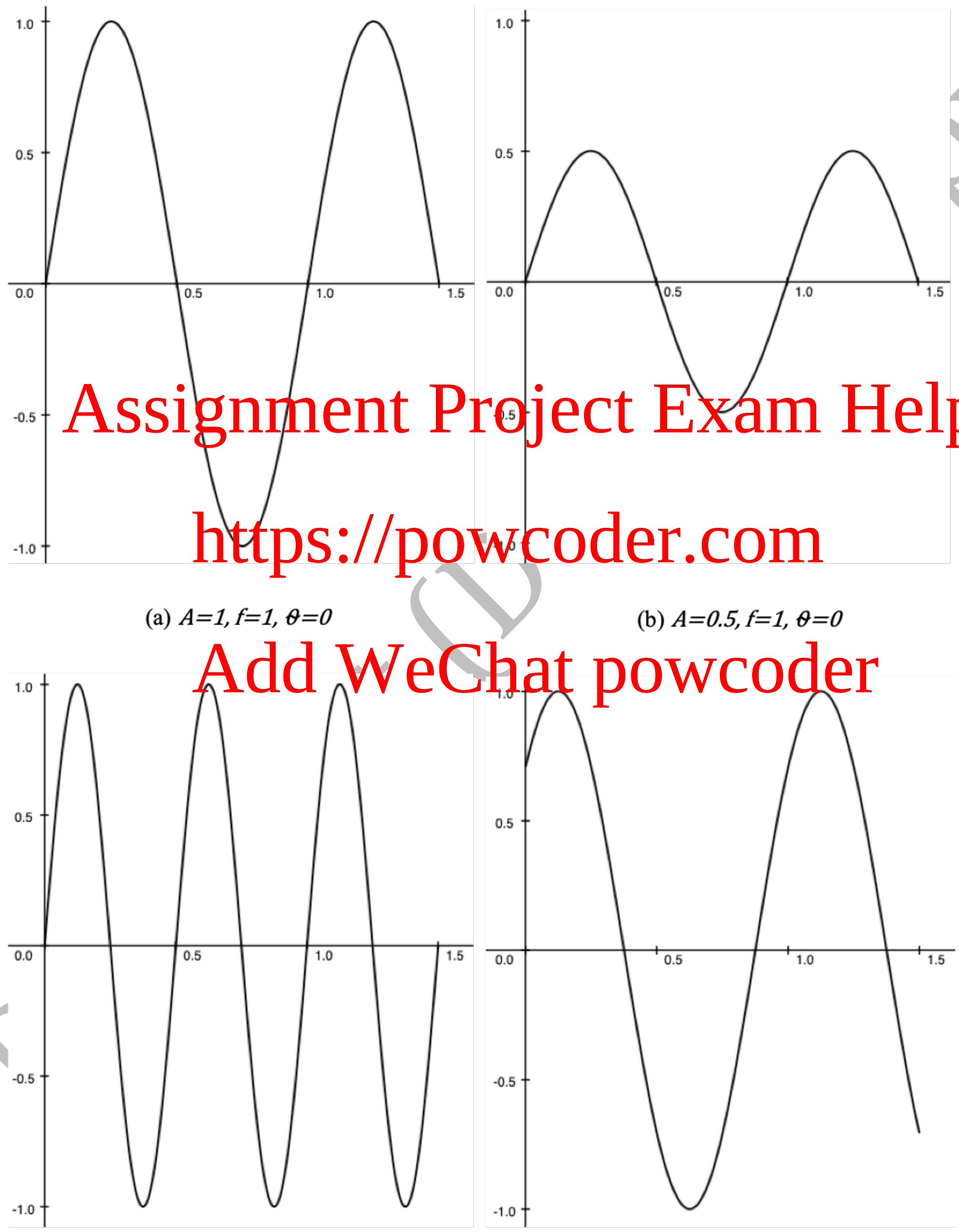
In the simplest form, a wave is mathematically represented by a sine wave $A\sin(2\pi ft + \theta)$, where A = Amplitude, f = Frequency, θ = Phase, and Period T = $1/f$. The small t is the current time, which allows us to obtain the value of the wave at any time using this formula.

Figure 2.2 illustrates the frequency, amplitude, and phase of a sine wave. Amplitude is the height of the wave, measured from zero to the maximum value, either up or down. Note that the sine wave is cyclic, i.e., it keeps repeating the pattern. One complete pattern is called a *cycle*.

Frequency is measured in cycles/sec or Hertz or simply Hz. For example, if a wave completes 1 cycle per sec, like the wave shown in Figure 2.2(a), then it has a frequency of 1 Hz. On the other hand, the wave in Figure 2.2(c) has a frequency of 2 Hz.

Phase is the amount of shift from a given reference point. For example, if we consider zero amplitude as our reference point, then the waves that start at zero while gaining their amplitudes, like most of the cases in Figure 2.2, then their phase is zero. The maximum phase shift is 360°, i.e., if the wave is shifted by 360°, then its phase is back to zero again. For example, in Figure 2.2(d), phase is shifted by 45°. Usually, the phase is measured in *radians*, where 360°=2π radian. Therefore, a 45° phase in radian would be π/4.

Figure 2.1 Waves in the water.



(a) $A=1, f=1, \theta=0$ (b) $A=0.5, f=1, \theta=0$

(c) $A=1, f=2, \theta=0$ (d) $A=1, f=1, \theta=\pi/4$

Figure 2.2 Frequency, amplitude, and phase of a sine wave $A\sin(2\pi f t + \theta)$.

2.1.1 2D Representation of Phase and Amplitude

The phase can be represented on a 2D graph. A sine wave can be decomposed into its *sine* and *cosine* parts. For example, a sine wave with a phase of 45°, can be written as the summation of two parts:

$$
\begin{aligned}
\sin(2\pi ft + \frac{\pi}{4}) &= \sin(2\pi ft)\cos(\frac{\pi}{4}) + \cos(2\pi ft)\sin(\frac{\pi}{4}) \\
&= \frac{1}{\sqrt{2}}\sin(2\pi ft) + \frac{1}{\sqrt{2}}\cos(2\pi ft)
\end{aligned} \tag{2.1}
$$

The first part is called the **In-phase component I**, and the second part the **Quadrature component Q**. In this case, we have I=$\frac{1}{\sqrt{2}}$ and Q = $\frac{1}{\sqrt{2}}$. I and Q are then plotted in a 2D graph as shown in Figure 2.3. In this figure, the *x*-axis is sine and the *y*-axis cosine. With I=$\frac{1}{\sqrt{2}}$ and Q = $\frac{1}{\sqrt{2}}$, we get a single point in the graph, which has a length (amplitude) of 1 (=$\sqrt{I^2 + Q^2}$), and an angle (phase) of 45°.

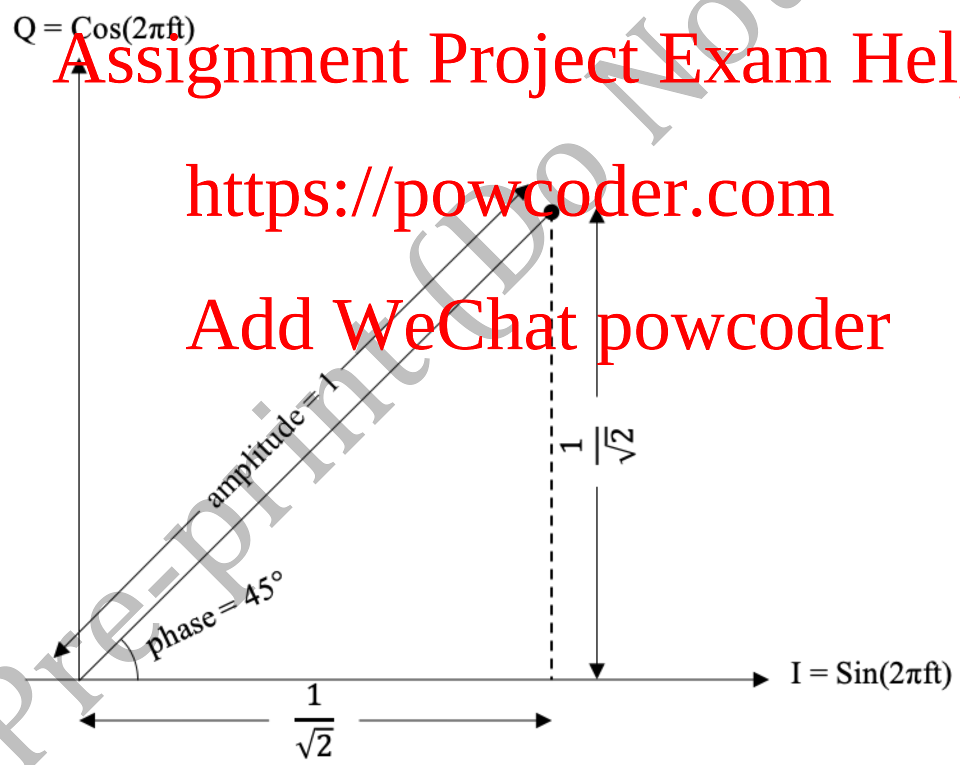


Figure 2.3 Phase and amplitude representation of sin(2πft+ π/4) in I-Q graph.

2.1.2 Wavelength

Waves propagate through space and cover distances over time. The distance occupied by one cycle is called the *wavelength* of the wave, and it is represented by λ. This is the distance between two points of corresponding phase in two consecutive cycles as shown in Figure 2.4.

In the air or space, all electromagnetic waves, irrespective of their frequencies, travel at the speed of light, which is a universal constant of *300 m/μs*. Given that it takes T sec for the wave to complete a cycle (T is called the period of the wave), and that $T=1/f$, we have

$$\lambda = cT = c/f \qquad (2.2)$$

Now we see that the wavelength is inversely proportional to its frequency.

Equation (2.2) is a universal formula that can be used to derive the wavelength for any types of communication medium. For example, for acoustic communications, which use sound waves to transmit data, the parameter c in Equation (2.2) should represent the speed of sound, which is only 343 m/s in dry air at 20º Celsius. Table 2.1 lists the wavelengths for some of the popular electromagnetic frequencies.

Table 2.1 Wavelengths of popular electromagnetic frequencies

Wireless Technology	Frequency	Wavelenth
IoT	915 MHz	32.7 cm
Bluetooth/WiFi	2.4 GHz	12.5 cm
Advanced WiFi	700 MHz	42.8 cm
	5 GHz	6 cm
	60 GHz	5 mm
Celluar (3G/4G/5G)	1.7 GHz	17.6 cm
	2.1 GHz	14.2 cm
mmWave and Terahertz for 5G and beyond	28 GHz	1 cm
	73 GHz	4 mm
	140 GHz	2.1 mm
	1 THz	0.3 mm
	10 THz	0.03 mm

Example 2.1
What is the wavelength of a 2.5 GHz electromagnetic signal propagating through air?

Solution

$$\begin{aligned} \text{Wavelength } = \lambda &= \frac{c}{f} \\ &= \frac{300 \text{ m}/\mu\text{s}}{2.5 \times 10^9} \\ &= 120 \times 10^{-3} = 120 \text{ mm} = 12 \text{ cm} \end{aligned}$$

Example 2.2
What is the frequency of a signal with 5 mm wavelength?

Solution
Wavelength = λ = 5 mm
Frequency = $f = c/\lambda$
= $(3 \times 10^8$ m/s) / $(5 \times 10^{-3}$ m)
= (300 x 109) / 5 = 60 GHz

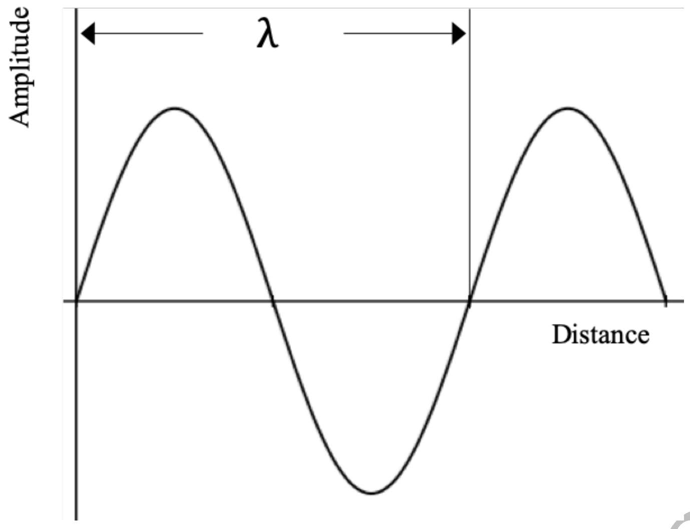


Figure 2.4 Wavelength of a sinusoidal signal.

2.2 Time and Frequency Domains

So far, we have seen how to represent waves in time domain. It turns out, that every wave can be represented in both time and frequency domains. Given the time domain representation, we can convert it to its frequency domain representation, and vice versa.

Using three different sine waves, Figure 2.5 illustrates the conversion from time domain representation (left hand side) to frequency domain (right hand side). The top sine wave has a frequency 1 (*f*) and amplitude A. In the frequency domain it is therefore just a pulse at frequency *f* (*x*-axis) having a height A (*y*-axis). The second sine wave has three times the frequency as the original one, but one third its amplitude. Therefore, in the frequency domain, its pulse is located at 3*f* and has a height of A/2.

The third sine wave is actually a combination of the first and the second waves. One can actually just add the wave values at each time instant to derive the third one. In the frequency domain, it therefore has pulses at two frequencies. The pulse at frequency *f* has a height of A and the pulse at 3*f* has A/2.

The transformation of a wave from time domain to frequency domain is called **Fourier transform** and from frequency domain to time domain is called **inverse Fourier transform**. There are fast algorithms to do this, such as Fast Fourier Transform (FFT) and Inverse FFT (IFFT). Most mathematical packages, such as MATLAB has library functions for FFT and IFFT. In recent years, general purpose programming languages, such as Python, are also offering library functions for FFT and IFFT.

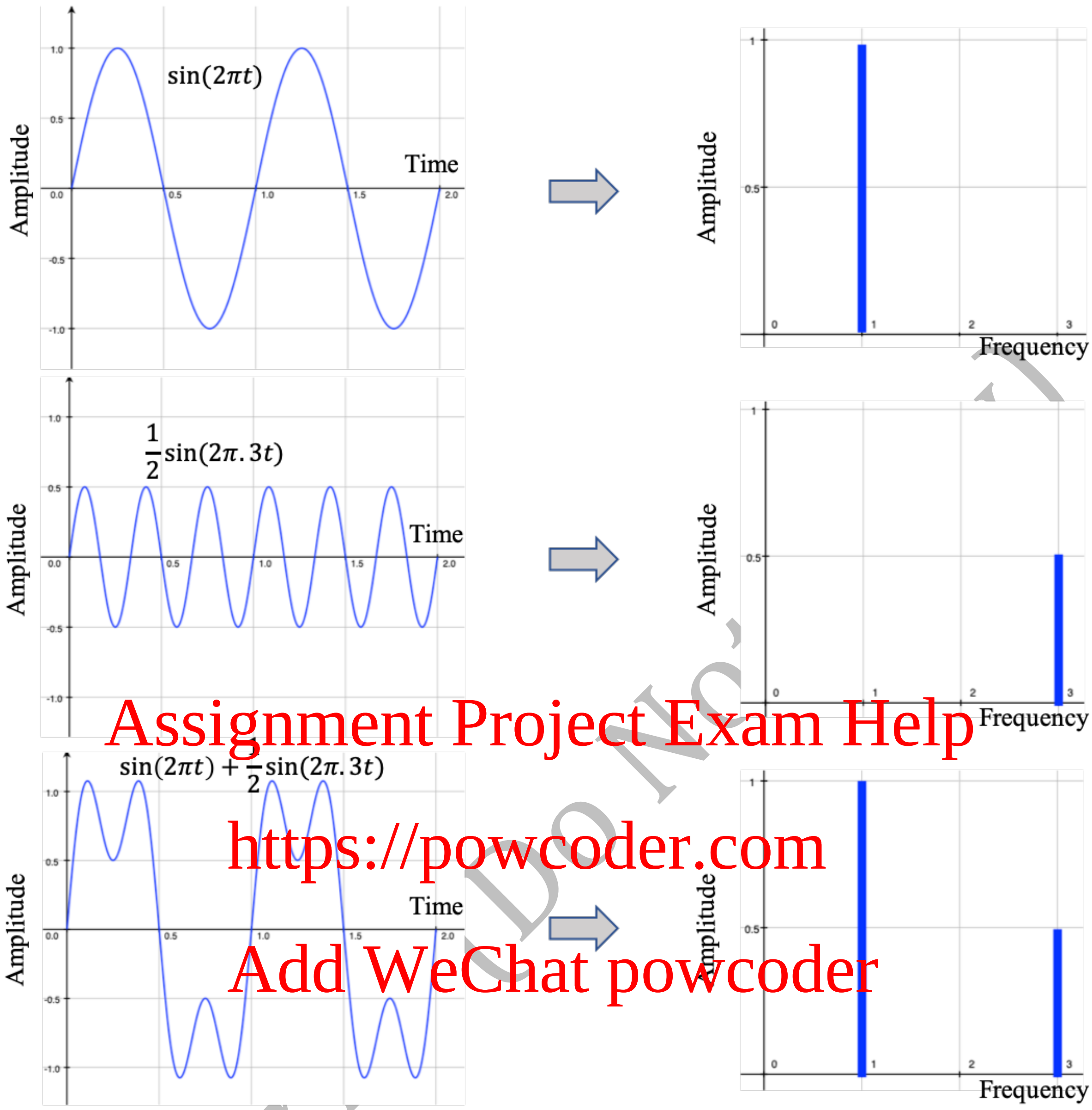


Figure 2.5 Time domain to frequency domain conversion.

2.3 Electromagnetic Spectrum

Wireless communications use the airwaves, which are basically electromagnetic waves that can propagate through the air or even in the vacuum. Any electricity or current flow will generate these electromagnetic waves. Therefore, many things we use generate or utilize some forms of electromagnetic waves. TV, power supply, remote control, microwave oven, wireless router, etc. all use or generate electromagnetic waves of different frequencies. Even the light is basically electromagnetic waves as we use electricity to generate light.

Electromagnetic waves can have a frequency of just 10 Hz, or 300 THz! The *spectrum* is all of the 'usable' frequency ranges. It is a natural resource and like most natural resources, it is limited. Spectrum use is therefore highly regulated by government authorities, such as the FCC in the US or ACMA in Australia.

A large portion of the spectrum is reserved for various government use, such as radar, military communications, atmospheric research and so on. The rest of the spectrum is often licensed to competing network operators, which give the operators exclusive rights to specific parts of the spectrum. For example, different TV channels or radio stations license different frequencies. Interestingly, part of the spectrum is also allocated for use without having to license it. Such spectrum is called licensed-exempt and sometimes referred to as 'free' spectrum. The spectrum used by Wi-Fi, such as the 2.4 GHz band, is a good example of such license-exempt spectrum. Table 2.2 lists some of the currently available license-exempt bands.

Table 2.2 Examples of license-exempt spectrum and their use

License-exempt Spectrum	Example use
433 MHz	Keyless Entry
900 MHz	Amateur Radio, IoT (e.g., LoRaWAN)
2.4 GHz	WiFi, Microwave Oven
5.2/5.3/5.8 GHz	WiFi, Cordless Phone

It is important to note that although manufacturers of any product can use license-exempt frequencies for free, they are subject to certain rules, such as power limitation for transmitting those frequencies. For example, the maximum transmit power of Wi-Fi products are often limited to about 100 mW depending on the region of operation.

Given the diverse needs of spectrum, spectrum allocation authorities must follow a set of principles when allocating spectrum. These principles include maximizing the spectrum utilization, adapt to market needs by promoting new technologies that may require some specific spectrum, promote market competition by strategically making certain spectrum license-exempt, ensure fairness in licensing spectrum among competing operators, as well as allocating spectrum to satisfy core national interests such as public safety, health, defense, scientific experiments and so on.

Certain services use specific *bands of frequencies*. For example, the basic Wi-Fi services use a frequency band of 2.4 GHz, which contains frequency in the range of 2.4 GHz to 2.5 GHz. Figure 2.6 shows the spectrum uses for different services. As we can see, historically wireless communications mostly use the spectrum between 100 kHz to 6 GHz. To meet the exponential demand for mobile data, the industry is now exploring spectrum beyond 6 GHz. For example, 60 GHz is now used in some Wi-Fi standards and the fifth generation (5G) mobile networks are targeting spectrum beyond 60 GHz.

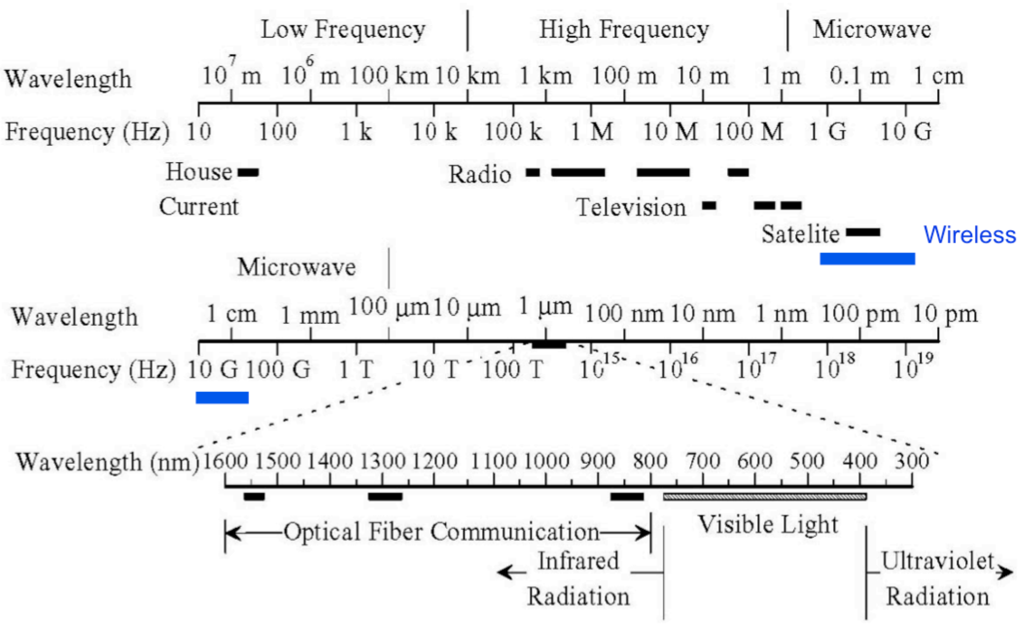


Figure 2.6 Spectrum allocation for different services. Wireless communication mostly uses 100 kHz to 6 GHz

2.4 Decibels

When waves travel, they lose power. We say that the power is *attenuated*. The question is: what would be a practical unit to measure power attenuation that is universal in all wireless communication systems?

Power loss for electromagnetic waves can be many orders of magnitude. For example, Wi-Fi chipsets can decode signals as weak as pico Watts, which allows them to offer reasonable communication coverage and range around the house or office buildings. Now imagine a signal that was transmitted by the Wi-Fi access point at full power of 100 mW but was received at a distant laptop with only 1 pW of power. The loss is one trillion folds!

Because the power loss can be many orders of magnitude, the attenuation is measured in logarithmic units. After the inventor Graham Bell, power attenuation was originally measured as *Bel*, where $Bel = \log_{10}(P_{in} / P_{out})$ with P_{in} representing the transmitted powered and P_{out} the attenuated power.

Bel was found to be too large for most practical systems. Later, a new quantity called *decibel*, written as dB, was introduced to measure power loss, where

$$\text{dB} = 10\log_{10}(P_{in} / P_{out})$$

Example 2.3
What is the attenuation in dB if the power is reduced by half (50% loss)?

Solution
Attenuation in dB $= 10\log_{10}(2) = 3$ dB

Example 2.4
Compute the loss in dB if the received power of a 100mW transmitted signal is only 1 mW

Solution
Power is reduced by a factor of 100. Attenuation = $10\log_{10}(100)$= 20 dB

The concept of decibel is also used to measure the *absolute* signal power, i.e., decibel can be used to measure the strength of a transmitted or received signal. In that case, it is a measure of power *in reference to* 1 mW and the unit is dBm. In other words, dBm is obtained as:

$$\mathbf{dBm = 10\log_{10}(power\ in\ milliwatt)}$$

Example 2.5
Convert 1 Watt to dBm.

Solution
We have $10\log_{10}(1\,\text{W}/1\,\text{mW}) = 10\log_{10}(1000\,\text{mW}/1\,\text{mW}) = 30\,\text{dBm}$

Example 2.6
Express 1 mW in units of dBm

Solution
$10\log_{10}(1) = 10\times 0 = 0\,\text{dBm}$ (ZERO dBm does not mean there is no power!)

Now we have a way to convert various electrical measurements, which are measured in Watts, into "networking measurements", which is in dB and dBm. Another important networking measurement is *noise*, which is often produced at the receiver due to the movement of electrons in the electronic circuits. The presence of such noise can make it difficult to decode data from the received signal if the signal-to-noise (SNR) is too low. Because decibel (dB) is basically a method to measure a ratio, it is also used to measure the SNR in wireless communications.

Example 2.7
With 100 µW of noise, what would be the SNR in dB if the received signal strength is 1 mW?

Solution

P_{signal} = 1 mW (received signal strength), P_{noise} = 100 µW

$$SNR = 10\log_{10}(1000/100) = 10\log_{10}(10) = 10\,\text{dB}$$

Example 2.8
Received signal strength is measured at 10 mW. What is the noise power if SNR = 10 dB?

Solution

$$SNR = 10dB = 10\log_{10}(10mW / P_{noise})$$

$$P_{noise} = 1mW$$

Now we can see that decibel is a versatile method to measure three different wireless communication phenomena, path loss, signal strength, and SNR. We have also seen that transmission or received power is measured in dBm, while path loss or antenna gain is measured in dB. It should be noted that dB can be added or subtracted to dBm, which would produce dBm again. For example, for a transmission power of 20dBm and a path loss of 25dB, the received power can be calculated as 20dBm-25dB = -5dBm. Similarly, if there was a 5dBi antenna gain for the previous example, the received power would be calculated as 20dBm+5dBi-25dB=0dBm.

It **does not,** however, make any sense to add dBm with dBm. For example, given P1=20dBm and P2=20dBm, it would be incorrect to say that P1+P2=20dBm+20dBm = 40dBm. The correct way to add powers would be to add them in linear scale and then convert the final value back to dBm. For the previous example, we have P1=100mW and P2=100mW, so P1+P2=100mW+100mW=200mW, which is only 23dBm!

2.5 Coding Terminology

The following terminology is often used to explain the coding of digital data on the carrier signal.

Symbol is the smallest element of a signal with a given amplitude, frequency, and phase that can be detected. Shorter symbol duration means that more signal elements carrying bits can be transmitted per second, and vice versa.

Baud rate refers to the number of symbols that can be transmitted per second. It is the inverse of symbol duration and hence sometimes referred to as the symbol rate. It is also called the modulation rate because this is how fast the property of the signal, i.e., its amplitude, frequency, or phase, can be changed or modulated.

Data rate, measured in bits per second, is the number of bits that can be transmitted per second. For example, for a binary signal, only 1 bit is transmitted for a given signal status, i.e., only 1 bit is carried over a baud and hence baud rate and data rate are equivalent. However, an M-ary signal has M distinct symbols and hence can carry $\log_2 M$ bits per baud or symbol. As we will see shortly, most modern wireless modulation techniques transmit multiple bits per symbol.

2.6 Modulation

Carrier waves are usually represented in sine waves. Data can be sent over a sine carrier by modulating one or more properties of the wave. As we have learned earlier in this chapter, there are three main properties of a wave, *amplitude*, *frequency*, and *phase*, that we can modulate.

Figure 2.7 shows how 0's and 1's can be transmitted over a sine carrier by modulating one of these three properties. When amplitude is modulated, it is called Amplitude Shift Keying (ASK). Similarly, we have Frequency Shift Keying (FSK) and Phase Shift Keying (PSK). In such modulations, the value of amplitude, frequency, and phase remains constant during a fixed period called *bit (or symbol) interval*. The receiver observes the signal value during this bit interval to demodulate the signal, i.e., extract the bit or bits transmitted during that interval. Note that the bit interval is essentially the inverse of the baud rate.

In Figure 2.7, we used only *two* different values of the amplitude, frequency, or phase to represent 0's and 1's. Therefore, we can send only 1 bit per different value of the signal, i.e., 1 bit per baud. In practical communication systems, usually more than 1 bit is transmitted per baud. For example, if we can modulate the amplitude in a way so we have four different values of the amplitude, then we need 2 bits to represent each amplitude, enabling us to transmit two bits per baud.

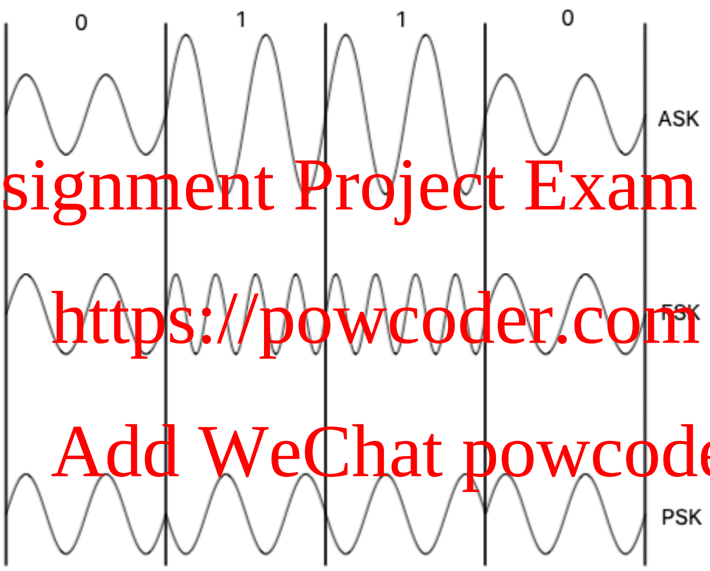


Figure 2.7 Modulation of a sine wave.

For PSK, the phase values can be *absolute*, or the *difference in phase* with respect to the previous phase. In Figure 2.8, the top graph shows that when there is no change in phase from the previous bit-interval to the next, it is treated as a 0 and when there is a change it is a 1. This is called *differential* BPSK. With differential, the receiver does not have to compare the phase against some pre-established value, but rather observe the change only, which is easier to implement.

The top graph in Figure 2.8 also shows that the phase is shifted by 180º and there is only one value to change. The corresponding 2D (I-Q) graph shows that the two dots are 180º apart. The bottom graph shows that the phase can switch to any of the four different values, which is called Quadrature Phase Shift Keying (QPSK). In QPSK, 2 bits can be sent per baud. Here in the I-Q graph, there are four dots and they are separated by 90 degrees.

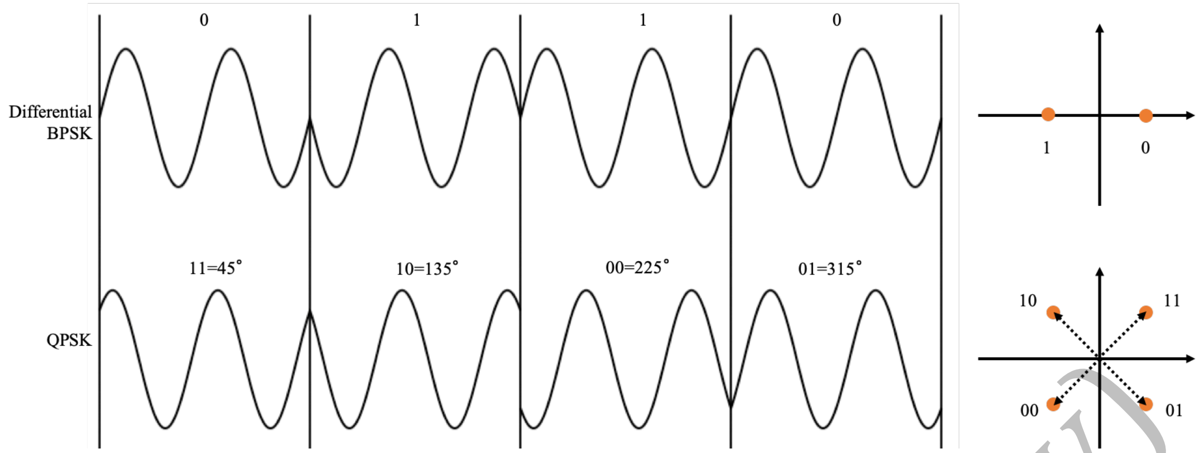


Figure 2.8 Differential BPSK and QPSK

2.7 QAM

To push the data rate even higher, we can combine amplitude and phase modulations together, which is called Quadrature Amplitude and Phase Modulation (QAM).

Note that in QPSK, the amplitude was kept constant. However, we could vary amplitude and get more than 2 bits per baud. A *constellation diagram* is often used to visually represent a QAM. Using constellation diagrams, Figure 2.9 shows three examples of amplitude and phase combinations to achieve different levels of QAMs. In the left most graph, we have constant amplitude, but 2 different phases. In total we have 1x2 = 2 combinations, so we get 1 bit per baud or 1 bit per symbol.

In the middle graph, we have four different phases, but just 1 amplitude. This is actually QPSK, but we could call it 4-QAM. It has a total of 1x4 = 4 combinations, so we have 2 bits per symbol. In the third graph (16-QAM), we have 3 different amplitudes; there are 4 different phases for each of the smallest and the largest amplitudes while the medium amplitude has 8 different phases. Thus, from 3 different amplitudes and 12 different phases, we use a total of 4+4+8 = 16 combinations of amplitudes and phases, which allows us to transmit 4 bits per symbol.

It is clear that we can increase the bit rates by going for higher QAMs. Table 2.3 lists the use of different QAMs in practical wireless networks, which shows that latest wireless standards employ as high as 1024 QAMs. As hardware and signal processing technology improves, we can expect even higher QAMs in the future.

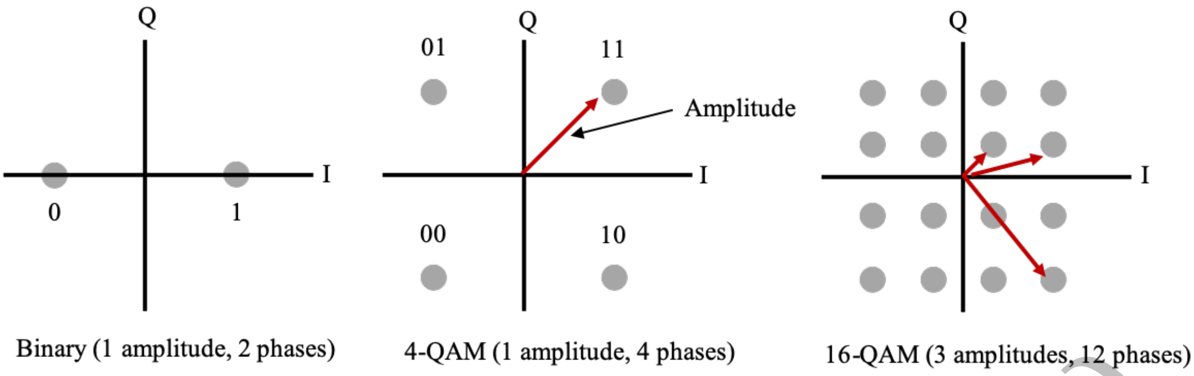


Figure 2.9 Constellation diagrams illustrating 2, 4, and 16 QAMs.

Table 2.3 QAMs used in practical wireless technologies

Wireless Technology	Supported QAM Technique
4G	256-QAM
5G	1024-QAM
WiFi 802.11n	16-QAM, 64-QAM
WiFi 802.11ac	256-QAM
WiFi 802.11ax	1024-QAM

2.8 Channel Capacity

The capacity of a channel basically refers to the maximum data rate or the number of bits that can be reliably transmitted over the channel. There are two basic theorems that explain the capacity, one by Nyquist and the other by Shannon. Both provides formulas to calculate channel capacity in terms of bits per second, but in slightly different contexts.

2.8.1 Nyquist's Theorem

Nyquist defines channel capacity under **noiseless** environment. In the absence of noise, receiving hardware can easily differentiate between a large number of different values of the symbol. As such, channel capacity according to Nyquist theorem is mainly constrained by the channel bandwidth, which defines the number of Hertz available in the channel. The higher the bandwidth, the more data rate can be achieved. Nyquist capacity is also dependent on the number of signal levels, i.e., the number of distinct symbols, used in encoding. More precisely, Nyquist capacity is obtained as:

$$\mathbf{Data\ rate = 2 \times B \times \log_2 M}$$

Where B is the channel bandwidth and M is the number of signal levels.

Example 2.9
Assume that you have discovered a novel material that has negligible electrical noise. What is the maximum data rate that this material could achieve over a phone wire having a bandwidth of 3100 Hz if data was encoded with 64-QAM?

Solution
We have

B= 3100
M = 64
Data rate=$2\times3100\times\log_2 64$=37,200 bps

2.8.2 Shannon's Theorem

Nyquist's theorem is valid for perfect noiseless channel. With perfect channel, we can get an infinite number of bits per Hz by coding data with an infinite number of signal levels. However, in reality, channels are noisy, which makes it difficult for the receiver to confirm exactly what signal value was transmitted. The noise, therefore, puts an upper limit on the number of bits we can transmit reliably. This *upper limit* is called Shannon's capacity and is obtained as:

Shannon's capacity=B $\log_2(1+S/N)$ bps

where B is the bandwidth of the channel, S is the received signal strength in Watt and N is the noise power in Watt.

Example 2.10
For an SNR of 30 dB, what is the maximum data rate that could be achieved over a phone wire having a bandwidth of 3100 Hz?

Solution
$10 \log_{10} S/N = 30$
$\log_{10} S/N=3$
$S/N = 10^3 = 1000$
Shannon's Capacity = $3100 \log_2(1+1000)$=30,894 bps

2.9 Hamming Distance and Error Correction

When data is transmitted over noisy channel, there is possibility of receiving the data in error. Using appropriate algorithms, such errors can be detected and even corrected. Hamming distance is a fundamental concept used by these error detecting and correcting algorithms. It is defined as the number of bits in which two equal-length sequences disagree. This can be easily obtained by applying the XOR operator between the two sequences.

Example 2.12
What is the Hamming distance between 011011and 110001?

Solution

Sequence 1:	011011
Sequence 2:	110001

Difference (XOR)	101010 → Hamming Distance = 3 (i.e., number of 1's in XOR output)

Data is usually *coded* and the codeword, which is longer than the data, is sent for error detection and correction purposes. Let us have a look at some examples.

Table 2.4 shows the codewords for the data bits where 2-bit words are transmitted as 5-bit words. Now let us assume that the receiver has received 00100, which is not one of the valid codewords. This means there was an error in the transmission.

Now let us look at the hamming distance between the received sequence and each of the valid codewords.

Distance (00100,00000) = 1   Distance (00100,00111) = 2
Distance (00100,11001) = 4   Distance (00100,11110) = 3

It is clear that most likely 00000 was sent, because it has the smallest hamming distance. Hence, the received sequence is corrected to data 00.

Now let us assume that the received sequence was 01010. We have, Distance (01010,00000) = 2 = Distance (01010,11110). There are two codewords at equal distance from the received sequence. In this case, error is detected but cannot be corrected.

Three-bit errors will not even be detected. For example, a 3-bit error could convert the transmitted codeword 00000 to 00111, which is also a valid codeword!

The lesson is, if we want to detect x-bit errors, any two codewords should be apart by a Hamming distance of at least x+1. Similarly, to correct x-bit errors, the minimum Hamming distance required is 2x+1. These rules provide valuable guidelines for designing codewords.

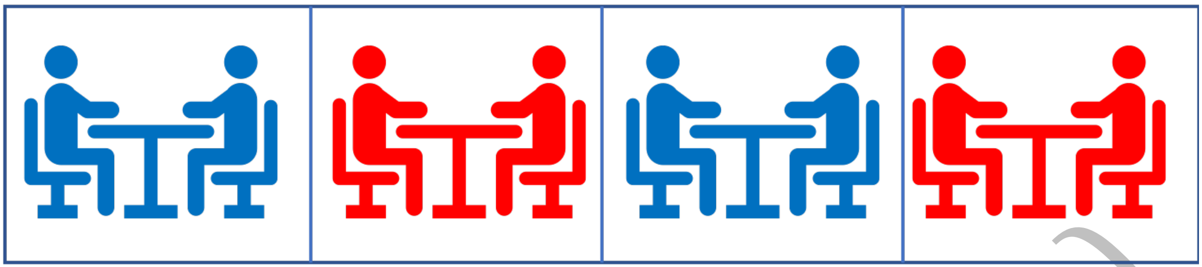
Table 2.4 Coding example for error correction

Data	Codeword
00	00000
01	00111
10	11001
11	11110

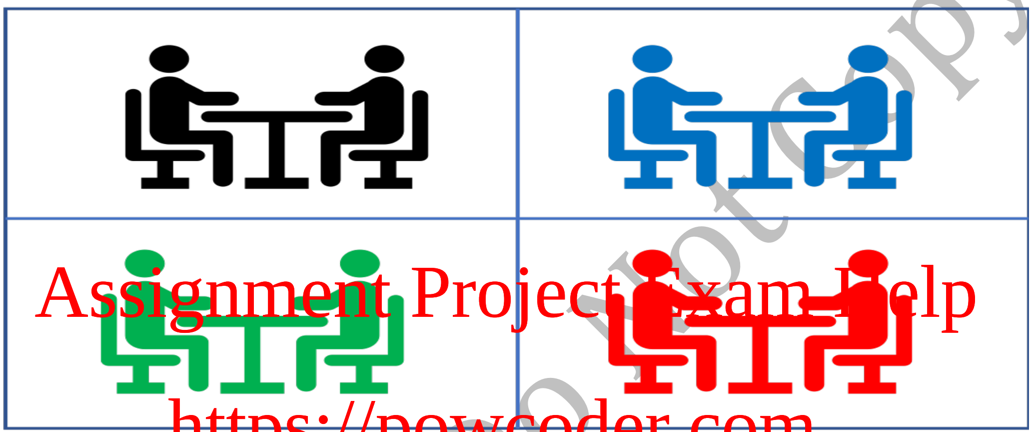
2.10 Multiple Access Methods

When the communication resources, such as a given frequency band, has to be shared by many devices, there has to be some rules to be followed, so all can enjoy interference-free communication. These rules constitute *medium access control*.

There are three fundamentally different medium access methods. They are based on *time*, *frequency*, or *code*. As such, they are called time division multiple access (TDMA), frequency division multiple access (FDMA), or code division multiple access (CDMA). Using an analogy of communicating groups of people, Figure 2.10 illustrates how these three multiple access methods differ from each other. At the top, we can see that the groups take turn, so they do not collide with each other. This is TDMA, because time is used to separate them. In the middle, the groups are located at different rooms, i.e., using different frequencies, so they do not collide with each other despite talking at the same time. Interestingly, at the bottom of the figure, all groups are talking at the same time in the same room, yet they do not really interfere with each other! This is possible because different groups are talking in different languages where people can still pick up their conversations because other languages simply appear as noise to them. Language is the *code* here to avoid interference.



(a) TDMA: Communicating groups are taking turns



(b) FDMA: Communicating groups are all talking at the same time, but in different rooms

(c) CDMA: Communicating groups are all talking at the same time in the same room

Figure 2.10 Multiple access methods

2.11 Spread Spectrum

Spread spectrum refers to techniques that spread an original narrow bandwidth signal over a much wider bandwidth. There are many purposes for spread spectrum, including improving security by making the signal harder to detect with a narrowband receiver, increasing resistance to interference, noise, and jamming, and to achieve CDMA. *Frequency hopping* and *direct sequence* are two popular spread spectrum techniques used in many communications systems.

2.11.1 Frequency Hopping Spread Spectrum

In frequency hopping spread spectrum (FHSS), both the transmitter and the receiver use a specific random sequence to switch frequency within the band every small interval as shown in Figure 2.11. Because the transmitter and the receiver use the same seed and random number generator, they have the exact same sequence, so they change the frequency at the right time to the right frequency and hence decode each other without any problem.

Because FHSS never stays on the same frequency for long, it is hard to jam the transmission using a particular frequency. For the same reason, it is also difficult to eavesdrop using a narrowband receiver that monitors a particular frequency. It can also be used for CDMA because the random seed for switching frequency becomes a code and those who do not have the code cannot decode the communication.

The most common and widespread use of FHSS is in Bluetooth, which switches frequency 1600 times per second or stays only 625 micro second over one particular frequency to avoid interference with nearby Wi-Fi devices that also use the same frequency band. FHSS is also used in military communications due to its anti-jamming and security features. Previous generations of cellular systems used FHSS to share the same frequency over many users at the same time.

Interestingly, FHSS was patented by an actress, Hedy Lamarr, who conceived the idea when she was playing on piano, where the tone is changed all the time. The patent, however, was not widely known until it was used in wireless communications.

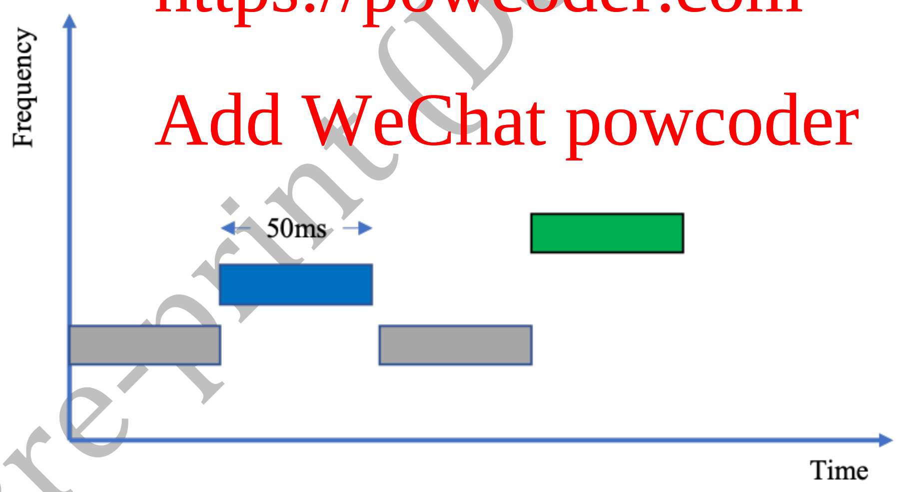


Figure 2.11 Frequency hopping. The transmitter switches frequency every 50 ms.

2.11.2 Direct-Sequence Spread Spectrum

Another way to achieve spread spectrum is to use a technique called Direct-Sequence Spread Spectrum (DSSS). The idea, as shown in Figure 2.12, is to expand a ‘0’ or ‘1’ by long series of 0’s and 1’s by applying a secret code to them. Then the resulting codeword is transmitted in place of a ‘0’ or ‘1’ in the data bit sequence.

The receiver knows the secret code used by the transmitter; hence it can retrieve the data bits from the codewords transmitted. In the example of Figure 2.12, a 10-bit code is used to create the codewords by applying the operation XOR to the data bits. In this case, 10 bits are actually transmitted over the channel to transmit a single data bit.

Spreading factor is defined as number of code bits per 0 or 1. FCC mandates that minimum should be 10, but for better privacy, much larger, such as hundreds are used. Military may choose to use thousands. The signal bandwidth therefore has to be orders of magnitude higher than the data bandwidth.

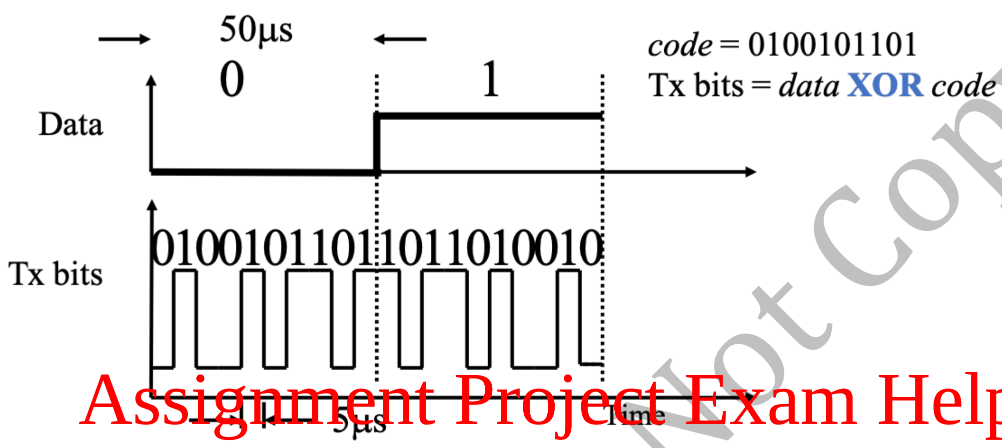


Figure 2.12 Direct-Sequence Spread Spectrum

2.12 Doppler Shift

The Doppler effect says that if the transmitter or receiver or both are mobile, the frequency of the received signal changes or shifts. This principle is named after Christian Andreas Doppler, an Austrian mathematician and physicist who first proposed the principle in 1842. If the transmitter is moving towards the receiver, then the receiver will receive higher frequency, i.e., the frequency is shifted to a higher value than originally transmitted. If the transmitter is moving away, the effect is opposite. This effect is illustrated in Figure 2.13, where a car is approaching a pedestrian, who is hearing higher frequency than what was generated by the car.

The next question is: how much the received frequency increases or decreases compared to the transmitted frequency? It turns out that this shift in frequency is a function of the relative velocity, v, between the transmitter and the receiver as well as the transmitter frequency or wavelength, which is obtained as follows:

$$\boldsymbol{Doppler\ shift = velocity/Wavelength = v/\lambda\ =\ vf/c}\ \textbf{Hz}$$

Example 2.13
Assume that a car travelling at 120km/hr is transmitting a packet to a roadside access point (AP) using 2.4 GHz Wi-Fi. If the car is approaching the AP (i.e., the AP is directly in front it), what is the frequency received by the AP?

Solution

The wavelength of the frequency is: $3 \times 10^8 / 2.4 \times 10^9 = 0.125$m
The velocity of the car is: 120km/hr = 120x1000/3600 = 33.3 m/s
Freq diff (Doppler shift) = 33.3/0.125 = 267 Hz

Therefore, the receive frequency at the AP is 2.4GHz + 267 Hz = 2.400000267 GHz

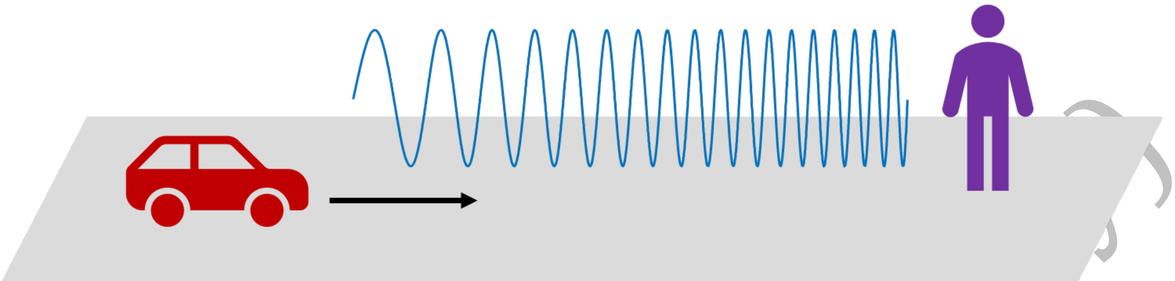


Figure 2.13 Effect of Doppler shift

2.13 Doppler Spread

Besides receiving the signal from the original source, such as a base station, a mobile object also receives signals reflected from other objects. Thus, a mobile object experiences Doppler shifts in both positive and negative directions as illustrated in Figure 2.14 in case of a moving vehicle. Thus, the received frequency is spread equally on both sides of the original frequency. Hence the Doppler spread is obtained as:

$$\boldsymbol{Doppler\ Spread = 2 \times Doppler\ shift = 2vf/c = 2v/\lambda}$$

2.14 Coherence Time

Coherence time refers to the time interval during which the channel does not change. This interval is important in optimizing many communication parameters. For example, during packet transmissions, usually there are some preamble signals used to probe the channel. Channel statistics gathered from this probe is then used to optimize the transmission parameters for the rest of the bits in the packet. Obviously, in this case the packet size has to be optimized so that it can complete within the channel coherence time, otherwise part of the packet will experience a different channel. Conversely, for long packets, multiple probes must be inserted within the packet to obtain the most up-to-date channel information.

Doppler spread is a frequency domain measure that influences the coherence time in a reciprocal relationship as follows:

$$\boldsymbol{Coherence\ time = 1/Doppler\ spread = \lambda/2v}$$

Therefore, higher the Doppler spread, shorter the coherence time, and vice versa. For example, doubling the frequency of transmission would halve the channel coherence time.

Example 2.14
What is the coherence time for a 2.4 GHz Wi-Fi link connecting a car travelling at 72 km/hr?

Solution

V= (72 x 1000)/3600 = 20 m/s
Doppler spread = 2*vf*/c = (2x20x2.4x109)/(3x108) = 320 Hz
Coherence time = 1/320 = 0.003125 s = 3.125 ms

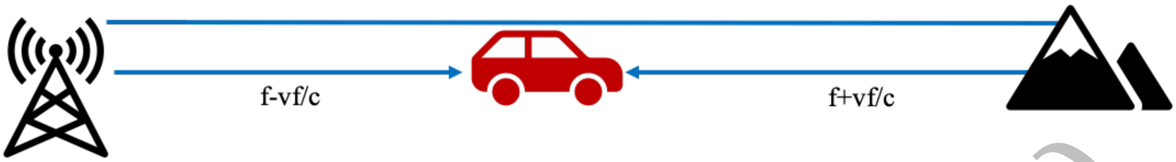


Figure 2.14 Illustration of Doppler spread experienced by a motorist.

2.15 Duplexing

Duplexing attempts to answer the following question: how the resource should be allocated between the transmitter and the receiver so they both can exchange information with each other, i.e., both can transmit and receive? Figure 2.15 shows that there are two ways to achieve this. One way is to allocate different frequencies for different directions. In this case both can talk at the same time, achieving full-duplex communications. This is called frequency division duplexing (FDD).

The other method is to use the same frequency for both directions, but only one entity can talk at a given time. For example, when the base station talks, the subscriber listens and when the subscriber talks, the base station listens. This method is called time division duplexing (TDD). Clearly, TDD cannot achieve full-duplex, but provides only a half-duplex communication. Despite this, many cellular deployments use TDD because it allows more flexible sharing of downlink (base station to subscriber) and uplink (subscriber to base station) resources without requiring paired spectrum allocation, which is wasteful if data is asymmetric in these two directions.

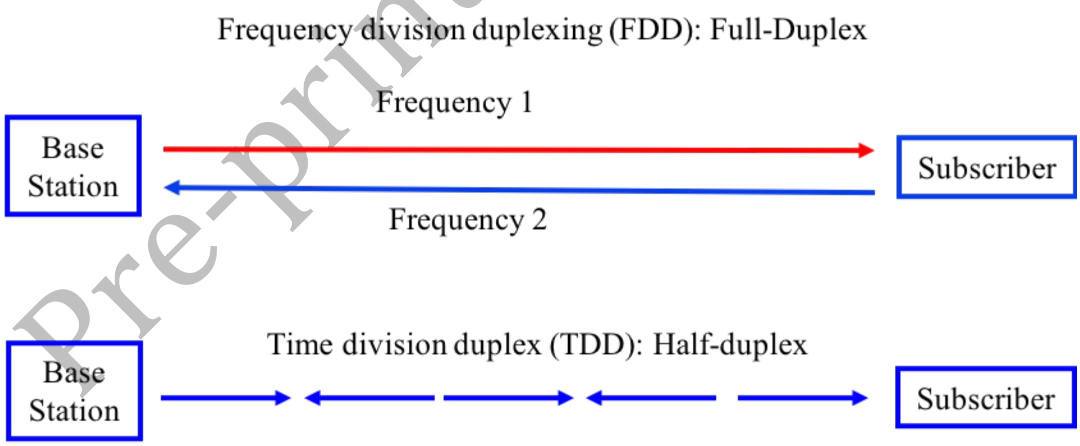


Figure 2.15 Frequency division vs. time division duplexing

2.16 Chapter Summary

1. Electric, Radio, Light, X-Rays, are all electromagnetic waves
2. Wavelength and frequency are inversely proportional ($\lambda = c/f$)
3. Historically, wireless communications mostly used frequencies below 6 GHz, but beyond 6 GHz is actively explored in modern wireless networks.

4. Hertz and bit rate are related by Nyquist and Shannon's Theorems
5. Nyquist's theorem explains capacity for noiseless channels
6. Shannon's capacity takes SNR into consideration
7. Power is measured in dBm and path loss or antenna gain in dB
8. dB can be added or subtracted to dBm to produce dBm, but dBm cannot be added or subtracted with dBm
9. By spreading the original signal bandwidth over a much wider band, spread spectrum can provide better immunity against interference and jamming as well allowing multiple parties to communicate over the same frequency at the same time.
10. FHSS and DSSS are two fundamental methods of realizing spread spectrum
11. Doppler effect explains the shift in frequency experienced by mobile objects
12. Doppler spread is twice the Doppler shift
13. Channel coherence time is inversely proportional to doppler spread
14. FDD and TDD are two fundamental methods of resource allocation between the transmitter and the receiver so they both can exchange information with each other

End of Chapter 2