

Signals

Alex S.*

1 Communications

Communication is generally accomplished via signals. Those can either be analog or digital. Now, whenever data travels through a wire (or air), it is analog. These notes provide an overview of how signals work.

2 Intuitive Intro

Think of how a guitar sounds. You pluck a string, and it vibrates. Each string vibrates at its own *frequency*. It might get louder or quieter depending on how you play it, but the frequency remains the same. Each one of those vibrations pushes and displaces air. This displacement of air is what reaches your ears—where it displaces a membrane (at that same frequency) that lets your brain know that there is a sound of a certain frequency.

Interesting fact: It turns out your ear can hear sounds from about 20Hz to approximately 20kHz. However, our voice only generates frequencies from 80Hz to about 3kHz.

3 Introduction To Signals

In the context of communications¹, a signal is basically some information somehow encoded as a wave. Everything travels as a wave, so let's look at some of the simple ideas of that. Figure 1 presents a basic *Sin* wave.

The vertical axis is the volume, or air-pressure, or strength, or energy, etc., depending on what this wave is. For example, if this is an ocean wave, then the vertical axis may be the wave height. If this is a sound wave in the air, then it is the air-pressure. If it is the sound wave in a wire, then it's the voltage.

The horizontal axis often represents the *time* or *distance* (or *space*) that the wave traverses in.

*alex@theparticle.com

¹Good reference: <http://www.phy.davidson.edu/instrumentation/NEETS.htm>

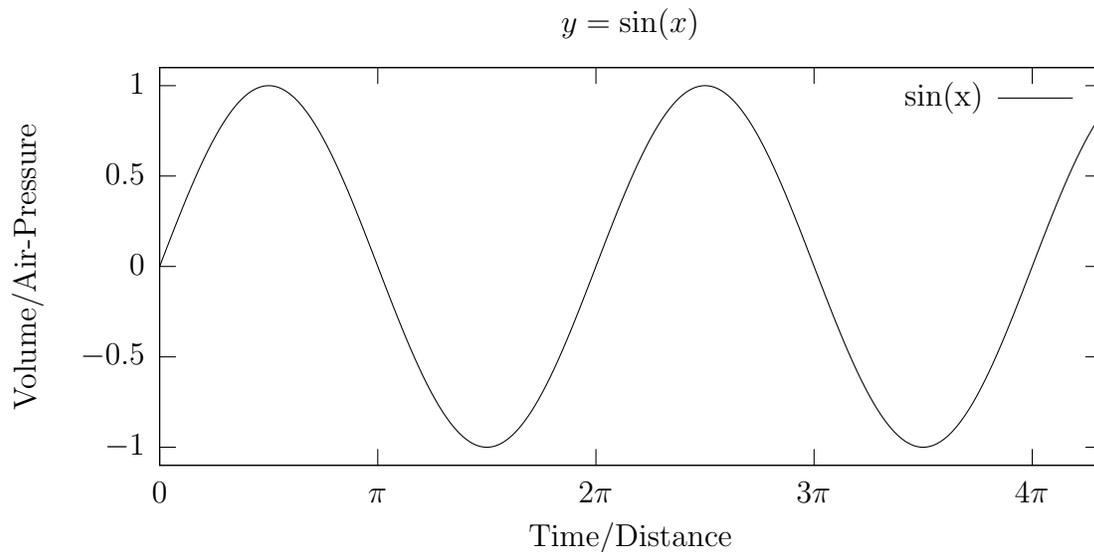


Figure 1: Analog Sin Wave.

Waves generally have a period, which can be viewed as the time/distance of when the wave repeats itself—starts the next iteration. In Figure 1 the period is 2π .

A closely related property of the period is the *frequency*. Frequency is the speed at which the wave cycles its periods, and is always expressed in proper speed units, as in ‘cycles per second’ (commonly known as Hz), and ‘kilo-cycles per second’ (commonly known as kHz)².

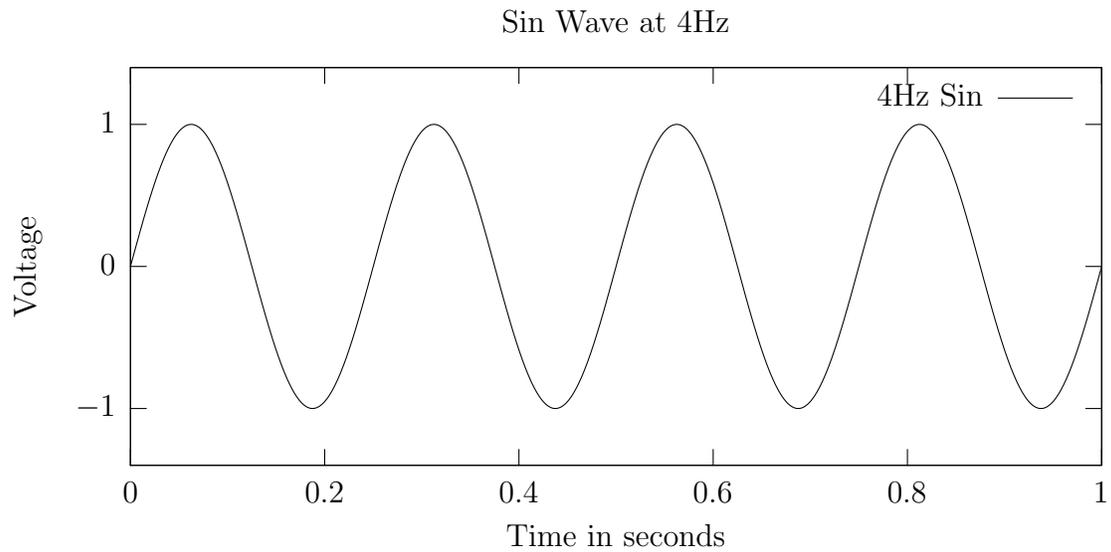
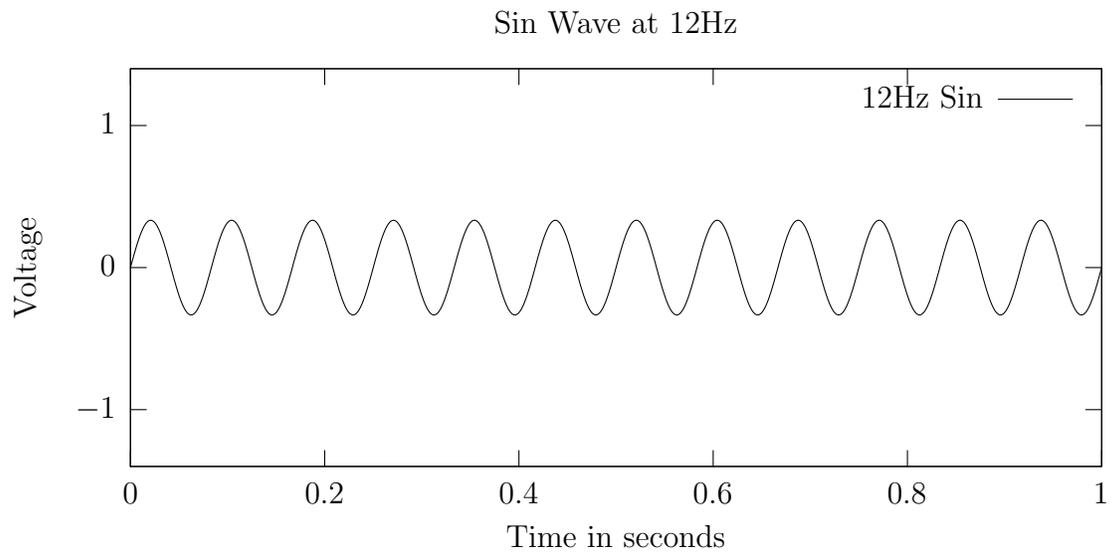
4 Time Domain

Normally, when we think of signals or waves, we think of the familiar ‘time domain’ wave. Figure 2 illustrates a *Sin* wave at 4Hz. Notice that the horizontal axis is *time*, and is expressed in *seconds*. You can count the number of times the wave cycles, and you’ll see that it’s 4, ie: 4Hz.

Given the frequency, you can easily find the wave period. For example, if the wave cycles itself 4 times per second, then the period is $\frac{1}{4}$ seconds, or simply $1/\text{Frequency}$. Similarly, knowing the period, you can find the frequency.

Another closely related value is the *wavelength*, which is basically the distance occupied by the wave period. Given the speed of propagation (light in a vacuum, or sound in the air, etc.), we can calculate that too. For example, light/electricity travel around 3×10^8 meters per second. Let’s imagine that Figure 2 represents electricity in a wire, then the wavelength is just $3 \times 10^8 \div 4$ or just 7.5×10^7 meters, or simply $\text{velocity} \div \text{frequency}$. Obviously you can find velocity if you know the wavelength and frequency.

²Naturally ‘kilo’ stands for 1000, so if something is 8kHz, then it’s 8000Hz.

Figure 2: 4Hz Sin Wave; $\sin(2\pi 4t)$ Figure 3: 12Hz Sin Wave; $\frac{1}{3} \sin(2\pi 12t)$

Continuing on with the discussion, Figure 3 shows a wave of 12Hz, and with $\frac{1}{3}$ amplitude. Notice how the period got smaller as the frequency increased.

Real life signals aren't such nice sin waves though. However, as it turns out, any wave is just a sum of many simple sin waves.

5 Spectrum

Very often we think of ‘energy’ (or radiation) in terms of frequency spectrum. For example, frequencies in increasing order would be: audio, radio, tv, microwaves, infrared, visible light ($10^{14} - 10^{15}$ Hz), ultraviolet, x-rays, gamma rays, etc.

6 Frequency Domain

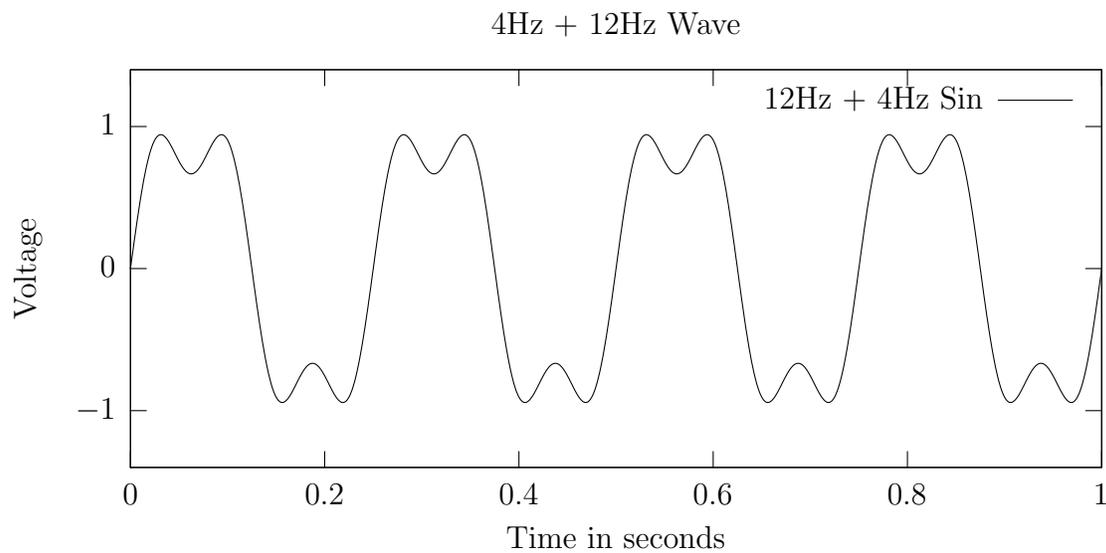


Figure 4: 4Hz + 12Hz Sin Wave.

Figure 4 illustrates the idea of taking two waves and adding them together. In this case, we’ve added waves from Figures 2 and 3.

There are primarily two ways of viewing any type of a wave; in the *time* domain, or in the *frequency* domain. The frequency domain of Figure 4 might look something like Figure 5.

It is important to realize that the two representations are really of the same thing. Notice that since we’ve added a 4Hz and a 12Hz signals, we have two peaks at those same frequencies. Normally, however, the frequency domain doesn’t look this clean.

6.1 Frequency Domain vs. Time Domain

It turns out that viewing signals in frequency domain is usually a lot more useful than viewing them in the time domain.

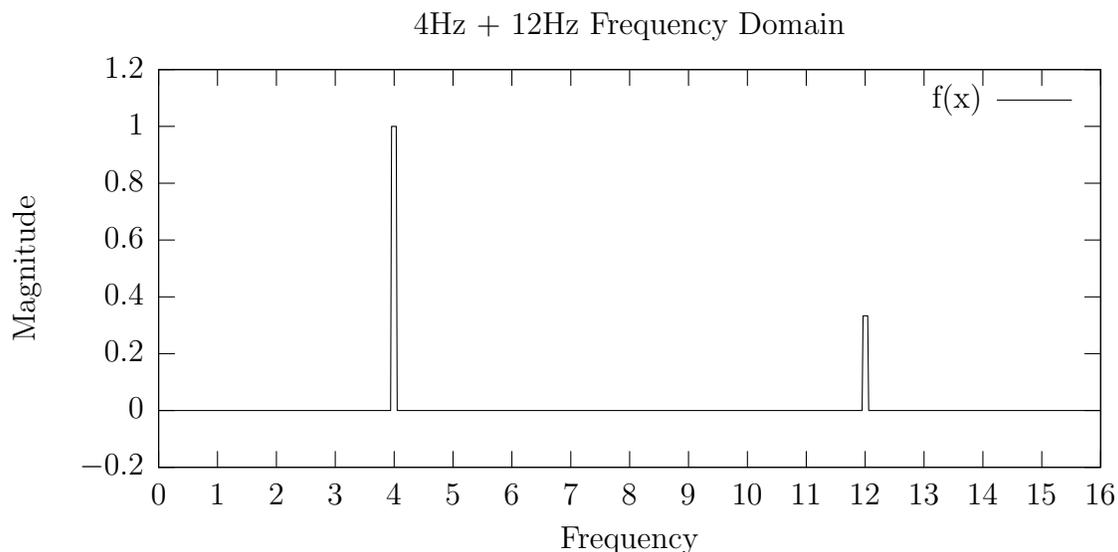


Figure 5: Frequency Domain of 4Hz + 12Hz Sin Waves.

6.1.1 Fourier Transform

We can go between frequency and time domains via the Fourier Transform (in this case, Discrete Fourier Transform):

$$H_n = \sum_{k=0}^{N-1} h_k e^{2\pi i k n / N}$$

6.1.2 Inverse Discrete Fourier Transform

We return to the original time-domain signal via the inverse fourier transform.

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{-2\pi i k n / N}$$

The transforms given are *discrete* transforms: means that they'd work on discrete data but not on continuous signals.

Now, there are many useful applications of these transforms outside of communications. It turns out that some mathematical operations are simpler in frequency domain. For example, things like multiplication.

These transforms are also used greatly for video/audio compression. A form of fourier transform is used in MP3 compression.

6.2 2D Discrete Cosine Transform

Other ‘similar’ transforms have been used for image compression. JPEG uses a transform known as Discrete Cosine Transform:

$$DCT(i, j) = \frac{1}{\sqrt{2N}} C(i) C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} Pixel(x, y) \cos \left[\frac{(2x+1)i\pi}{2N} \right] \cos \left[\frac{(2y+1)j\pi}{2N} \right]$$

Where $C(x) = \frac{1}{\sqrt{2}}$ if x is 0, else 1 if $x > 0$.

The inverse is just:

$$Pixel(x, y) = \frac{1}{\sqrt{2N}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C(i) C(j) DCT(i, j) \cos \left[\frac{(2x+1)i\pi}{2N} \right] \cos \left[\frac{(2y+1)j\pi}{2N} \right]$$

Where $C(x) = \frac{1}{\sqrt{2}}$ if x is 0, else 1 if $x > 0$.

A similar transform known as Discrete Wavelet Transform (DWT) is used in JPEG2000 standard.

To make a long story short: This concept is *very* important, and you should get a book on it and learn it. If you want to do anything that involves multimedia or any sort of engineering work, you *need* to know these things.

6.3 Spectrum Analyzer

A spectrum analyzer is a piece of equipment that lets you observe the frequency domain of a signal (ie: lets you see the component frequencies of a signal).

7 Bandwidth

Whenever we’re talking about signals and their frequencies, we must also mention *bandwidth*. Bandwidth is just a range of frequencies. Now, signals can be of various frequencies, for example, TV needs 6Mhz signal for each channel. There are 68 channels that we can tune to—which means we need what kind of bandwidth capability in our tuning equipment? That’s right: 6×68 Mhz.

Digital signals have an ‘infinite’ bandwidth—so all transmissions of digital signals need to account for the fact that the medium has a limited transmission capability.

8 Modulation

More often than not, whenever we’re transmitting anything over a wire, there is already a signal there; called the *carrier* signal. It has a specific frequency³. Interrupting or changing the carrier signal somehow enables the transmission of data. There are many different ways of accomplishing this.

³The signal is generated by an Occilator—a circuit that outputs a constant frequency; a ‘tone’.

8.1 Morse Code

For example, morse code, which is a series of ‘dits’ and ‘dahs’, is just a series of interruptions in the carrier signal.

8.2 AM and FM

In every case, we have a signal to transmit. We also have a constant frequency carrier signal. Amplitude Modulation is a way of varying the amplitude of the carrier signal. Frequency Modulation is a way of varying the frequency of the carrier signal.

There is also Phase Modulation, but we haven’t mentioned phases, so we’ll just let it be. It’s sort of similar to FM.

8.3 Frequency Domain

Modulation is relatively easy to visualize in frequency domain. For example, lets say we have a signal to send, which is 8Hz, as in Figure 6.

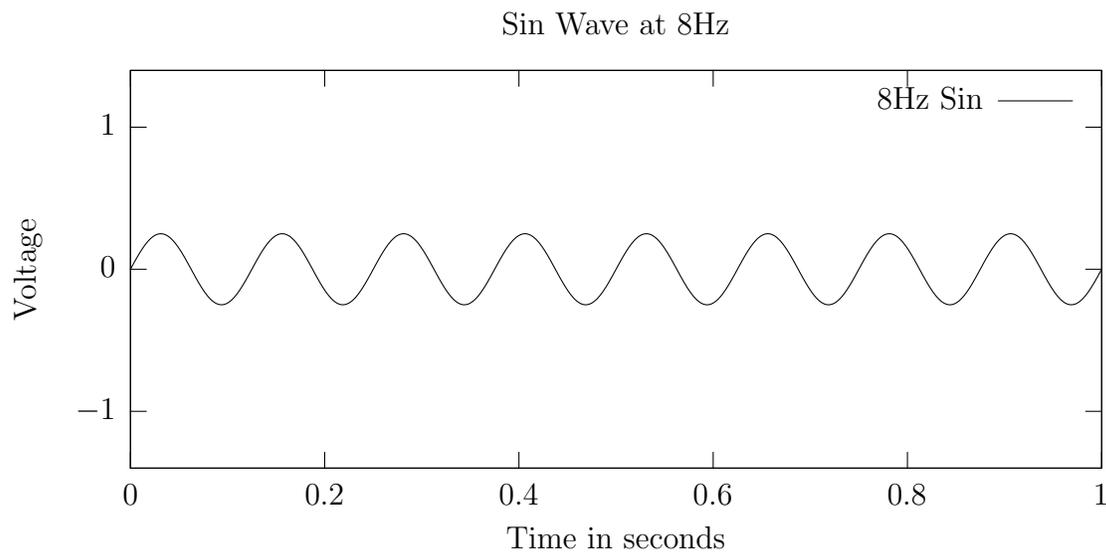


Figure 6: 8Hz Signal

Our carrier signal is something like 64Hz, as in Figure 7.

Now, Figure 8 shows a modulated signal. It is just a sum of both signals (Amplitude Modulation).

The frequency domain of this modulation may look something like Figure 9. This is basically the carrier frequency with a + and – the signal frequency.

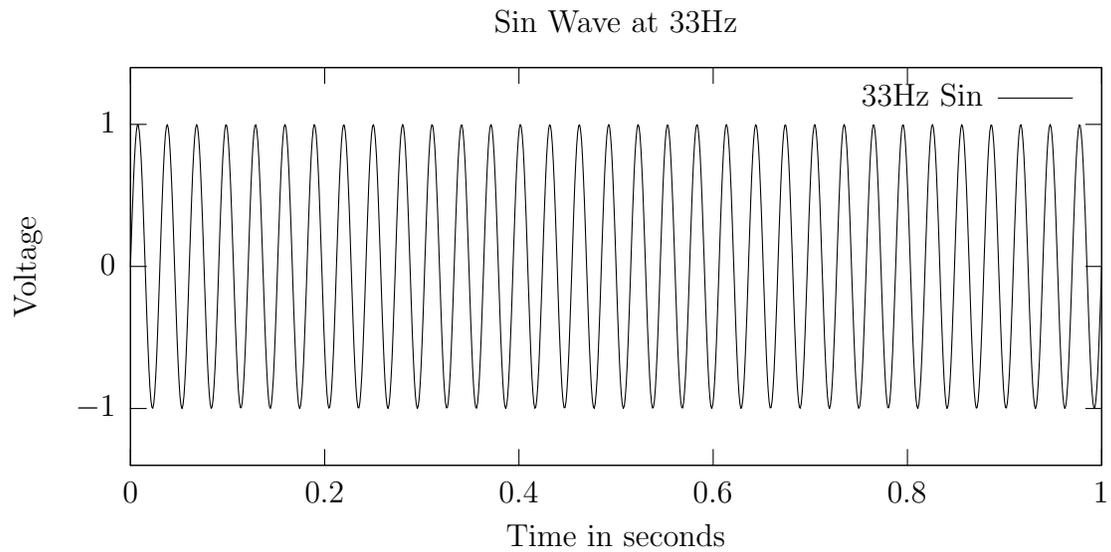


Figure 7: 33Hz Signal

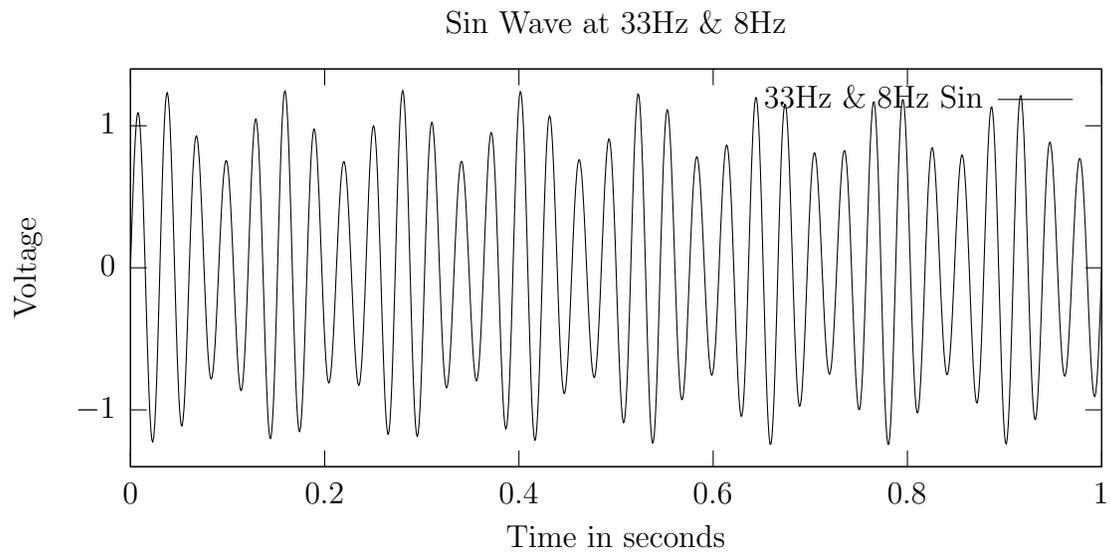


Figure 8: 33Hz & 8Hz Signal

9 Mixing, Channeling

Very often, we need to send more than one signal over the same wire. Now, we cannot use the same frequency for both signals, but we can certainly use different frequencies. What we can do is allocate a band of frequencies, then using a *mixer*, let each signal have it's own

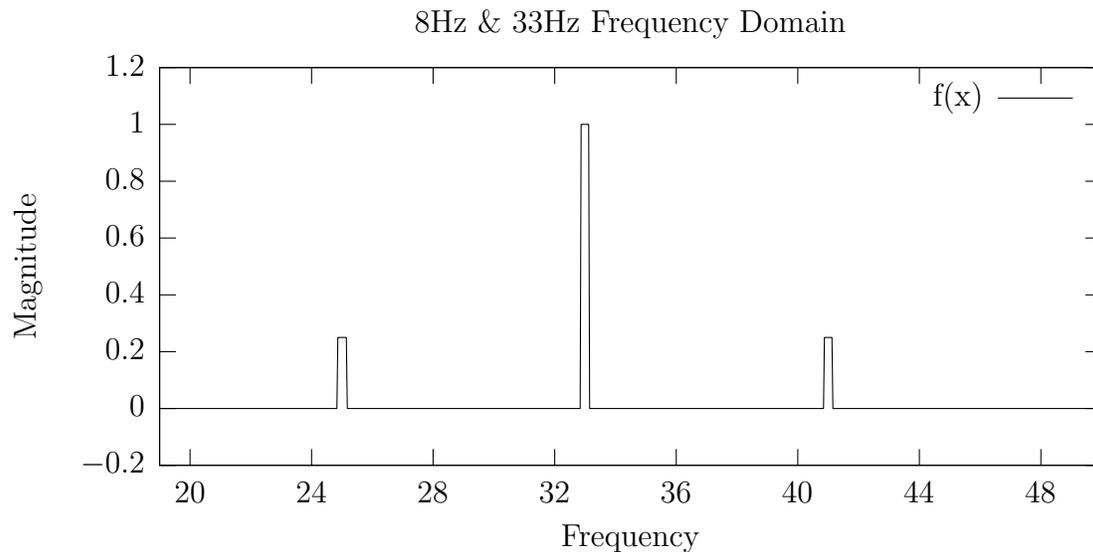


Figure 9: Amplitude Modulated 8Hz in 33Hz carrier; frequency domain.

‘channel’ (sorta like TV channels have their own frequency spot).

10 Analog vs. Digital

When we’re thinking of an analog wave, we’re thinking of something of the sort seen in Figure 4. When we’re thinking of a digital wave, we’re thinking of the sort illustrated in Figure 10.

Now, we can’t send a ‘digital’ signal over the wire, but we can send its approximation via an analog signal. So our goal is to create an analog signal that looks like a digital one. Let us start with Figure 4, and add on more frequencies to it. The result would look something like in Figure 11.

Dumping a few more frequencies, as in Figure 12, we have a relatively ‘square’ looking analog wave.

It turns out that a digital signal is just an *infinite* sum of various analog signals at various frequencies: $\sum \frac{1}{k} \sin(2\pi kft)$ for odd k .

11 A2A, A2D, D2A, D2D

There are all sorts of devices that do all sorts of things. Here, I’d like to mention a few:

A2A Analog to Analog: This would take the form of a Modulator. As seen before, modulators can do many things—like modulating the carrier signal, etc.

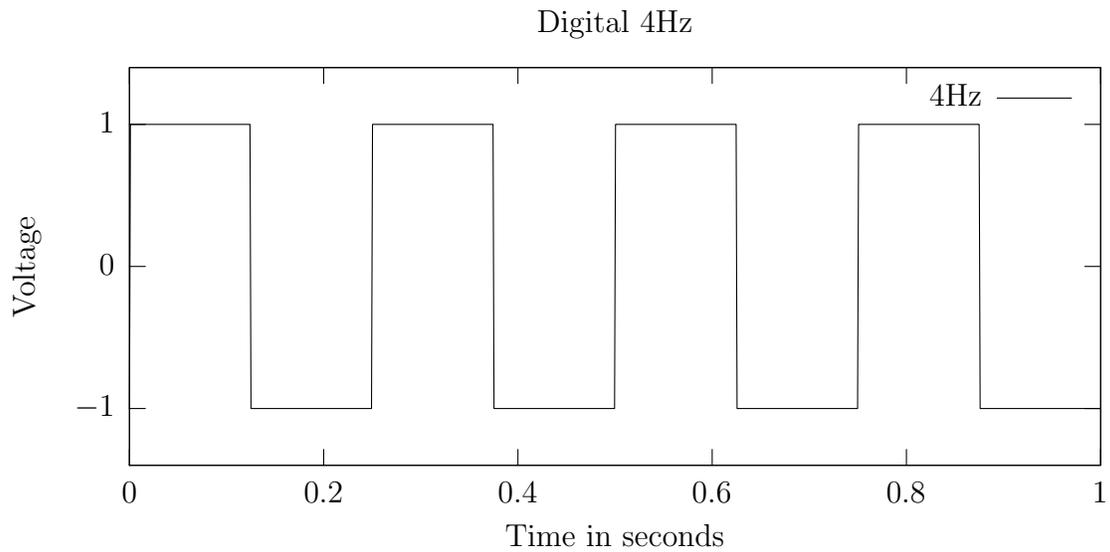


Figure 10: Digital 4Hz signal.

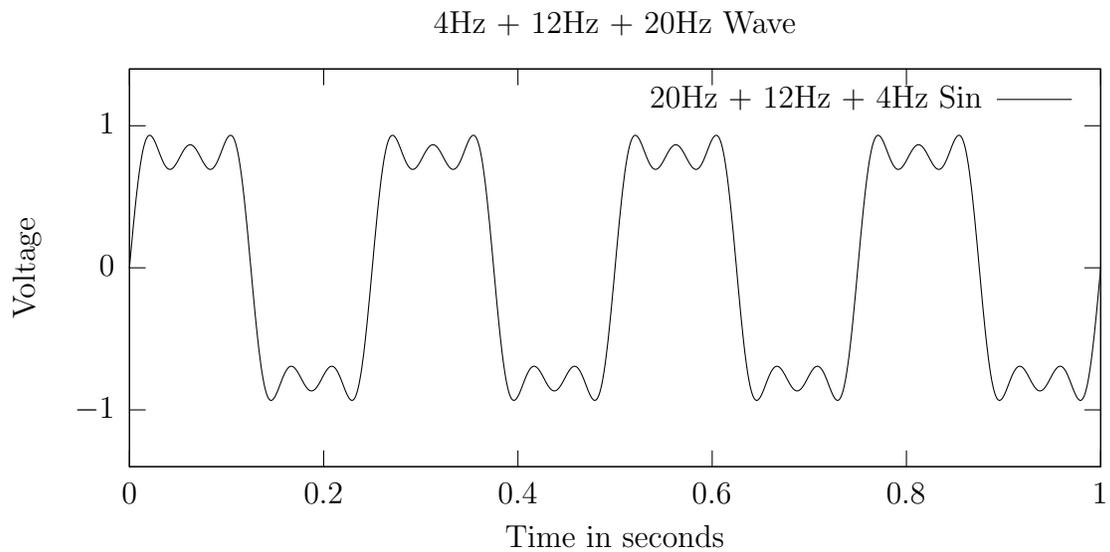


Figure 11: Composite of 20Hz, 12Hz, and 4Hz.

A2D A Codec, used to convert analog signal to a digital one. There is one in your modem.

D2A A 'Modem', used to convert from digital to analog. There is definitely one of these in your modem. There is also one on your graphics card—to turn a digital computer signal into one usable by the CRT monitor.

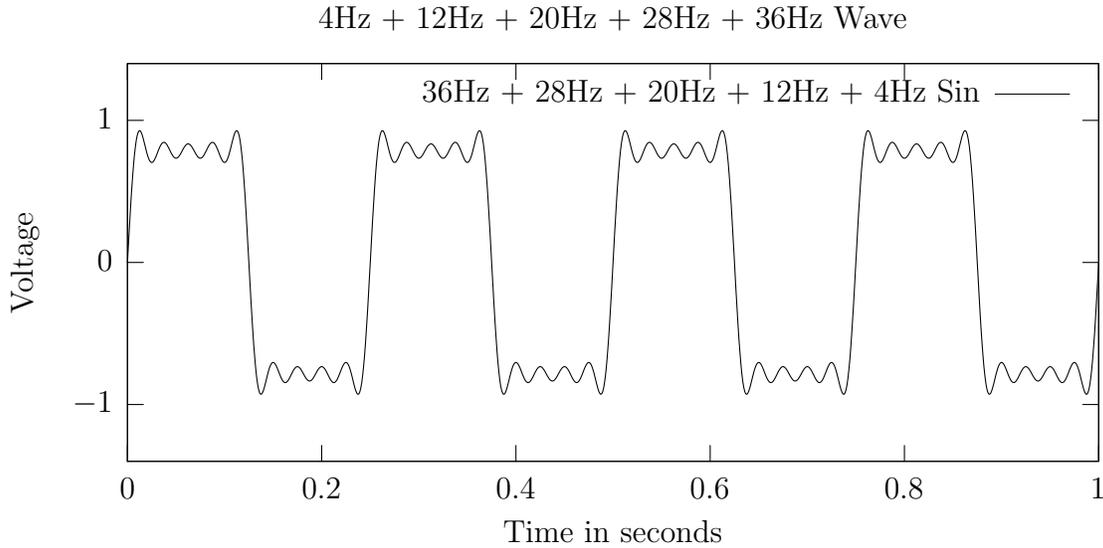


Figure 12: Composite of 36Hz, 28Hz, 20Hz, 12Hz, and 4Hz.

D2D This is a Digital Transmitter (transceiver). It is used to retransmit digital signals, or to change encodings, etc.

12 Transmission Impairments

12.1 Attenuation

The strength of the signal falls off with distance over any transmission medium. Over guided media, this is usually an exponential function, and is generally expressed in decibels per unit distance.

This means that the sender must ensure that the signal has enough strength to get to the end and be detected—or get to the nearest amplifier or repeater.

Higher frequencies usually suffer more than lower frequencies (think of a digital signal—the closer the pulses, the less ‘power’ is transmitted per pulse). This is often solved by having the amplifier amplify higher frequencies more than lower ones.

We can use the decibel (dB) to calculate the loss (or gain) of power during the transmission

$$dB = 10 \log_{10}(P_r/P_s)$$

where P_r is received signal power, while P_s is the sent signal power. Note that dB will be negative if we lost power, and positive if we gained power (and exactly the same if there was no change). When a signal is amplified, dB becomes positive.

12.2 Delay Distortion

Delay distortion occurs because the velocity of propagation of a signal through a guided medium varies with frequency. (different frequencies flow through the wire at different rates).

12.3 Noise

Noise is the major impairment on transmission. There are several types of noise:

12.3.1 Thermal Noise

Noise caused by temperature in electronic circuits. It is generally unavoidable. Sometimes referred to as ‘white noise’.

12.3.2 Intermodulation Noise

Medium carrying signals at several frequencies may combine to cause interference at other frequencies. For example, signal at f_1 and another signal at f_2 may produce some energy at frequency $f_1 + f_2$, which may interfere with some other signal at that frequency.

12.3.3 Crosstalk

Crosstalk happens when two signals cross paths. This is similar to hearing someone else’s conversation on your telephone/radio.

12.3.4 Impulse Noise

Spikes in the signal are called Impulse Noise. They’re not predictable, and cause huge amplitudes in the signal (these can be caused by faulty equipment, by some sudden event like lightning, or someone switching on/off some other equipment).

13 Hz vs. bps

13.1 Nyquist Bandwidth

Previous section should have illustrated something about the unequal nature of bandwidth and data rate. Usually they are related. More bandwidth usually implies higher data rate.

It turns out (due to Nyquist), we can get about 2bps for every 1Hz. Each wave has an upside (the 1bit), and a down side (the 0bit). So if a channel has say 8000Hz bandwidth, we can transfer bits at 16000bps. That is, assuming the channel is error free.

Obviously there are more things we can do—such as instead of just worrying about +1 and -1 voltages, we can also have different levels of voltages. Ie: increasing number of

different signal levels. The formula in such a case becomes:

$$C = 2B \log_2 M$$

where M is the number of distinct voltage levels, B is the bandwidth in Hz, and C is the bps of the signal.

Note that some books use the “Pulse Rate” instead of bandwidth. Pulse rate is just twice the bandwidth:

$$\text{BitRate} = \text{PulseRate} \log_2 L$$

where BitRate is bit rate in bps, PulseRate is number of pulses per second, and L is number of distinct voltage levels.

13.2 Shannon Capacity

Once noise gets involved, things change dramatically. Usually, increasing bandwidth, increases data rate, but since waves become more frequent, any slight noise will disturb them. That’s why we have signal-to-noise ratio (which is just a number). We can assign ‘decibels’ units to it via:

$$SNR_{dB} = 10 \log_{10} \frac{\text{signal power}}{\text{noise power}}$$

Claude Shannon setup a formula for the theoretical maximum data-rate that can be achieved via:

$$C = B \log_2(1 + SNR)$$

where C is data-rate in bps and B is bandwidth in Hz, and SNR is just $\frac{\text{signal power}}{\text{noise power}}$. So given a certain signal to noise ratio, on a channel of certain bandwidth, we can figure out what bps we can achieve. In practice, we get much lower rates—the formula gives us a theoretical maximum we can ever hope to achieve; it also doesn’t account for non-continuous (Impulse) noise—the noise is assumed to be ‘white noise’.

Note that the formula implies that we can increase data rate by either increasing bandwidth or increasing signal strength. In practice, increasing signal strength also increases Intermodulation Noise.

13.3 E_b/N_0

There is a similar expression to SNR , but relates signal energy *per bit* to noise power density per Hertz, E_b/N_0 .

The energy per bit, $E_b = ST_b$, where S is signal power, and T_b is the time required to send one bit. The data rate $R = 1/T_b$. In short:

$$\frac{E_b}{N_0} = \frac{S/R}{N_0}$$

this tells us that if we’re doubling the data rate, and want to maintain the same E_b/N_0 , then we also need to double the signal strength.

13.4 Intuition

Entropy is a measure of uncertainty. If an outcome is very uncertain, then entropy is high, etc. If things are easily predictable, then entropy is low. We can picture every ‘information source’ to have some measure of entropy.

For example, if we have an information source that *always* produces 1s, then that data source has entropy 0. We don’t need to transmit any bits to be able to reproduce an infinite sequence of 1s. Similarly, if we have an information source that produces seemingly random numbers between 0 and 1, then that information source will have entropy 1.

Entropy is a function of the form:

$$H = -K \sum_{i=1}^n p_i \log p_i$$

The constant K can often be ignored. For two possibilities, this function becomes:

$$H = -(p \log p + (1 - p) \log(1 - p))$$

Which is plotted in Figure 13. Notice that when probability is exactly at 0.5, we’re completely uncertain—which makes entropy high.

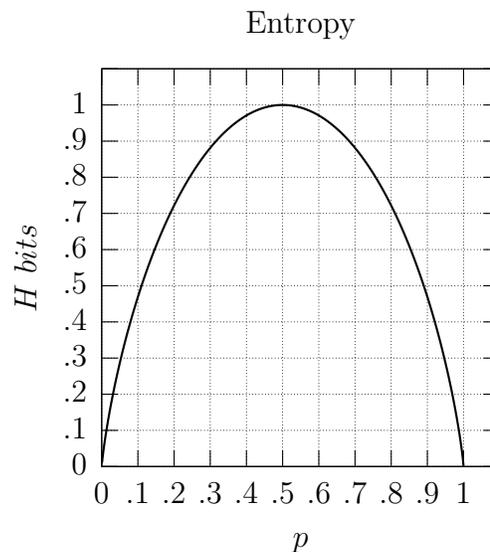


Figure 13: Entropy in the case of two possibilities with probabilities p and $(1 - p)$.

Now to the capacity. Entropy is essentially the measure of information we must transmit. Whatever channel we use, it *must* have higher capacity than the information source entropy.

13.4.1 Correction Data Channel

(This example is directly from Shannon’s Mathematical Theory of Communication.)

It is useful to picture a channel that contains ‘correction’ information. Picture a channel with rate of 1000 bits per second. There is a 1% probability of error (meaning, that approximately 1 in 100 transmitted bits will be flipped).

Now, consider what it would take to construct a channel to transmit just the correction information. What is the entropy of this correction information source? Using the ‘two possibilities’ graph, we can come up with

$$H_{\text{correction_channel}} = -(p \log p + (1 - p) \log(1 - p)) = -(0.99 \log 0.99 + 0.01 \log 0.01) = 0.081$$

That’s about 0.081 bits per *symbol*. If our rate is 1000 bits (symbols) per second, 81 of those bits must be used for this correction channel—leaving us with error free rate of $1000 - 81 = 919$ bits per second.

13.4.2 More Intuition

Let us now be silly and define the noise free channel capacity as:

$$C = B \log_2 M$$

but instead of M being ‘number of distinct voltage levels’, let us define it as ‘power’ (this isn’t exactly right—but it works for intuition).

$$H_{\text{received}} = B \log_2(M + N)$$

where B is bandwidth, M is signal ‘power’, and N is noise ‘power’.

The noise entropy is:

$$H_{\text{noise}} = B \log_2 N$$

Now, to get the signal entropy (ie: capacity of this channel), we do: $H_{\text{received}} - H_{\text{noise}}$, which gets us:

$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

which is the same as the ‘Shannon capacity’ presented above.⁴

14 Transmission Media

There are basically two forms of media... guided and unguided. Every time we have a wire, cable, etc., it is ‘guided’. Every time we just send out radio signals in all directions, then it is unguided. Generally, there are only a few transmission media that we have to be aware of: cable (sorta your cable TV wire), wire (sorta your telephone wire), fiber (glass/plastic wire,

⁴ $C = B \log_2(1 + SNR)$

used with lasers/leds), microwave (directed antennas), air (radio). There are also a bunch that we can consider like sonar, etc., but generally the above listed ones are what we have to work with.

Cable has very high bandwidth, as can be evidenced by it carrying all that video information for every channel. Telephone wire is cheap, but has a high bandwidth. The reason our telephones are so bad is because traditionally, the telephone company used to place many phone conversations on a single wire, allocating a chunk of bandwidth of each user. Fiber is possibly the winner in terms of bandwidth.

15 ISO's OSI Reference Model

Since we are getting to the point of transmitting bits over the wire, Now is a good place to mention some standards, and why they evolved.

As you can probably guess, anyone can lay wires anywhere, and as you've seen, anyone can begin to send digital signals to places without major technical issues. Ie: it's all just voltage in a wire (or light in the fiber).

Now, you can also devise a way to read those signals where they arrive. The problem is that if you want anyone else to be able to read/interpret them, then they also need to know what you're doing. Thus, we need a standard that people can use in communication.

The OSI reference model is sorta the ideal and failed model. It fails because it never got off the ground—designed by a committee; also, if it were to be implemented, it would be less efficient (but more convenient) than say TCP/IP (which we'll talk about a bit later in the course). Anyway, with all the short-comings, the OSI ref model is still a good place to look at the 'how it should be' of a protocol.

There are 7 layers:

1. Physical Link: this is the wire. This part would specify electric characteristics of a signal, etc.
2. Data Link: Responsible for establishing a link between 2 nodes, and sending data between them. This may possibly split things into distinct messages. Provides error detection.
3. Network: Responsible for routing information from one part of the network to another.
4. Transport: This layer identifies the source/destination end-points. It has no routing information.
5. Session: Sets up a reliable communication session.
6. Presentation: Handles things like data encryption, compression, etc.
7. Application/User: This is the user application using the channel.