An Overview of Satellite Communications: Topics in Satellite Television

Ferrel Atkins
Advised by Janice Hudgings

Thesis submitted to
Pomona College
in partial fulfillment for the award of the degree of
Bachelor of Arts
in Physics

2015-2016

Contents

Acknowledgements

This work would not have been possible without the help of my thesis advisor Janice Hudgings. I would also like to thank my college advisors Bill Swartz and Phil Choi for guiding me through college and my readers Scott Tan, Carson Witte, and Andy Wright.

Chapter 1: Introduction

1.1 Motivation

The communications industry today is worth over $5.6 trillion a year [1]. This industry involves the transfer of information, including anything from telephone services to wireless internet. One such service is the satellite communication industry. Of the $5.6 trillion worth in total communications, $195 billion is associated with the satellite industry [2]. Today satellites are used for the anything from commercial services such as television, to science experiments created by NASA. They provide the basis of modern day wireless communications. Through the use of electromagnetic waves, complex information can travel long distances without a network of wires. In my research, I explain the process of transmitting information through satellites. Within this subject, I focus on satellite television services, which make up a large part of the 40% of the industry set aside for commercial use [2]. I explore how electromagnetic waves hold such a large amount of information and how the receiver takes and converts the wave back into useable information. After exploring these topics, I comment on their combined advantages and disadvantages. I determine if the current system is optimized for the process of providing satellite television because of the haphazard manner in which frequencies were allotted.

This research was meant to aid my Harvey Mudd clinic team's project. DIRECTV has tasked us with improving their antenna installation process. We have created a web application designed to determine the feasibility of placing an antenna at a given location. Currently, DIRECTV must send someone to the customer in order to decide if an antenna with direct, unobstructed line of sight access to their satellites can be placed. These visits are often costly and sometimes unsuccessful due obstacles that prevent an antenna from being installed. With this web application, both the customer and DIRECTV will be able to determine remotely the feasibility of placing an antenna, removing the necessity of site visits.

From this project stems an interest in the sending and receiving of information via satellites. I have considered the current setup of information transfer and have researched the following questions. Is the frequency allotted for television broadcast ideal for its purpose? Do the algorithms that encode and decode the information improve the signal transfer? What significant affect does the receiver have on information transfer performance?

1.2 Background

DIRECTV provides television services wirelessly through satellites. The company owns five satellites in geosynchronous orbit at latitudes between 99º and 119º each of which provides a different channel package. In order for a customer to receive signal from these satellites, there must be an unobstructed path between the antenna and the satellite. Because there are many obstacles such as trees, other buildings, and other obstructions, finding a suitable location for the antenna is difficult. However, in my research, I hoped to have determined if the technology available mitigates the effects of noisy signals and scattering on electromagnetic waves. After research, I have determined that the current process is idea for the purpose of providing satellite television.

First, I explore frequency space. When an electromagnetic wave is scattered, the signal is weakened. This signal disruption can allow noise to dominate, creating significant errors in the data being transmitted. Noise makes the signal hard to detect, causing problems in the receiving process. In order to determine signal attenuation, I will explore different scattering affects and compare them to see which part of the electromagnetic spectrum would have the least signal loss when going through the atmosphere, weather, leaves, and other obstacles. This knowledge was meant to help the feasibility program by determining if a given obstacle would cause enough scattering to significantly attenuate the signal. If the specific broadcast frequency is not scattered, the antenna could still be installed. With the possible improvements that the decoding algorithms and the receiver can provide, a line of sight to the satellite might not be absolutely necessary.

Next, I explored the algorithms designed to convert a signal into useable information. These algorithms are designed to deal with noisy signals where the information might be changed during transmission. When they process the complex information, they are designed to pick out the actual signal. Because they are designed to deal with bad signals, they are capable of getting the information out of a weak signal if used properly. Although they can deal with bad signals, it takes a lot of extra data to ensure error detection and correction. The hope was that an obstacle would not completely destroy the signal so that the algorithms could recreate it from fragments. An antenna could therefore be installed in an unlikely place due to the signal correction power of the algorithms, improving the feasibility application.

Finally, I explored the antenna technology. This involves the parabolic dish designed to focus as much of the signal as possible, the feed horn designed to collect the entire reflected signal, and the low noise block converter designed to transform the electromagnetic waves into an electric signal. These are all designed to amplify the weak signal coming from the satellites. After travelling through the earth's atmosphere and possibly weather, the signal has been significantly weakened by scattering. The dish, in combination with other technologies, helps the antenna receive the strongest signal possible. Because obstacles weaken the signal, these technologies can possibly overcome signal strength issues, allowing for antenna placement in previously excluded areas.

Chapter 2: Topics in Frequency

2.1 Frequency and Satellite Assignment

Soon after it was discovered that electromagnetic waves could hold information, people took advantage of the ability to communicate wirelessly at the speed of light. Currently, these waves blanket the world, making every form of wireless communication possible. Both land based transmitters as well as space based transmitters constantly bombard people with information receivable with the proper equipment. Each receiver is designed to pick up a specific frequency range that the intended information is sent through. These frequency ranges each have advantages and disadvantages that determine their use. In order to regulate the soup of electromagnetic transmissions, the world created the International Telecommunications Union.

Because similar frequencies interfere, the type of information sent through each frequency range must be regulated. The International Telecommunications Union serves this purpose, designating specific frequency bands for specific purposes. The organization has broken the world into 5 regions, each of which has organized frequency spectrum use differently [3, 4].

In the region that includes the United States, waves with frequencies ranging from 500kHz to 900MHz are used for terrestrial communication. Signals ranging from 500kHz to 3MHz, known as medium frequency, propagate between the earth's surface and a part of the earth's atmosphere known as the ionosphere. This medium acts as a waveguide, allowing these signals to cover distances as far as several hundred miles as long as there is not weather or other attenuating conditions. Because these waves can travel large distances, these frequencies are used for maritime and AM radio broadcasting. The 3MHz to 30MHz range of the frequency spectrum is reflected by the ionosphere, allowing for long distance communication. These wavelengths are used for the purpose of international radio due to their ability to reflect around the globe. Once frequencies are above 30MHz, the signal passes through the ionosphere with little loss. The signal is therefore relegated to skywave propagation, meaning it can only travel shorter distances and satellite communication where line of sight access is needed. Frequencies below 900MHz, both very high and ultra-high frequencies, are used for the purpose of local communication. FM radio and local television providers use this range because different regions can use the same frequencies with minimal interference. The International Telecommunications

Union has assigned the higher frequencies, ranging from 2.5GHz to 22GHz, for satellite use [4, 5].

Many of these satellites orbit the earth at a height of 22,300 miles above sea level in what is called a geosynchronous orbit. In geosynchronous orbit, a satellite orbits the earth at the same speed as the earth's rotation, allowing them to remain stationary over the Earth's surface. They can provide a continuous stream of information to an area as large as 42.4% of the earth's surface [4]. Because of the large coverage potential, a large portion of the electromagnetic spectrum is set aside for their use. However, because these satellites must orbit over the equator, there is limited space for them. In order to prevent two satellites transmitting at the same frequency from interfering with each other, they must be separated by at least 2 degrees in the sky. This separation limits the number of geosynchronous satellites that can be put into orbit. The International Telecommunications Union therefore regulates geosynchronous satellite space. Countries must petition the organization for a specific latitude before they can launch a satellite [4, 6].

Satellites used to provide satellite television emit signals in either the K_u band (10 GHz to 17 GHz) and K_a band (18 GHz to 22 GHz) frequency ranges. In addition to transmitting through the atmosphere with minimal signal loss, these frequencies also have the advantage of requiring a smaller antenna. Because signals from different satellites are easily distinguished, they can transmit at a higher gain, meaning they send more powerful signals. Satellite television therefore requires smaller reflecting dishes than C-band transmission, which is transmitted at a weaker gain [4, 7]. For a parabolic dish, the maximum possible focused gain received goes as,

$$G_{max} = \varepsilon \frac{4\pi}{\lambda^2} A \tag{1.1}$$ [8]

Where ε is the efficiency of the dish, λ is the wavelength of the incident signal, and A is the aperture of the dish. Smaller wavelengths therefore require a much smaller dish in order to collect the same gain, making the K_u and K_a bands convenient for commercial use. Smaller antennas and reflecting dishes are easier to commercially create and mount, allowing for companies such as DIRECTV to provide satellite television service to many people [4, 5]. In order to provide this service, however, these companies must also determine methods of sending information via electromagnetic wave.

2.2 Electromagnetic Wave Modulation

There are several methods of modulating an electromagnetic wave to carry information. The first, known as AM modulation, involves changing the amplitude of a given wave to carry the information. The carrier wave's frequency remains constant but the different amplitude levels relay the information. Given the relation,

$$P \propto A^2 \tag{1.2}$$

Where P is power and A is wave amplitude, these modulations affect signal power. Because shorter wavelengths are more susceptible to attenuation due to scattering (see section 2.3), longer wavelengths of the radio spectrum use amplitude modulation [9].

Another method of wave modulation is frequency modulation. The carrier wave's frequency, as opposed to its amplitude, is modulated to carry information. The range of frequencies used for the final signal is known as the bandwidth. Receivers are designed for a specific bandwidth, limiting interference from other signals. The size of the bandwidth is determined by the Shannon-Hartley theorem, which is derived fully in section 3.4

$$C = Blog_2\left(1 + \frac{S}{N}\right) \qquad (1.3)[10]$$

Where C is the channel capacity in bits per second, B is the bandwidth in hertz, and S/N is the signal to noise ratio the receiver receives. Large bandwidths are therefore required to carry frequency modulated information. Although bandwidths of the same size can carry the same amount of information, providers do not limit the bandwidth to the minimum size needed [6, 10].

Due to the vast amount of data required for satellite television, a much larger than needed bandwidth is used. Video, several audio channels, and permission information are all packaged into a set of frequency channels within the designated frequency bands. In order to provide information fast enough for quality television, providers use wideband transmission, or wider than necessary bandwidths. As the Shannon-Hartley equation describes above, this allows for a much larger channel capacity and therefore the ability to send information much faster. When compared to the entire frequency spectrum, these larger than necessary bandwidths take up only a small fraction of the frequency spectrum allotted for satellite communication. In addition to a large capacity for many channels in the satellite radio frequencies, any frequency can also be reused, increasing the number of possible signals. Signals can share the same frequency without interference through the use of polarization. Electromagnetic waves can be polarized oppositely through a basis of linear directions or circular directions. The polarization of electromagnetic waves prevents signals of the same frequency from interfering, increasing the usefulness of wireless communication farther [6, 10, 11].

2.3 Scattering

A limiting factor in the use of electromagnetic waves to carry information is scattering. When a wave is scattered it losses power, attenuating the signal. As a wave travels through a medium, it interacts with the particles that make up that medium. This interaction, known as electromagnetic scattering, causes the incident wave to be reemitted in many directions weakening the signal. The type of scattering a wave experiences is based on the relative size of the scatterer to the signal wavelength. When electromagnetic waves propagate through the atmosphere, they undergo a type of scattering known as Raleigh scattering.

Raleigh scattering occurs when the incident wave hits particles much smaller than the wavelength. For satellite television wavelengths, which are on the order of centimeters, molecules in the atmosphere such as water droplets cause scattering. Modeling these scatterers as a dielectric radiating sphere, the Raleigh equation becomes the ratio of the scattered power to the incident power density. It states,

$$\sigma_{total} = \frac{8}{3}\left(\frac{\epsilon-\epsilon_0}{\epsilon+2\epsilon_0}\right)^2 (k^4a^4)(\pi a^2) \qquad (1.4)[12]$$

Where ϵ is the dielectric constant, a is the radius of the sphere, and k is the wavenumber meaning,

$$k = \frac{2\pi}{\lambda}. \qquad (1.5)$$

The total power scattered to incident power density ratio therefore goes as radius to the sixth over wavelength to the fourth. The total power scattered over the incident power density therefore goes as,

$$\sigma_{total} \propto \frac{a^6}{\lambda^4}. \qquad (1.7)$$

Taking an example from Applied Electromagnetism by Laing and Jin, it is possible to determine the total scattering cross section of a raindrop for a wavelength on the order of satellite television. Because a typical raindrop has a cross section much smaller than the wavelength, equation 1.4 can be used. If a raindrop is assumed to be 3 mm in diameter, to have a dielectric constant of $61\epsilon_0$, and have an incident wave of 10 GHz, the total scattering cross section is,

$$\frac{\sigma_{total}}{\pi a^2} = \frac{8}{3}\left(\frac{60\epsilon_0}{63\epsilon_0}\right)^2 \left(\frac{2\pi}{0.030}^4 * 0.0015^4\right) = 0.024, \qquad [12]$$

Meaning a raindrop scatters 0.024 of the power in the incident wave [12]. These small particles do not cause a significant amount of scattering on their own therefore allowing the signal to pass through the atmosphere with limited power loss. However, if the wave is incident on enough of these particles, it will be significantly weaker. In an effect known as rain attenuation, satellite television signal can be scattered enough in rainy weather to cause service issues. The many rain droplets can even add in effect to weaken the signal to the point of signal loss [4, 13].

The scattering effect of ice particles on an incident wave can also be modeled using equation 1.4. Using the same frequency wave, and the same sized ice particle with a dielectric constant of $3.2\epsilon_0$the scattering cross section is,

$$\frac{\sigma_{total}}{\pi a^2} = \frac{8}{3}\left(\frac{2.2\epsilon_0}{5.3\epsilon_0}\right)^2 \left(\frac{2\pi}{0.030}\right)^4 (0.0015)^4 = 0.0045. \qquad [12]$$

This result shows that ice scatters less than rain [12]. However, with the additive effect of enough particles, significant signal attenuation and possible signal loss can occur [4, 13].

Because scattering affects the power, and therefore the amplitude of electromagnetic waves, non-terrestrial based communication all use frequency modulated signals. Scattering can significantly affect amplitude modulation, making weather and other particles affect more than the signal strength but the information itself. For these reasons, longer wavelengths, which are scattered less, are amplitude modulated [5].

Although small particles weaken gigahertz frequency signals, other objects that are sized on the order of the wavelength often completely attenuate the signal. Scattering by objects on the order of the wavelength is known as Mie scattering [12]. The most common Mie scatterer for satellite television is leaves. Because the Mie solutions are computationally complex, they will not be covered here, however, studies on object’s scattering coefficients have occurred [15].

According to Helhel et. al., leaves scatter based on their water content. Similar to Raleigh scattering, this proves that the dielectric constant of the scatterer plays a large role in signal attenuation. Their studies, which simulate a leaf as a resistive strip, suggest that the moisture content of a leaf can cause as large as a 72% increase in signal attenuation, rendering the signal basically useless [14]. At this level of power loss, the antenna cannot use the signal. This signal loss explains the necessity of a clear line of sight between the antenna and the satellite.

2.4 Conclusion

Electromagnetic waves experience many forms of scattering, which cause power attenuation. Different conditions are problematic for different wavelength waves, exemplifying the importance of using each part of the frequency spectrum to its full advantage. The longer wavelengths, which experience little signal loss due to scattering, are used for longer distance communications, while the shorter wavelengths are used for satellites because they go through the atmosphere with relative ease. Satellite television frequencies are therefore allotted for the proper use due to their ability to travel through the atmosphere despite the limiting factors of signal attenuation due to scattering.

Chapter 3: Topics in Information Theory

3.1 Introduction

Information can be sent in two forms, digital and analog. In analog transmission, the whole electromagnetic wave that represents the information is processed both in transmission and receiving. This method uses unnecessarily large amounts of data, rendering it obsolete for the purpose of satellite television. Currently, satellite television uses digital communication in order to transmit information. Digital information processing involves transforming the analog signal into a set of discrete values, which are then sent as modulations through the carrier wave. In order to determine these values, an electrical signal is sampled fast enough to capture the necessary information. These discrete values reduce the amount of information sent over a given frequency channel.

The sampled discrete values are transformed into a binary system of information storage. Values of 0 or 1, known as bits, are strung together in order to represent information in the simplest form possible. These strings of 0s and 1s have many advantages, including a technique known as compression. Strings of bits can be packaged into smaller strings, allowing for faster information distribution.

The process of transforming these values into a waveform also simplifies the process. For satellite television, a form of frequency modulation known as frequency shift keying is used. The carrier wave of a specific frequency is modified by a square wave resembling either a value of 0 or 1 (see figure 2). The carrier frequency is either increased or decreased to represent a discrete sample value. As these waves travel, they are subject to various forms of disturbances. These changes to the wave changes the discrete values they represent and therefore the information transmitted. Another advantage to digital information representation is that these changes, error in bit strings, can be easily found and corrected. Through techniques created by Richard Hamming, methods of correcting these errors have been found and implemented to create a robust form of information transfer.

3.2 Analog to Digital Conversion

3.2.1 Analog Sampling

In digital communications, a continuous signal must be transformed into discrete values that completely represent the information contained within the signal. This process, known as sampling, converts analog waveforms into a set of discrete values containing the key information being transmitted. Most discrete values take on the form of a string of 0s and 1s in a binary system. When sampling, the longer the bit string representing a sampled amplitude, the better the resolution of the digital signal in approximating the original signal (see figure 1). In addition to the amplitude resolution of the discrete samples, which carries the actual information, the digital signal must also be sampled fast enough to be able to accurately represent and then recreate the original signal [16].

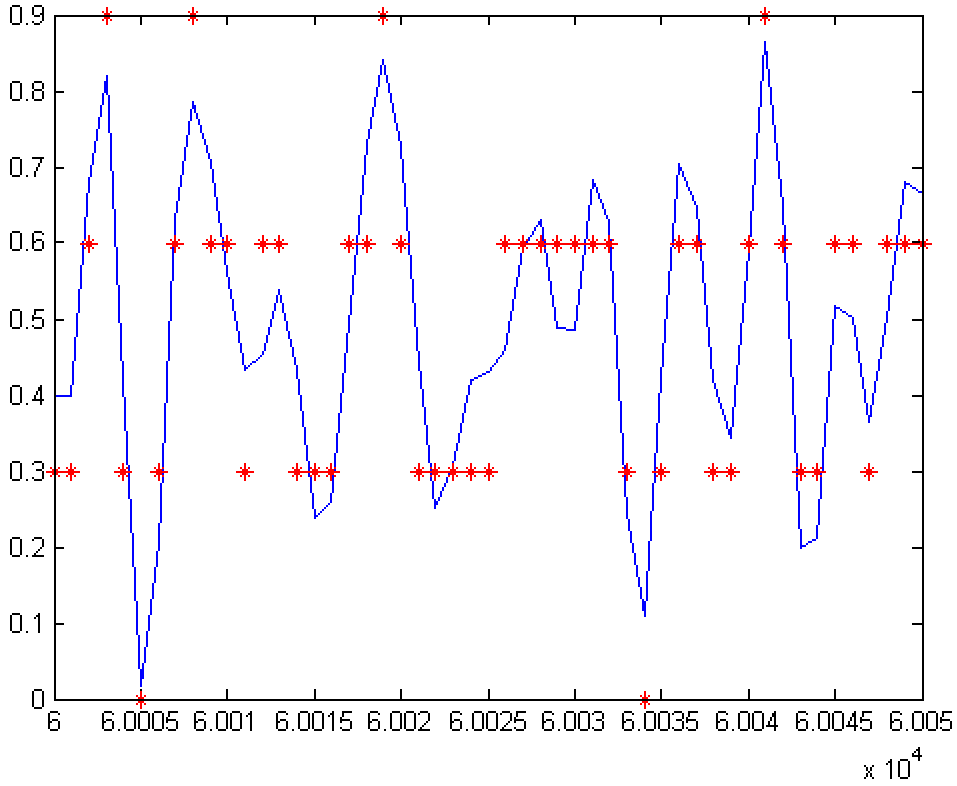

Figure 1: An example of a sampled signal where the x-axis is time and the y-axis is amplitude. The red points represent the sampled values of the blue waveform at 2 bit resolution. Points at 0

would be represented by 00, at .3 would be 01, at .6 would be 10, and at .9 would be 11. Increasing the bit resolution to 3 would increase the number of possible discrete values to 8.

3.2.2 Frequency Shift Keying

Once a signal has been converted into a bit string, the information must still be sent via electromagnetic wave. In a process known as frequency key shifting, these waves are modified to hold either a 1 or a 0 as figure 2 shows. Within a given bandwidth, the carrier wave's frequency will be adjusted to represent a bit value, forming a simple waveform information storage system. Because scattering affects the amplitude of a signal, it affects the aspect of the wave that holds information in an amplitude modulated analog signal. For a frequency based signal, however, scattering causes little information distortion and only affects the actual strength of the received signal [17].

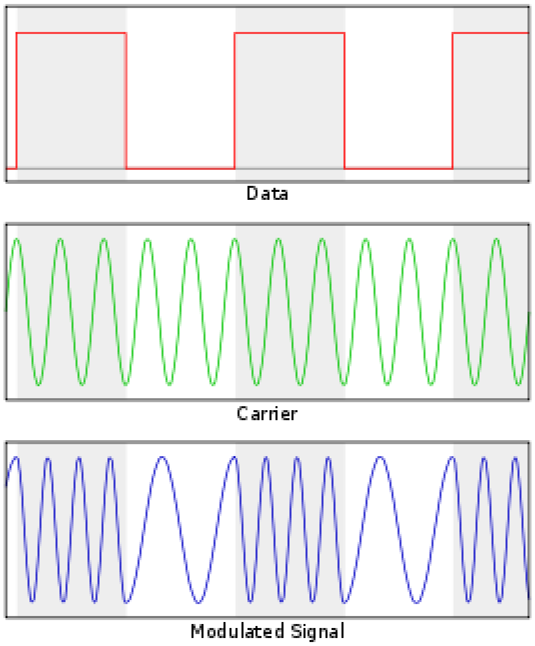

Figure 2: An example of the simplest form of frequency shift keying. A binary value of 1 is represented by a higher frequency than that of a binary value of 0. This figure shows an extreme version of frequency differences. A much smaller difference is used in practice [17].

3.2.3 The Nyquist Sampling Theorem

In order for the receiver to obtain the bit values represented by the incoming wave's frequency, the incoming signal must be sampled fast enough. If the signal is sampled too slowly, it would be incorrectly reconstructed. Figure 3 displays an example where the signal is not sampled fast enough and therefore reconstructed incorrectly. This reconstruction would produce a completely different signal from what is meant to be obtained, rendering the measured information useless [16].

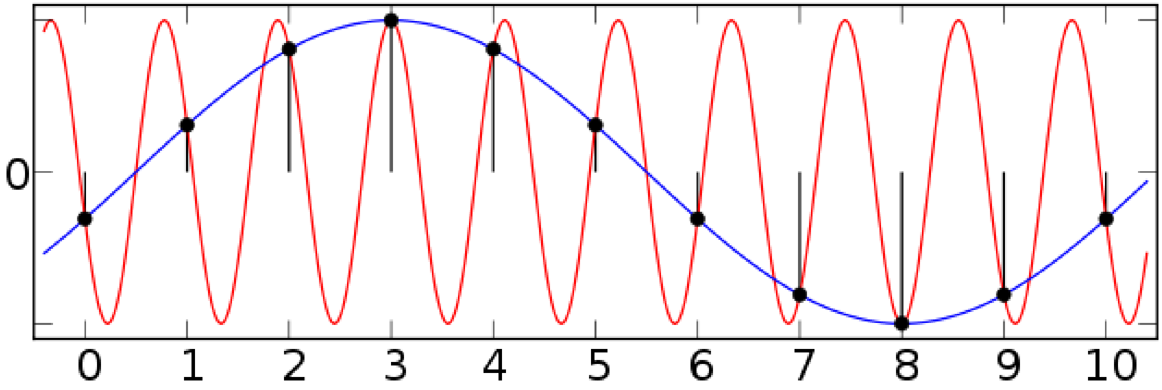

Figure 3: An example of aliasing where the waveform is sampled too infrequently. The original signal is red and the blue wave is the reconstructed signal from the black discrete points sampled [18].

The Nyquist Sampling Theorem gives the sample rate necessary to reconstruct the correct signal. In the time domain (amplitude vs time), this theorem states that in order to be able to reconstruct the signal, at least half the period of the wave must be sampled. Intuitively, it states that ignoring the phase, a sinusoid must be sampled at least at each amplitude maximum in order to reconstruct the signal. In the frequency domain (amplitude vs frequency), the theorem states that in order to be able to reconstruct the signal, it must be sampled at least twice the incoming signal frequency (see figure 4). Because period and frequency are inversely related, the theorem is the same for both domains [10, 16].

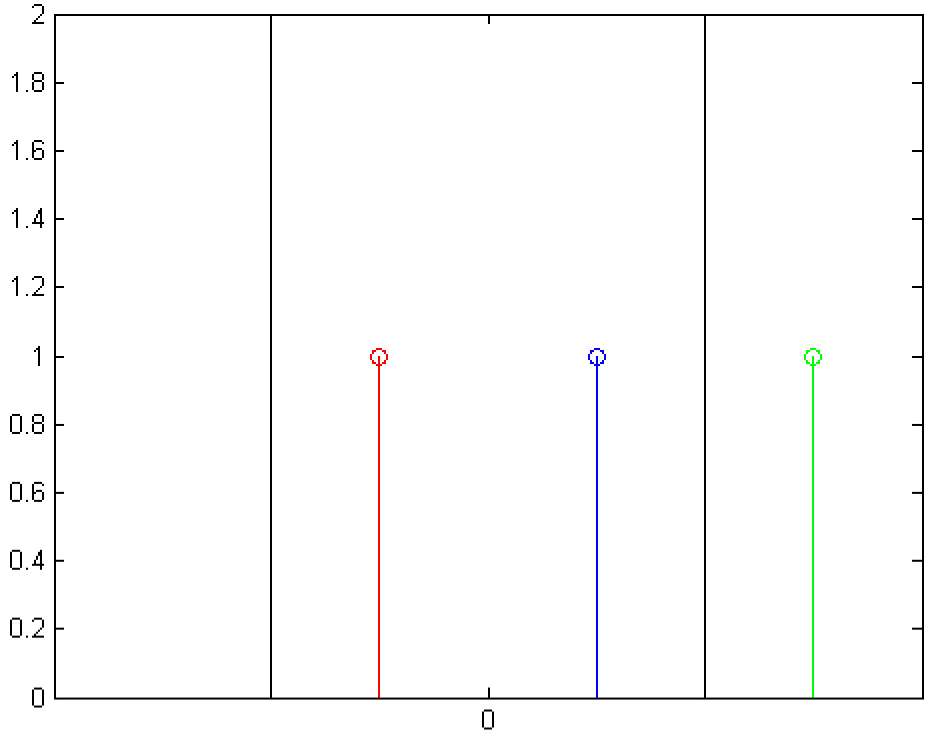

Figure 4: This plot shows sampling in the frequency domain where the x-axis is the frequency (Hz) and the y-axis is the signal amplitude. The black lines represent the bounds for frequencies will appear in the sampled signal. If the green signal were the original signal and the sampling rate was less than 2 times the signal frequency then it would appear as the red frequency (a completely different frequency) in the plot above. If the signal fell on one of the black lines then the signal would appear at both black lines when sampled, creating two matching signals representing the proper signal. If the blue signal were the original signal and the signal rate was high enough, then the red signal would also appear. This signal, however, is a replica of the blue signal as it is its negative frequency wave and therefore does not affect the construction of the intended signal.

3.3 Compression techniques

Although converting an analog signal into a digital one simplifies the information transmission process, it is often still beneficial to farther shrink the signal into a more manageable amount of information. In a process known as compression, digital information is father processed to reduce the amount of information required to send. For satellite television, compression can reduce the required transmission speed from rates as high as 200 Mb/s to 2 Mb/s [4]. Because a much smaller channel capacity is needed, the providers use much less bandwidth per channel, increasing the number of programming options available to customers.

The current method of compressing data for television services is known as MPEG-2. Through several tricks, television service providers can drastically reduce the amount of data required to send a channel to their customer. First, the signal goes through preprocessing. In this step, unimportant visual information is removed from the signal. Because people cannot perceive this information, it is removed before the encoding process. Next, the signal experiences temporal prediction and motion compensation. When coding information to be sent, this process gets rid of the unnecessary frame by frame detail. Because one frame usually follows from the previous, these processes encode the differences between frames. Instead of coding frame by frame information, only the moving pieces of information between frames are coded. Lastly, the signal goes through a process known as quantization coding. Here, pieces of visual information that are basically indistinguishable by the human eye are combined to approximate the original signal. In this step, colors that the eye is less sensitive to are combined in order to reduce data without limiting the visual experience. Through these processes, satellite television providers require a smaller channel capacity, making it possible to provide the most information possible [4].

3.4 Shannon's Proof of Channel Capacity

In order to use a channel to its full potential, its capacity given a noise level must be determined. When a signal is received, most added signal originates from thermal noise. Hot instruments add power to the signal and can therefore affect its amplitude and frequency. If a given signal has N average noise power, then the standard deviation of noise is,

$$\sigma = \sqrt{N}. \tag{3.1}$$

This noise adds to the power of the original signal for a total received signal power of $P + N$ at many frequencies [10].

Noise interference therefore makes distinguishing a specific frequency difficult. Since each individual frequency will be distributed normally with standard deviation σ due to noise, the specific frequency will map from a range of possible values. To combat this noise, the range of frequency values would match to a single discrete value based on the expected deviation. Through frequency shift keying these frequencies would relate to a specific amplitude of a sample from the analog waveform. Since the power of a signal is related to its amplitude as,

$$A \propto \sqrt{P}, \qquad (3.2)$$

The number of acceptable amplitudes such that the noise distribution of frequencies representing these amplitudes are distinguishable is,

$$K\frac{A_{max}}{\sigma_n} = K\sqrt{\frac{P+N}{N}} \qquad (3.3)[10]$$

Where K is an arbitrary value greater than 1 that is closer to 1 as errors are not allowed [10].

Since a given signal of bandwidth B must be sampled at least $2TB$ different times in a time T, the number of different possible amplitudes and therefore information holding frequencies must also be $2TB$ (given Shannon's derivation of the sampling theorem in Communication in the Presence of Noise). Therefore the number of distinct signals (M) that can be used for a given bandwidth is,

$$M = \left[K\sqrt{\frac{P+N}{N}}\right]^{2TB} \qquad (3.4)[10]$$

Given that $\log_2 M$ bits can be sent in time T, the channel capacity in bits per second is therefore,

$$C = \lim_{T\to\infty}\frac{\log_2 M}{T} = B\log_2\left(K^2\sqrt{\frac{P+N}{N}}\right) \qquad (3.5)[10]$$

As an example, assume that there are two possible sample amplitudes represented by a 1 or a 0. Given there are no errors (K = 1), $\sigma = 1$, $\sqrt{\frac{P+N}{N}} = 2$ meaning there are two bits, and $2TB = 3$ meaning 3 samples are taken, then the possible values (distinct signals) are:

000
001
010
100
011
110
101

Equation 3.4 shows that the distinct number of signals is,

$$M = 2^3 = 8$$

Therefore the M possible combinations represent the total number of signal values for a given number of amplitudes sampled a certain number of times.

Because these amplitudes convert into frequencies in frequency shift keying, a specific type of frequency pulse must be used. Shannon argues that the type of pulse required is,

$$\frac{\sin 2\pi Bt}{2\pi Bt} \qquad (3.6)[10]$$

This pulse has a specific maximum, tapers to zero as $t \to \infty$, and has a value of 0 at other sample points. In order to convert different amplitudes into frequencies, these signals are therefore overlapped so that the maximums line up with the zero points of the other signals as figure 5 shows [10].

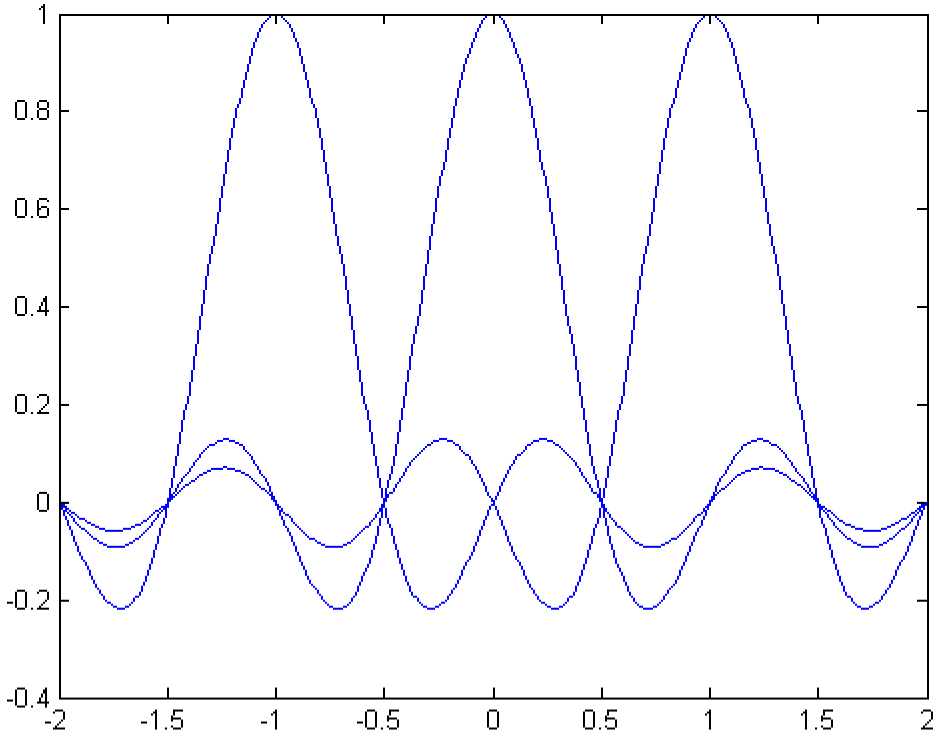

Figure 5: An example of 3 pulses where the x-axis is time (seconds) and the y-axis is amplitude. Each pulse has a maximum where each other pulse is zero and goes to zero away from the maximum.

3.5 Hamming codes

As information is sent between satellites and receivers, the carrier waves are subject to attenuation and noise. These disturbances in the wave can cause a change in the signal or even loss of the signal. Bits can change value due to noise or the addition of other signals or lost in transmission. In order to make digital data transmission productive, codes were created to detect and correct these errors. In 1950, Richard Hamming created a system of bit error detection and correction, which was published in The Bell System Technical Journal [19]. Although better methods are now used, these codes provide a basis for correcting bit errors.

For the next sections, n refers to the total number of bits used in each code block. These n binary digits are broken into two groups. The symbol m refers to the number of bits used for the purpose of information storage while $k=n-m$ refers to the number of bits used for the purpose of error detection and correction. Because some of the information sent is only used for the purpose of error detection and correction, code efficiency is also important. A value known as redundancy,

$$R = \frac{n}{m}$$

Measures this efficiency and determines how effectively the channel capacity is being used. As this value increases, more of the bits sent are used for error detection and correction and less for the purpose of the information. The amount of information sent is therefore less than the channel capacity of the range of frequencies used [19].

3.5.1 Error Detection

In order to solve problems associated with sending bits over large distances where the electromagnetic wave could change, errors in the strings of bits must first be detected. The simplest method of determining a single error in a string is to use a parity check. The last value of the n digits is given a value of 0 or 1 not associated with the actual information, creating this simple form of error correction. The single value creates a string which includes an even number of 1s, allowing for the receiver to easily determine an error. If there is an odd number of 1s within the n bits of the received signal, there is an error in transmission. For example, if a string of 6 bits starts with the first 5 of,

$$11001$$

Then the last digit, used for the parity check, would be assigned the value of 1 [19].

Although this method of error detection is simple, it behaves poorly for large values of n. As n increases, the likelihood of two errors, which would appear to something detecting a parity error as a correct signal, increases. According to Hamming, “a double error has probability 1/2e = 0.1839,” [19] when the probability of a single error is small and n is large. Therefore parity checks are often spaced at equal intervals within large strings of bits [19].

3.5.2 Error Correction

In order to properly detect and then correct an error in a set of bits, the number of bits used for parity checks, k, and the number of bits used for information storage, m, must be determined. In a string of n=k+m bits, k parity checks must take place. If a single check passes then a separate string of bits k digits long is assigned a 0, otherwise this string is assigned a 1. This group of bits is known as the checking number. For each subsequent parity check the value determined is added from right to left to give a final checking number which returns the position of the error in bit form. From Hamming, given $n=m+k$, the inequality,

$$2^m \leq \frac{2^n}{n+1}$$

Determines the maximum number of digits that can be used to hold information from an n-digit string while still allowing for error correction. For a given m value for information usage this gives,

$$(n, k) = (2^m - 1, 2^m - 1 - m)$$

The next step after determining the number of bits needed for error correction is to decide where to place the information in the string of bits, such that the checking number works properly. Parity checks based on each binary number can be performed such that the digit in the spot of that number check is a one causing it to fail. For example, for the first binary check we check spots 1, 3, 5, 7, 9… because their binary representations have a 1 in the far right spot.

1 = 01

3 = 11

5 = 101

7 = 111

9 = 1001 …

For the second parity check, the second digit from the right must include a 1, giving spots 2, 3, 6, 7, 10, 11…. Table 1 displays the checking positions based on the check number [19].

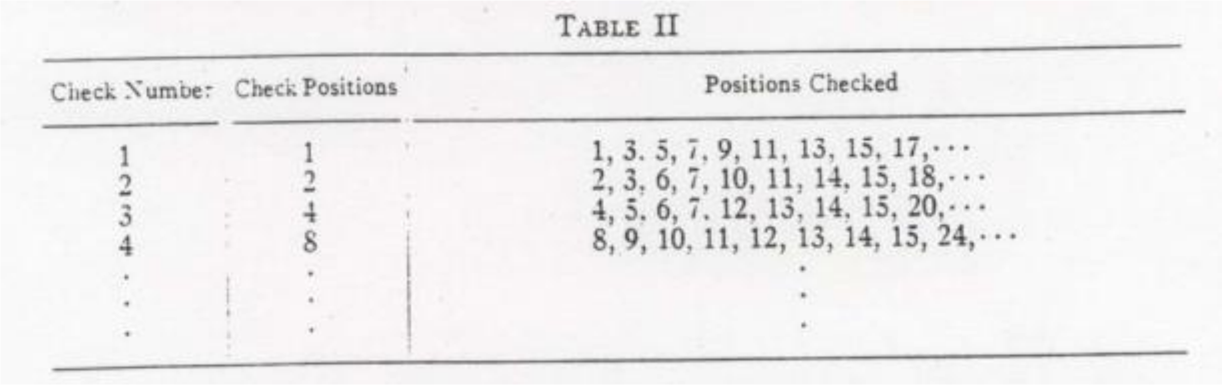

TABLE II

Check Number	Check Positions	Positions Checked
1	1	1, 3, 5, 7, 9, 11, 13, 15, 17,· · ·
2	2	2, 3, 6, 7, 10, 11, 14, 15, 18,· · ·
3	4	4, 5, 6, 7, 12, 13, 14, 15, 20,· · ·
4	8	8, 9, 10, 11, 12, 13, 14, 15, 24,· · ·
.	.	.
.	.	.
.	.	.

Table 1: This table shows the positions included in a given parity check [19]

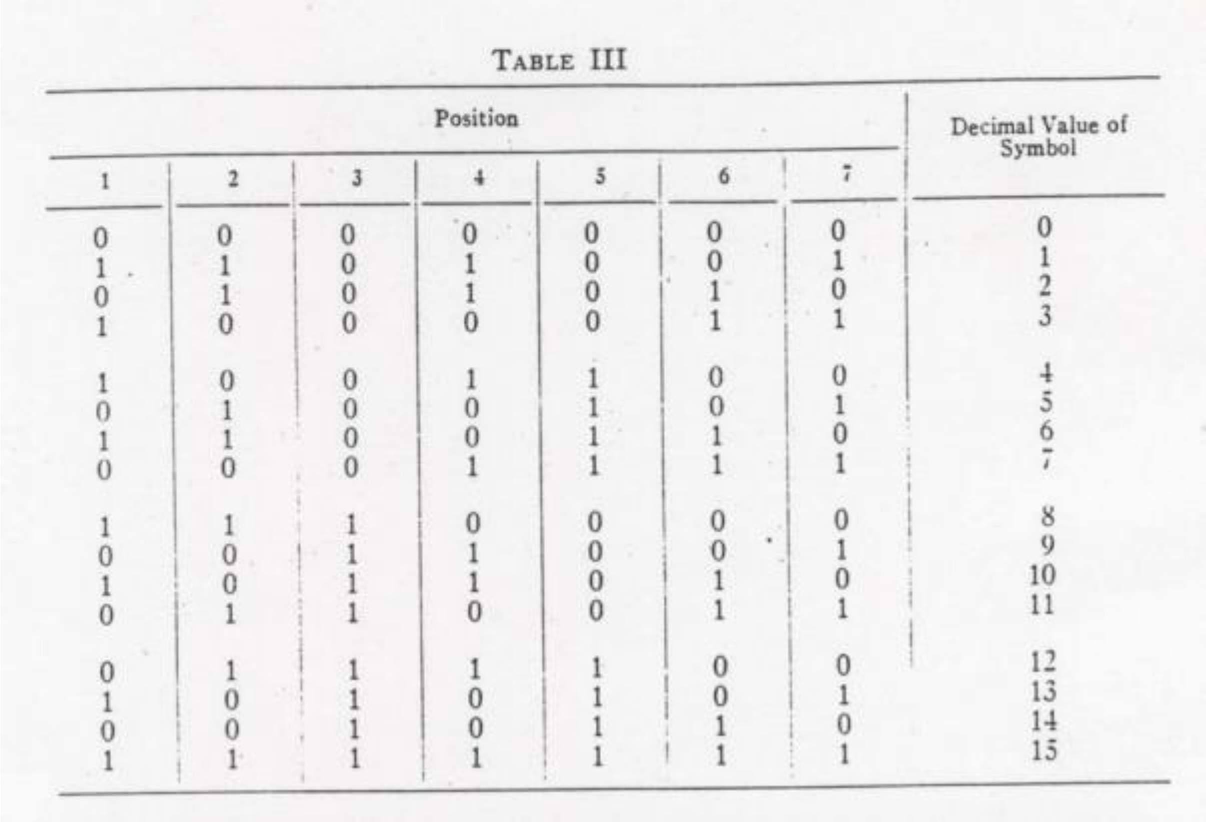

TABLE III

Position							Decimal Value of Symbol
1	2	3	4	5	6	7	
0	0	0	0	0	0	0	0
1	1	0	1	0	0	1	1
0	1	0	1	0	1	0	2
1	0	0	0	0	1	1	3
1	0	0	1	1	0	0	4
0	1	0	0	1	0	1	5
1	1	0	0	1	1	0	6
0	0	0	1	1	1	1	7
1	1	1	0	0	0	0	8
0	0	1	1	0	0	1	9
1	0	1	1	0	1	0	10
0	1	1	0	0	1	1	11
0	1	1	1	1	0	0	12
1	0	1	0	1	0	1	13
0	0	1	0	1	1	0	14
1	1	1	1	1	1	1	15

Table 2: This table shows the bit values of a 7 bit code such that there are 3 bits for error correction and 4 for information storage [19].

Hamming gives an example of how these codes work in his paper. If there are 7 positions for bits then parity checks one, two, and three can be applied, meaning 3 of the bits are for error correction. The spot each of these parity checks begins are used for the error correction positions while the others can be used for information. Therefore positions 1, 2, and 4 are for error correction. Table 2 displays these values for the different information showing that a 7 bit code can hold 16 pieces of information. In Hamming's example, the fifth position of the bit representation of 12 (see table 2) is changed giving the new code,

0111000.

The three parity checks from 1 to 3 then give a checking number built from right to left of

101

Which shows that the fifth position has the error to be corrected [19].

3.6 Conclusion

Digital communications provides a robust system to transfer information. Because analog signals are cumbersome to transmit as they contain a large amount of unnecessary information, a system of information transfer that greatly lowers channel capacity costs is beneficial. Through a good approximation, digital signals can provide all of the necessary information, while limiting the amount of signal space needed. Unlike analog signals where frequency and amplitude hold the information, these digital signals can be corrected if errors arise from transmission. The use of digital information communication instead of analog transmission has made digital communications the current standard of information transfer.

Chapter 4: Topics in Antenna Hardware

4.1 Introduction

The hardware that makes up the receiver system only ensures that the received signal remains intact. Antennas of different dimensions are used to transmit and receive electromagnetic waves of different wavelengths. Through the resonance or standing waves, wires at proper dimensions when compared to the desired signal wavelength can both send and receive electromagnetic waves. Electromagnetic waves are created by the propagation of electric field changes from an antenna. The direction these receivers point affects the signal as added noise. In order to minimize noise, receivers are pointed at colder temperatures. Once pointed properly, the low noise block downconverter converts the signal from radio frequency waves to electric L-band frequency signals that can be sent through normal wires (see figure 5). Each piece operates with the purpose of minimizing signal interference so the signal is intact when it finally reaches the television.

4.2 Resonant Antennas

In order to send and receive signals, television providers must use resonant antennas. These antennas use either one or two prongs in order to create or receive electromagnetic waves. The two pronged antennas, known as dipole antennas, are used in satellite television to create the signal to be sent to the satellite. These signals are focused by a parabolic dish that directs the signal to the satellite. In this process, the dish must produce a focused enough beam that the signal received by the satellite is strong enough to redirect back to Earth. Once the signal beam is transmitted from the satellite, single pronged antennas, known as whip antennas, receive the signal. Because weaker signals are simply electromagnetic waves with smaller electric field amplitudes, these antennas work better with length. Length, however, is not the only factor. In order to receive a signal of a specific frequency, the antenna must be the proper dimensions compared to the signal wavelength. This dimension effectiveness relation minimizes interference due to different frequencies for antennas [20].

4.2.1 The Dipole Antenna

Service providers use dipole antennas to create the electromagnetic waves that contain the information to send to satellites. These antennas have two prongs, which point opposite each other. One end of these prongs is free while the other end is connected to the transistor, which deals with the electronics required to make an antenna work effectively. In order to create an electromagnetic wave, a voltage is induced on the free end of each wire. Because the wires act like a conductor, a voltage can be induced across each wire, creating either a negative or positively charged free end. The transistor charges both wire ends such that they always have equal and opposite voltages. This voltage difference produces an electric field pointing from the end of the negatively charged wire to the end of the positively charged wire (see figure 1). Stronger voltage differences therefore create stronger electric fields [21].

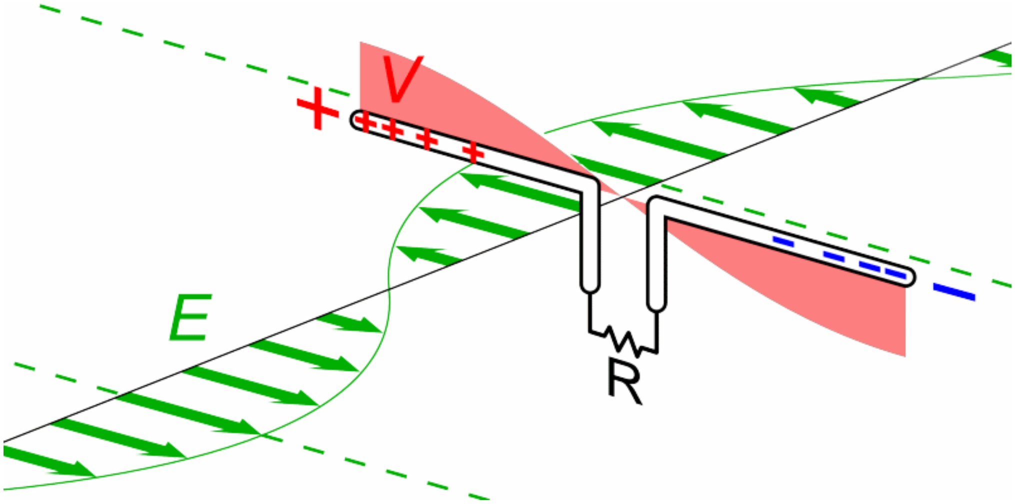

Figure 1: An example of a dipole antenna. One end is positively charged while the other is negatively charged. This apparatus creates an electric field with amplitude shown in the figure. The electromagnetic wave travels perpendicular to the electric field vectors outward from the antenna in both directions [21].

The dipole antenna uses changes in voltage differences to create an electromagnetic field. In order to change the amplitude of the electric field, the transistor must adjust the current creating the voltage differences across the two wires. Because these changes in voltage involve moving electrons, the antenna also creates a magnetic field that changes as the voltage changes. As these voltages are adjusted, electromagnetic waves are formed. Their waveform is based on the speed and method that the voltages are adjusted. The change in voltages propagates the electromagnetic wave outward from the antenna with a geometry shown in figure 2 [21].

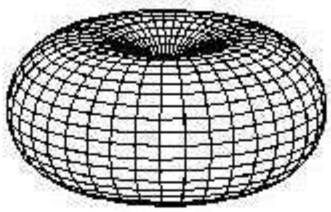

Figure 2: The three dimensional radiation pattern of a dipole antenna. Here the antenna is pointed along the z-axis (up) and the signal radiates outward as shown above. As the radiation distance increases, the signal gets significantly weaker [22].

Because antenna electromagnetic wave radiation is semi-isotropic, the gets significantly weaker the farther away it travels. The maximum power of an electromagnetic wave is,

$$P = \epsilon_0 E_0^2 c \qquad (4.1)[23]$$

Where $\epsilon_0$is the vacuum permittivity, E_0 is the maximum strength of the electric field, and c is the speed of light. Therefore, since electric field amplitude goes as,

$$E = -\frac{dV}{ds} \qquad (4.2)$$

Where s is the distance the voltages are separated, the power of an electromagnetic waves is directly related to the voltage difference across the antenna. If the signal were simply radiated, then distances on the order of geosynchronous satellites would require large voltage differences across these small antennas. Due to air conductivity and therefore sparking, these voltages are infeasible since the length of the wires must be on the order of the intended signal wavelength. Instead of creating incredibly strong electromagnetic waves, service providers combat this signal loss through the use of reflecting dishes. If the antenna radiates at the focus of the parabolic dish, the signal can be focused into a beam directed to the satellite intended to receive the signal. Therefore the voltages required to use the dipole antenna to create an electromagnetic signal are reasonable [21].

4.2.2 The Whip Antenna

Because a dipole antenna is too large to reasonably put inside a waveguide, only a single wire antenna, known as a whip antenna, receives the incoming signal. These monopole antennas use wave resonance, similar to the dipole electromagnetic wave creation process, in order to receive the signal. As the incoming wave passes the whip antenna, its electric field induces a voltage across the antenna. The electrons within the wire move based on the strength of the electric field to create a voltage difference between the free end of the antenna and ground. These voltages vary as the frequency of the wave, allowing for the antenna to turn the electromagnetic signal into an electrical signal (see figure 3). Because the strongest voltage difference would be induced across half of the wavelength of an incoming wave, these antennas work best when they are on the order of the signal wavelength. In order to perceive the strongest signal possible, these antennas must also be aligned with the polarization of the electric field of the incoming electromagnetic wave [24].

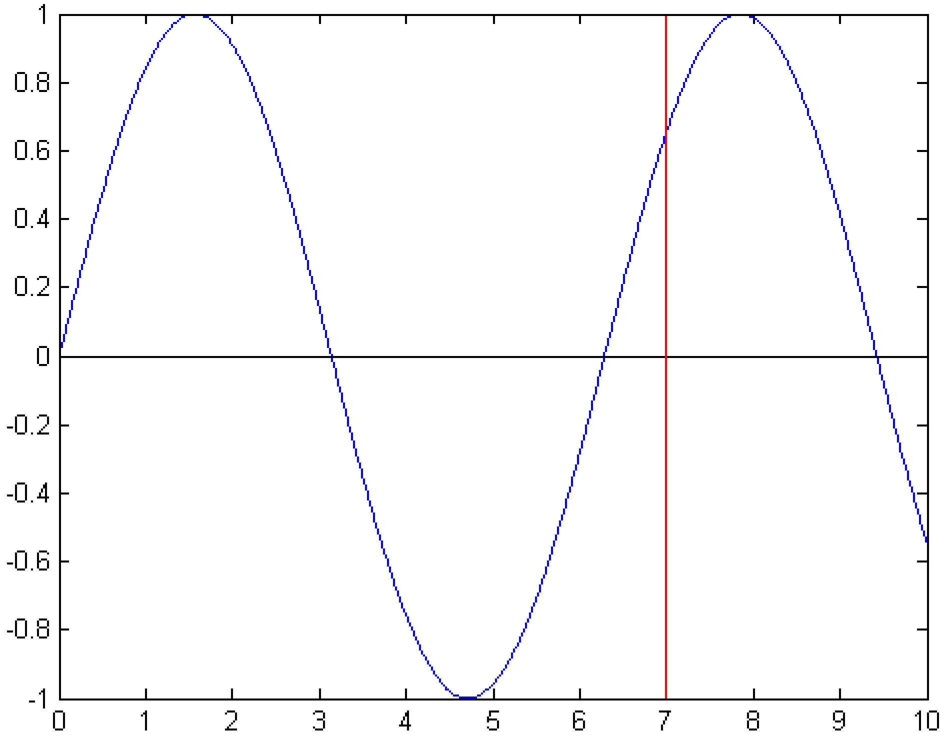

Figure 3: An example waveform received by a whip antenna. The red line representing the location of a whip antenna to receive where the x-axis represents distance and the y-axis

represents electric field amplitude. If the wave propagates to the right with time, the voltage induced across the antenna would change. Currently, the top end of the antenna would be positively charged when compared to ground. The antenna could be as little as 1/10th the amplitude and still receive the signal.

4.3 Noise

In order to collect the best signal, the receivers limit the amount of noise. This noise is the addition of frequencies not related to the original signal. When the noise level is low enough compared to the signal, these added frequencies do not affect signal detection (see figure 4). However, if noise levels rise to the same order as the signal, then errors in decoding the signal can arise. Frequencies signifying a completely different piece of information can take precedent over the original value when their amplitudes are within an order of magnitude of the signal intended signal frequency, causing errors or possibly even the loss of the entire signal. Noise originates from many different sources, each of which produces a different interfering spectrum. Often, for satellite communications, the greatest noise source originates from the operating temperature of the system and the temperature of the surrounding medium [4, 25].

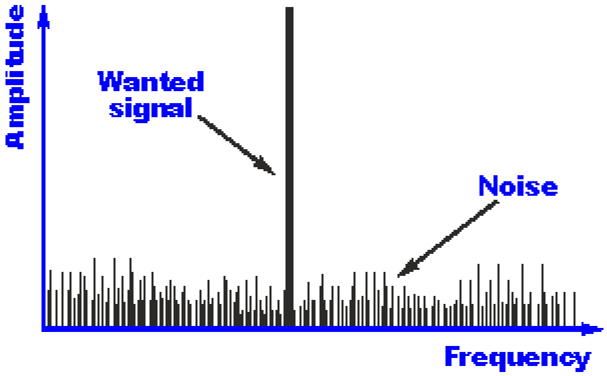

Figure 4: An example of noise added to a signal. This is weak noise compared to the signal as the signal has a much higher comparative amplitude. However, if the noise were the same power as the wanted signal then it would be likely to cause errors in the recreated signal [25].

Thermal noise originates from the temperature of the receiver system. As an object or medium heats up, the charged particles (such as electrons) within gain more energy. These charge carriers move constantly changing energy levels producing electromagnetic radiation across the frequency spectrum. With the increase of energy, this movement increases, causing stronger radiation across the frequency spectrum. Because internal circuit elements and the receiver system's environment have temperature, they affect the amount of noise added to the signal [25].

When the section of the antenna that collects the signal, known as the feedhorn, (described in 4.4.1) is connected to the parabolic reflecting dish, it points toward the antenna. This orientation can cause problems, depending on the temperature of the medium the feedhorn points to. Because the reflecting dish points to the sky, it only sees a noise temperature of about 30 Kelvin [4]. The feedhorn, however, must point at the dish, causing it to see temperatures higher than the sky temperature. Although the reflecting dish acts as a barrier between the feedhorn and the Earth, which has a noise temperature of 290 Kelvin, the antenna still detects some Earth noise. It is therefore designed to limit this excess noise. The feedhorn collects more of the signal from the center of the dish, where there is little Earth noise affecting the signal, than from the corners where the hot noise is much stronger [4].

In addition to thermal noise from the sky and Earth, circuit components also create noise. Because resistors within a circuit dissipate energy in the form of heat, they directly attribute to the circuit's noise level. Energized charge carriers again emit different frequencies as they drop energy levels, which combine to create system noise [25].

For a given bandwidth (B), temperature in Kelvin (T), and resistance impedance (R), the root mean square voltage noise is given as,

$$V^2 = 4k_BTBR \qquad (4.3)[25]$$

Where k_B is the Boltzmann constant. Since power is related to voltage squared, the noise power is,

$$P = k_BTB \qquad (4.4)[25]$$

Which can then be compared to the original signal for decibel power. Following an example from Radio Electronics, at room temperature (290 K), a bandwidth of 1 Hz, and an impedance of 50 Ω, a circuit system can cause 0.9nV of noise [25],

$$V = \sqrt{4 * 1.3803 * 10^{-23} * 290 * 50 * 1} = 0.9nV \qquad [25]$$

Because noise significantly affects service, companies do their best to install antennas such that the noise received is limited. After a successful installation, noise does not cause service issues. According to Long, thermal noise, the most significant source of noise, adds only hundredths of decibels to unwanted frequencies when the antenna is correctly installed [4].

4.4 The Low Noise Block Downconverter

At the focus of a parabolic reflector is a device that converts the electromagnetic wave into an electrical signal for televisions to convert to audio and video. This device, known as a low noise block downconverter, converts the signal into an electrical signal while adding minimal noise to the signal frequency spectrum. The electrical signal is converted from high radio frequencies that require special cables to transmit to lower frequencies that can be transmitted through basic wire as it goes through the low noise block downconverter. This final signal is transmitted to the set top box where it is converted into an approximation of the original analog signal through the television and its speakers. A diagram of the whole process is shown in figure 5 [26, 27].

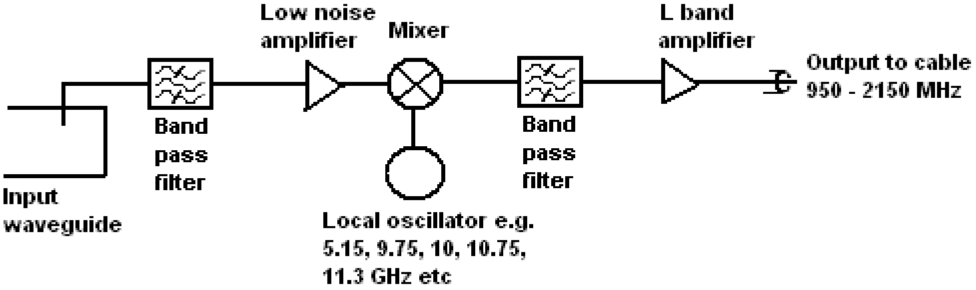

Figure 5: A basic diagram of the low noise block downconverter. The waveguide collects the focused signal and brings it to a pin that acts as a whip antenna. The first bandpass filter then filters out all unnecessary signals so signal can be amplified and mixed to a lower frequency that is again filtered. This L-band signal is then amplified again and sent to the set top box to create television [26].

4.4.1 The Feedhorn and Waveguide

Once the incoming signal has been focused by the reflector dish, it must still be guided to the actual low noise block downconverter. In order to collect this signal, the receiving equipment begins with a feedhorn. The feedhorn orients the electromagnetic wave so that it propagates properly through the waveguide minimizing signal loss [28].

When a radio frequency electromagnetic wave needs to be transmitted in a single dimension or guided somewhere, a waveguide is used. Because in free space these waves propagate isotropically, radio waves can experience signal loss at a rate related to the inverse-square of the distance it has traveled. Although satellite signals are more focused than isotropic radiation, they still transmit an expanding beam that experiences signal attenuating due to distance traveled. Once a portion of this beam has been focused by the parabolic dish, it propagates through the waveguide. Waveguides remove signal loss by limiting the dimensions of motion of an electromagnetic wave. This process of focusing the signal into a waveguide allows the strongest possible signal to reach the low noise block downconverter [4, 26, 30].

In order to reduce signal loss, waveguides propagate radio waves in only one dimension. These hollow rectangular tubes with dimensions on the order of the wavelength of the wave use total internal reflection to trap the electromagnetic wave. The materials within a waveguide create an environment where the signal does not get attenuated as it reflects down the tube. At any point inside the tube, the electric field amplitude of the electromagnetic wave will vary in time as a sinusoid based on the frequency. At the end of the waveguide, there is a resonant pin that acts as a whip antenna. This antenna transforms the electromagnetic wave signal into an electric current signal that the low noise block downconverter can process. If there are multiple signals at different polarizations, the waveguide can contain two orthogonal pins each meant to collect a different signal without any interference from the other signal. Once the electric field has affected the electrons to make an electric current, this high frequency electrical signal continues onto other sections of the low noise block downconverter for further processing [4, 26].

4.4.2 The Bandpass filters and Mixer

After the signal has been collected by the whip antenna, the low noise block downconverter must filter out any unnecessary noise. The first bandpass filter eliminates any frequencies that fall outside of the allowed bandwidth of the incoming signal. Once this noise has been eliminated, the signal is amplified to match amplitude of the signals outputted by the local oscillator. The mixer then takes the inputted signal and adds frequencies added by the local oscillator [26].

The local oscillator operates as a waveform creator. With a small dielectric resonator, the local oscillator creates a specific frequency for the mixer to use. Because these are mass produced, these oscillators can have frequency fluctuations due to temperature as large as 2MHz However, when compared to the bandwidth, this difference does not significantly affect the signal read by the set top box. In order to convert the high frequency radio signal into a lower frequency, the mixer combines the local oscillator frequency with the incoming signal (in the K_a or K_u bands) in the forms of addition, subtraction, and multiplication. These signals are then sent through another bandpass filter designed to eliminate all but the L-band signal frequency. Because the L-band frequencies range from 1 GHz to 2 GHz, these signals can be sent through normal wires without problems. After amplification, this signal is sent via normal wires to the set top box where it is finally converted into television [4, 26].

4.5 Conclusion

Although the hardware converts the information into waves and then back into useable information, it does not improve upon signal correction. Antennas are only designed to collect a waveform and transform it into an electrical signal. They are designed to collect a specific part of the frequency spectrum, meaning they do receive the intended signal better than noise signals, however, they do not ensure the strongest signal received is the intended signal. Once in electrical form, the low noise block downconverter simply converts the signal to a lower frequency. Again this portion of the conversion process does not ensure that proper signal is passed through but simply lowers noise outside of the expected bandwidth. The electrical signal received by the set top box has the same information that the signal inputted to the low noise block downconverter contains.

Chapter 5: Conclusion

After research, I have determined that the current method of data transmission is the best providers can do for satellite television. My previous conceptual understanding of signals and noise led me to believe that signal loss and errors in data transmission are easily overcome with technology. Although technology can solve some of these issues, the current method of data transmission focuses on avoiding these errors reaching the television. As Long explains, a single bit error in satellite television happens about once every 4 hours with more efficient error correction than explained previously [4]. When possible, the providers have created a system to minimize error instead of letting the signal become too weak and forcing technology to correct errors. Although this provides a robust form of information transfer, it is clear that an unobstructed line of sight from the antenna to the satellite is required.

When weighing scattering issues versus propagation issues through the Earth's atmosphere, it seems clear that the correct frequency range has been chosen for satellite television. Although the signal is too severely attenuated by leaves or other obstacles to be useful and strong enough rain can cause signal loss, the fact that these high frequency radio waves can travel through the ionosphere makes them the clear choice. In order to combat loss of information, these signals are frequency modulated, which means the loss of amplitude due to scattering does not affect the information contained within the waveform. Therefore due to the limited choices, the frequency assignment for satellite television is the best option.

The process of converting analog information to digital information is also created to limit signal error. In analog transmission, the information is contained in both the waveform's amplitude and frequency, leaving it susceptible to loss of information due to scattering. Through the use of discrete values represented in frequencies, digital information does not depend on amplitude, making it much less susceptible to information loss. The process of detecting and correcting errors in this form of data is much simpler even though it does take a large amount of capacity to include these codes. This analog to digital conversion is ideal as it can be corrected if necessary, while also providing a much more compact form of representing information.

The hardware simply limits the amount of noise received. Because the most common origin of noise is heat, the receiver points to the coldest part of the reflecting dish in order to minimize incoming noise signal. Although it does not correct for noise or reduce the received noise, the antenna is set up so that it receives as little noise as possible preventing the signal from being drowned out.

Because the system of sending and receiving information via satellites is designed to avoid signal error, and when necessary fix it, antennas must be placed so that they have a clear line of sight to satellites. This research shows that the current method of satellite communication, although complicated, combines what seem to be the best possible options given the limiting factors.

References

[1] https://www.plunkettresearch.com/statistics/telecommunications-market-research/

[2] http://www.sia.org/wp-content/uploads/2014/05/SIA_2014_SSIR.pdf

[3] https://en.wikipedia.org/wiki/International_Telecommunication_Union

[4] Long M. (1999). *The digital satellite TV handbook.* Boston: Newnes.

[5] Proakis, John R (1983). *Digital Communications* (3rd edition). New York, New York: McGraw Hill.

[6] http://www.space-airbusds.com/en/news2/do-you-know-how-a-communications-satellite-works.html

[7] https://en.wikipedia.org/wiki/Direct-broadcast_satellite_television

[8] http://www.antenna-theory.com/antennas/reflectors/dish.php

[9] https://en.wikipedia.org/wiki/Amplitude_modulation

[10] Shannon, C.E. (1949). "Communication in the Presence of Noise," in *Proceedings of the IRE* , vol.37, no.1, pp.10-21.

[11] https://en.wikipedia.org/wiki/Wideband

[12] Shen, Liang Chi and Kong, Jin Au (1987). *Applied Electromagnetism* (2nd edition). Boston, MA: PWS publishers.

[13] https://en.wikipedia.org/wiki/Satellite_television

[14] Helhel et. al. (2009). “Measurement of dielectric constant of thin leaves by moisture content at 4mm band,” in *Progress In Electromagnetics Research Letters*, vol 7, pp. 183-191.

[15] https://en.wikipedia.org/wiki/Mie_scattering

[16] Cha, Philip P. and Molinder, John L. (2006), *Fundamentals of Signals and Systems*. New York, New York: Cambridge University Press.

[17] https://en.wikipedia.org/wiki/Frequency-shift_keying

[18] https://en.wikipedia.org/wiki/Aliasing

[19] Hamming, R. W. (1950), Error Detecting and Error Correcting Codes. Bell System Technical Journal, 29: 147–160.

[20] https://en.wikipedia.org/wiki/Antenna_(radio)

[21] https://en.wikipedia.org/wiki/Dipole_antenna

[22] https://en.wikipedia.org/wiki/Dipole_antenna#/media/File:L-over2-rad-pat-per.jpg

[23] http://farside.ph.utexas.edu/teaching/302l/lectures/node119.html

[24] https://en.wikipedia.org/wiki/Whip_antenna

[25] http://www.radio-electronics.com/info/rf-technology-design/noise/electronics-radio-frequency-rf-noise.php

[26] http://www.satsig.net/lnb/explanation-description-lnb.htm

[27] https://en.wikipedia.org/wiki/Low-noise_block_downconverter

[28] https://en.wikipedia.org/wiki/Feed_horn

[29] https://en.wikipedia.org/wiki/Waveguide_(electromagnetism)