Ionospheric Sounding During a Total Solar Eclipse

William Charles Lloyd

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Electrical Engineering

Dr. Robert McGwier
Dr. Gregory Earle
Dr. Joseph Baker

April 29, 2019
Blacksburg, Virginia

Keywords: ionosonde, ionosphere, ionospheric sounding, total solar eclipse, August 2017
total solar eclipse, ionogram, GNU Radio

Ionospheric Sounding During a Total Solar Eclipse

William Charles Lloyd

ABSTRACT

The ionosphere is a constantly changing medium. From the sun to cosmic rays, the ionosphere proves to be a continually interesting area of study. The most notable change that occurs in the ionosphere is the day and night cycle. The ionosphere is not a singular medium, but rather made up of different sections. The day side of the ionosphere consists of a D, E, F1, and F2 layer. The night day of the ionosphere consists of an E and F layer. These layers all have different properties and characteristics associated with them. A notable interaction is how radio waves propagate through the ionosphere. A radio wave can either reflect, refract, or pass through a layer of the ionosphere depending on the frequency of the signal, among other sources of disturbance. The ability to have a radio wave reflected back downwards is a core principle of an ionosonde, which measures the height of the ionosphere. A solar eclipse presents a night side ionosphere condition during the day. The change in the ionosphere that the eclipse will cause is something not a lot of research has gone into. This thesis aims to elaborate on the design and development of an ionosonde along with eventual ionosphere readings during the August 2017 total solar eclipse.

Ionospheric Sounding During a Total Solar Eclipse

William Charles Lloyd

GENERAL AUDIENCE ABSTRACT

The atmosphere that surrounds the earth is made up of various unique regions. The region of interest for this thesis is the ionosphere. The ionosphere plays an important role in wireless communication of radio waves. It follows that changes in the ionosphere are something of great interest and study. A notable change that the ionosphere undergoes on a daily basis is the shift from the day side to the night side. A solar eclipse serves not only as a spectacular sight, but also to bring a night side condition to the day side. This thesis aims to uncover the changes that will occur to the ionosphere during the August 2017 total solar eclipse.

Table of Contents

Page

Page

Introduction

The first project worked on during my graduate career involved SDR techniques with GNU Radio and a coming eclipse. This project aimed at furthering the understanding of how the height of the ionosphere changes in the presence of an eclipse. It is well documented that wireless transmission can be altered based on the medium it is traveling through. With air being the most common medium, a sudden change in the atmosphere from an eclipse will surely affect transmitting radio waves. The degree by which radio waves are affected will be seen by how the ionosphere changes during the eclipse.

Literature Review

GNU Radio

One of the tools used for this project is GNU Radio. GNU Radio is a signal processing program that allows the user to use software-defined radio (SDR) techniques on the computer [1]. The user is prompted to create what is known as flowgraphs using the various signal blocks provided as illustrated in [2, Fig. 1]. This is extremely useful as it provides a means of visual identification for signal processing. A flowgraph is an arrangement of blocks that do the signal processing. GNU Radio, at its core, is a signal processing tool based in C++ and Python with a general user interface (GUI) in the form of blocks that make a flowgraph. All of the blocks perform some sort of signal processing, which as stated, done under the hood as C++ and Python. Typically, the main runtime processes are done with C++ as Python is slower due to its object-oriented construction. While Python can be used for signal processing during runtime, more often is the GNU Radio interface exported to Python for scripting together the flowgraph [1].

Much like reading, flowgraphs run from left to right. This can be visually identified through the source and sink blocks. These blocks only connect with other blocks from the right and left respectively, indicating a visual start and stop point. An advantage that GNU Radio offers is that should a signal processing function not be present in its native repository, the ability to write, code, and implement a custom block exists. These blocks are known as out-of-tree (OOT) module blocks. Additionally, these custom blocks are able to be packaged and uploaded to online repositories, such as GitHub, through which others may download and install to use.

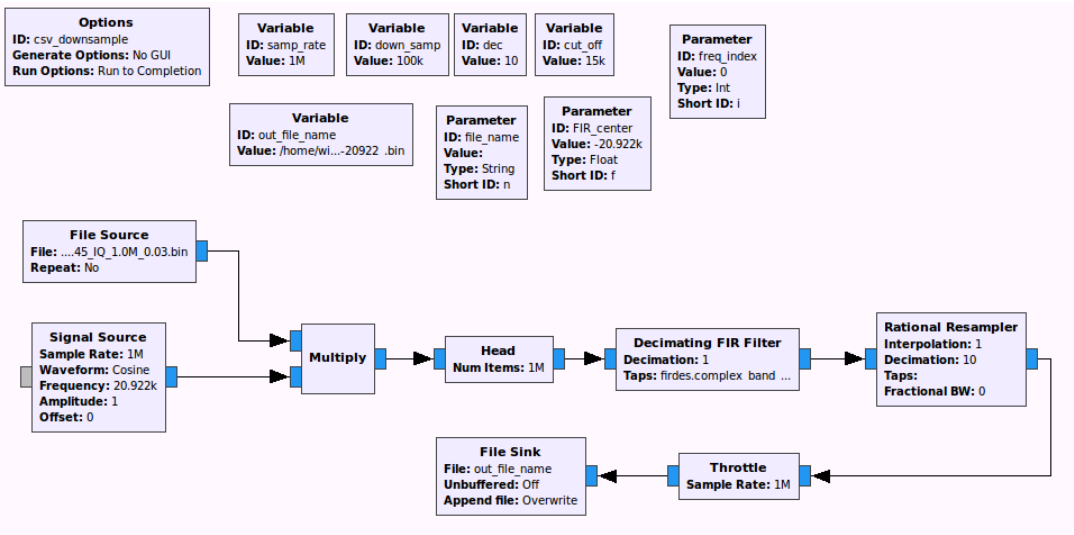

Fig 1: GNU Radio Flowgraph

The flexibility of GNU Radio cannot be understated. GNU Radio is Python and C++ based. Depending on the application of GNU Radio, it may be beneficial to code directly with the Python file generated during run time with the GNU Radio GUI. Working directly with the Python file has several advantages. The implementation of input and output files in various configurations as well as variable manipulation is more straightforward in Python. Additionally, the ability to execute the file from the command

line, as GNU Radio is mainly found on Ubuntu Linux machines, allows for scheduling and other timing options.

Signal processing in GNU Radio is not based in terms of samples, but rather in terms of items [1]. Not every block functions based on samples. Packets, frames, bits, bytes, and symbols are some of the other items that GNU Radio blocks can operate off of. The inputs and outputs of the blocks indicate, based on color, what item type they use, as illustrated in [3, Fig. 2]. Blocks will pass data streams through the flowgraph so long as the item type and bit size are consistent.

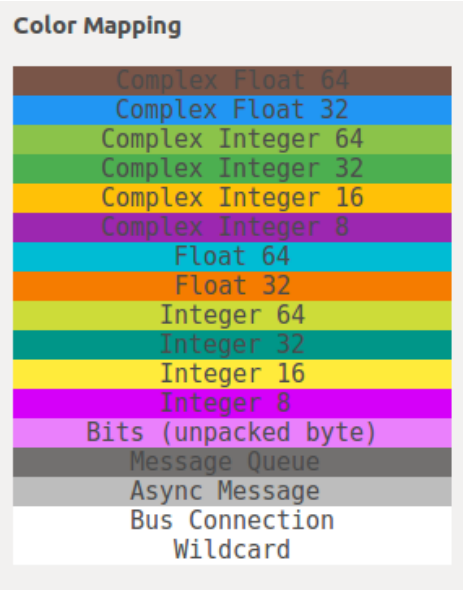

Fig 2: GNU Radio Item Color Map

Blocks in GNU Radio will vary in item type, but also how the items are passed through. There are four types of blocks within the GNU Radio framework. The first and most notable block is the synchronous block. Synchronous blocks are 1:1 in their data ports. The input consumes data and will produce data in equal amounts. The GNU Radio scheduler tells the block to take in N number of items to fill an input vector. The block takes the vector and applies signal processing on it. The output vector is filled with the N

number of processed items and streamed along to the next block. A synchronous block can have any number of input and output ports. There are two special kinds of synchronous blocks. Synchronous blocks without an input are called source blocks where signal generation is formed. Synchronous blocks without an output are called sink blocks where data streams are collected. The main characteristic is that the output and input vectors, should there be both, are the same length. The next two block types are the decimation and interpolation blocks. The key feature of these blocks is that the items consumed at the input and the items produced at the output will not be the same. The data ports are N:1 for decimation blocks and 1:M for interpolation blocks. This means that after whatever signal processing the block performs, the decimation block will operate based on the input and output relationship as seen in [4, eq. (1 – 2)] respectively.

$$input\ items = output\ items * decimation \qquad \text{eq. (1)}$$

$$input\ items = \frac{output\ items}{interpolation} \qquad \text{eq. (2)}$$

The output vector length of the decimation and interpolation blocks are modified by the decimation and interpolation factor and therefore these blocks are not 1:1. The last kind of block is the general block where there is no set relationship between the input vector and the output vector. The N:M relationship is what all the other block types are simplified from.

Plug in blocks in GNU Radio allows for a wide range of hardware applications. A notable hardware connection is the Universal Software Radio Peripheral (USRP). These SDR interfaces, produced by Ettus Research, offer wide ranges of applications such as FM transmitters and receivers, amateur radio uses, and passive radar usage among others. Other hardware devices can operate using similar plugin blocks or are found as user-

supplied blocks. This is all to say that GNU Radio applications with various hardware are readily available for a range of uses.

Ionosphere

The ionosphere is a conducting layer of the upper atmosphere and is considered to be the lower boundary to the magnetosphere. The discovery of such a conducting layer did not happen overnight. One of the earliest experiments that brought the attention of many scientists towards the sky was Marconi's transatlantic radio wave telegraphy experiment in 1901 [2]. This over-the-horizon radio communication was not thought possible at the time. Nevertheless, this experiment prompted Heaviside and Kennelly to develop theories predicting an ionized layer in the atmosphere in 1902 [2]. These theories were developed independently and thus the predicted layer was named the Heaviside-Kennelly layer. It was not until 1926 that Appleton and Barnett gave the theory of a conducting layer in the ionosphere experimental proof. The experiment done compared the fading of radio signals from one antenna to another through both a ground and sky wave [2]. Later in the same year, Briet and Tuve developed what is known as the ionosonde. This machine was able to obtain altitude measurements of the ionosphere [2]. The conducting layer in the atmosphere would eventually come to be known as the ionosphere.

Since its discovery and subsequent confirmation, the ionosphere has been studied at length to understand its composition, location, and how it interacts. The ionosphere is composed of various layers. There exists a D, E, F1, and F2 layer. Each of the layers of the ionosphere is distinct in their composition and particle interactions. The reason there are different layers to the ionosphere is a mixture of effects. Solar radiation has a specific

energy spectrum that is absorbed in different atmospheric altitudes based on atmospheric composition [3]. Additionally, recombination of ions with free electrons vary with height and density. This can be mathematically described as Chapman Theory, as seen in [6, eq. (3)], where $q(z)$ is the ion production rate, $n_a(z)$ is the neutral density, and $I(z)$ is the intensity of sunlight.

$$q(z) = n_a(z)\sigma_i I(z) \quad \text{eq. (3) [2]}$$

Chapman Theory states that sunlight will decrease in intensity going down the atmosphere while the density of the atmosphere will slowly decrease going upwards. These opposing vectors cause a maximization of ion production at some altitude, which is the ionosphere [4]. The layering of the ionosphere is further defined by how the different atmospheric molecules are affected by solar radiation. The methods by which ion production is done dictates which molecule species inhabit which layer of the ionosphere. The molecular variability by altitude is illustrated in [6, Fig. 3]

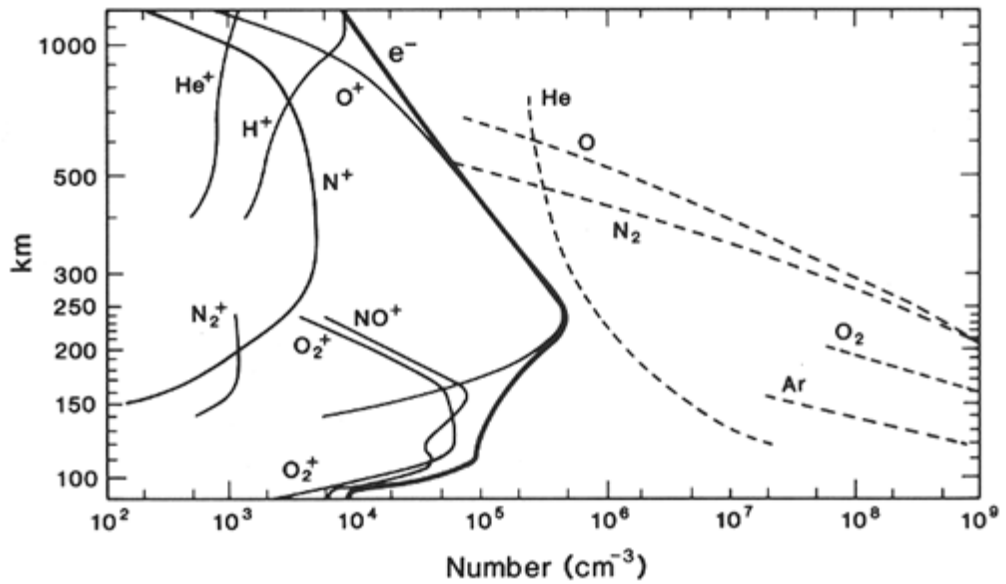

Fig 3: Molecular Distribution of Density vs. Altitude [4]

The two production methods that dictate molecular variability in the ionosphere layers are photoionization and impact ionization.

$$X + hv \rightarrow X^+ + e \qquad \text{eq. (4) [4]}$$

$$X + e \rightarrow X^+ + 2e \qquad \text{eq. (5) [4]}$$

Photoionization, as seen in [7, eq. (4)], occurs when photons from solar radiation interact with atoms to eject an electron out resulting in an ion and free electron. Impact ionization, as seen in [7, eq. (5)], occurs when an energetic electron impacts with an atom to eject an electron resulting in an ion and free electrons. These processes contribute to the ionization of atoms in the ionosphere. The day side is dominated by photoionization due to solar radiation and maximizes at noon. The night side of the ionosphere is dominated by impact ionization, regarding production, as the sun can no longer play a role and maximizes at midnight. Generally speaking, the total electron density decreases during the night side. The electron density in the ionosphere is also affected by loss mechanisms. There are four different loss mechanisms that serve to inhibit ion production.

$$X^+ + e^- \rightarrow X + hv \qquad \text{eq. (6) [4]}$$

$$XY^+ + e^- \rightarrow X + Y \qquad \text{eq. (7) [4]}$$

$$WX^+ + YZ \rightarrow WX + YZ^+ \qquad \text{eq. (8) [4]}$$

$$X^+ + YZ \rightarrow XY + Z^+ \qquad \text{eq. (9) [4]}$$

Radiative recombination is the reverse of photoionization, as seen in [7, eq. (6)]. This slow process occurs when an ion captures a free electron and emits a photon with equivalent energy. Dissociative recombination is a fast reaction that, as seen in [7, eq. (7)], occurs when a free electron interacts with a positive ion. The electron causes the

positive ion to dissociate into neutral atoms. The final two loss mechanisms have a zero net loss in electrons. Charge exchange, as seen in [7, eq. (8)], and Atom-Ion Interchange, as seen in [7, eq. (9)], are interactions between atoms rather than free electrons. These processes alter the ion concentrations which can lead to radiative recombination or dissociative recombination due to the differences in molecular bond strength that may occur. These production and loss mechanics play a large role in determining how a signal of a given frequency will behave when interacting with the ionosphere and its layers.

As stated previously, the ionosphere is composed of four layers. Not all of these layers exist all the time. During the day side, the D, E, F1, and F2 layers are present. However, the E layer and subsequently combined F layer are the only layers present during the night side as illustrated in [8, Fig. 4]. The increased plasma during the day side from high photoionization leads to the formation of the lowest layer in the ionosphere, the D layer.

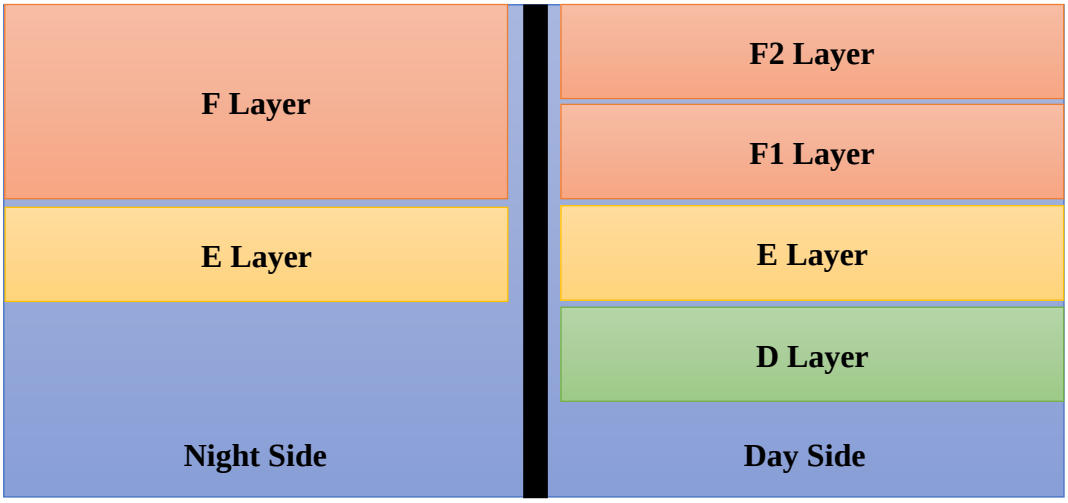

Fig 4: Ionosphere Layers for Day and Night

The D layer has an upper boundary of approximately 90 km and a lower boundary of approximately 70 km. The chemistry of this layer deals mostly with molecular

nitrogen and molecular oxygen. One of the main ionizations that occurs in the D layer is from Lyman-α radiation lines [5]. A Lyman-α line is the hydrogen spectral line that is emitted when an electron changes orbit. These lines from the solar spectrum are powerful enough to penetrate deep in the atmosphere to the D layer to ionize the particles there. Nitric oxide (NO) and nitrate (NO_3^-) are the dominant molecular structures that occur due to the ionization of the molecular nitrogen and molecular oxygen [5]. Additionally, hard x-rays are capable of penetrating the atmosphere to the D layer and thus will ionize the particles there. Cosmic rays from the universe will consequently ionize ionosphere particles as well. The E layer is located above the D layer with a lower boundary of approximately 90 km and an upper boundary of approximately 170 km. Ionization that occurs in the E layer due to soft x rays and extreme ultraviolet (EUV) rays. These processes interact with molecular oxygen and nitrogen to produce O_2^+ and NO^+ [4]. The abundance O_2^+ and NO^+ in the E layer is a result of the solar radiation absorption by the atmosphere to alter the wavelengths. This change allows for a different radiation band allows the molecular oxygen to ionize in such a way that allows O_2^+ and NO^+ to flourish. The F layer of the ionosphere has a lower boundary located approximately at 170 km and an upper boundary located approximately at 500 km. The upper boundary is generally agreed upon as an altitude of 500 km, but there is no hard cutoff defined. The sun is the primary source of ionization in the F layer. EUVs are capable of penetrating the outer atmosphere with ease to reach the F layer. As a result, the solar radiation is capable of breaking up molecular oxygen to produce O^+ [4]. There is enough significant difference within the F layer to subdivide it into an F1 and F2 layer. One significant difference is that the F1 layer is in photochemical equilibrium and the F2 layer is not. In fact, both the

F1 and E layers share this characteristic. Photochemical equilibrium for an ionosphere layer indicates that transport between within the layers is approximately zero [4]. Another difference is that the F1 layer, as well as the E layer, are considered Chapman layers. Chapman wished to discover how the ionosphere layers were formed. He derived a function that related the production rate of ions to the scale height and the slant angle. This simple Chapman function is able to describe the E and F1 layer [3]. The D and F2 layers are complicated by transport of plasma and therefore are not Chapman layers.

The ionosphere is composed of plasma generated by solar radiation, and other forms of production, to form an abundance of free electrons and ions of various type at various altitudes. A plasma is an ionized gas where the amount of positive particles is approximately equal to the number of negative particles [6]. A common thought is to associate plasmas as being hot. While true, the plasmas that are of relevance to the ionosphere are considered cold. Cold plasma is simply one where the thermal motion of the electrons is small enough to ignore in calculations. The variation in electron thermal motion gives multiple classes of plasmas. However, all plasmas will still exhibit collective behavior. Plasmas are a collection of charged particles which allows for collective interactions due to the attraction and repulsion of positive and negative particles. As a result, plasmas are subject to various behaviors because of this.

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\epsilon_0} \qquad \text{eq. (10)}$$

$$\nabla \cdot \boldsymbol{B} = 0 \qquad \text{eq. (11)}$$

$$\nabla \times \boldsymbol{E} = -\frac{\delta \boldsymbol{B}}{\delta t} \qquad \text{eq. (12)}$$

$$\nabla \times \boldsymbol{B} = \mu_0 \left(\boldsymbol{J} + \epsilon_0 \frac{\delta \boldsymbol{E}}{\delta t} \right) \qquad \text{eq. (13)}$$

A plasma can be well described through electromagnetic behavior. Maxwell's equations, as seen in [10, eq. (10) – (13)], can be used to describe the electrical behavior of a plasma. If an electrical field is applied to a plasma, the positively charged particles and negatively charged particles will separate which can allow for oscillations to occur [6]. Magnetic perturbations that occur within a plasma can cause gyration along field lines. These spiraling effects can cause sources of current. Describing the magnetic interactions within a plasma using the Maxwell equations requires the use of the Lorentz Force equation, as seen in [11, eq. (14)]. The Maxwell equations along with the Lorentz Force helps describe magnetized plasma. Due to magnetic fields having a preferred direction, this infers that magnetized plasmas exhibit anisotropic behavior.

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{V} \times \boldsymbol{B}) \quad \text{eq. (14)}$$

The thermal electron temperature plays a role in determining whether the plasma is considered hot or cold. It also plays a role in determining the effective collision rate of the plasma. Depending on the particle density of the plasma as well as the thermal electron temperature, a plasma could be considered to be collisionless or collisional. This becomes significant regarding the ionosphere layers as the varying particle density classifies some layers to be collisional and some layers to be collisionless. The characteristics of a plasma allow for the knowledge of how radio waves will interact.

The Sun's Impact

The sun is one of the biggest components of the constantly changing ionosphere plasma. At nearly 93 million miles away, the impact the sun has on the ionosphere cannot be understated. The sun is composed of several regions. Internally, the sun is made up of the core, the radiative zone, and the convective zone [7]. The core of the sun is where

nuclear fusion occurs to form helium from hydrogen. The sun is composed mainly of hydrogen and helium at a split of 75% hydrogen and 25% helium [7]. The hydrogen in the core of the sun reacts in nuclear fusion to form helium and energy as seen in [12, eq. (15)].

$$4H^1 \rightarrow He^4 + 26.73MeV \qquad \text{eq. (15) [7]}$$

This reaction to form He^4 can take millions of years, which gives the lifetime of the sun several billion years. The energy generated in this fusion process is transported out of the core through radiation in the form of photons in the radiative zone [7]. This process can take thousands of years for these energetic photons to reach the surface of the sun. The plasma density inhibits any direct path outwards from the core. The photon path from the core to the surface is similar to Brownian motion. Once the photons leave the radiative zone, they enter the convective zone where instabilities in heat magnify. The photons entering the convective zone have lost energy from collisions in the radiative zone. This loss in energy, as well as an increased distanced from the core, contributes to a strong temperature gradient as illustrated in [13, Fig. 5].

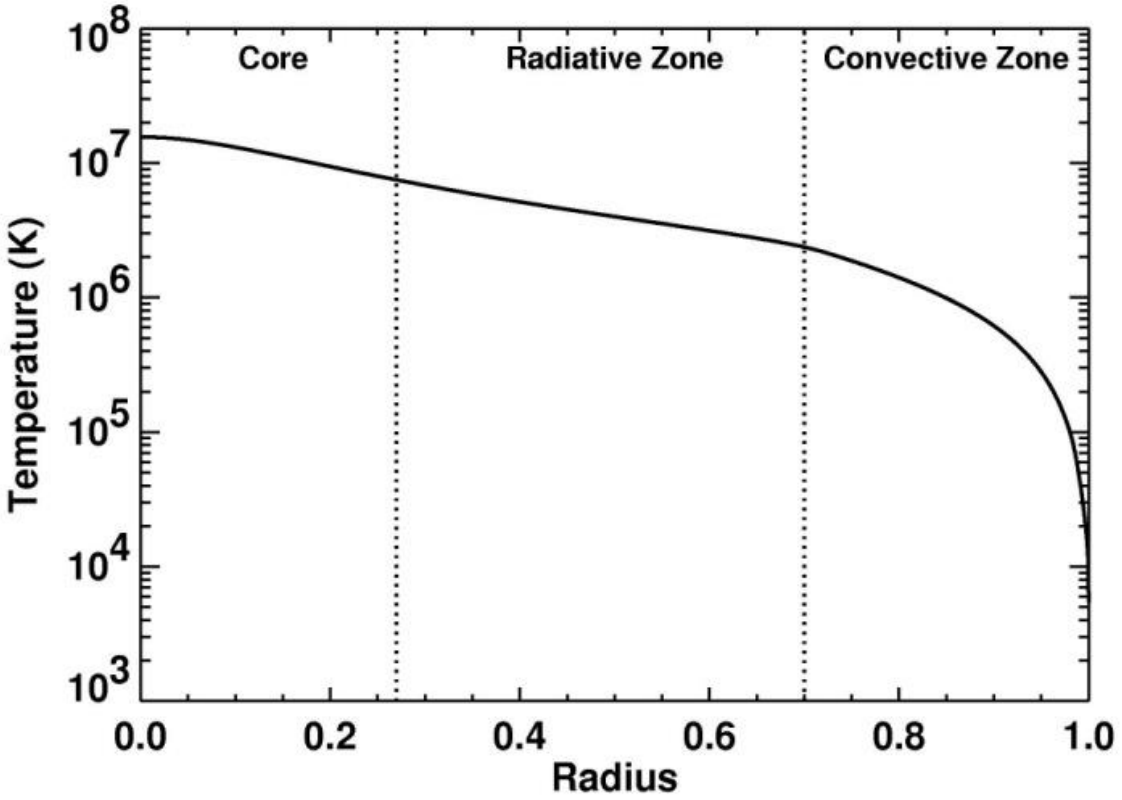

Fig 5: Solar Interior Temperature Gradient [7]

Much like the earth, the sun has an atmosphere where each region is distinct based on solar activity and temperature. There are four regions that compose the atmosphere of the sun. The first of these layers, and also the lowest, is the Photosphere. This region of the sun is characterized by sunspot and granule activity as well as being the coolest in temperature [7]. At a thickness of about 500 km, the activity that occurs here will impact ionospheric conditions on earth. Sunspots are regions of strong magnetic activity on the sun which serve to inhibit the upward convection of plasma from below [8]. These regions appear much darker than the surrounding surface as the temperature is significantly lower due to a lack of convection. Granules are a cell-like pattern on the surface of the sun due to plasma convection. These lighter regions of the granules indicate an upward flow of plasma while the darker regions indicate a downward flow of

plasma [8]. The next region in the solar atmosphere, right above the photosphere, is the chromosphere. The chromosphere is characterized by solar activity in the form of filaments, prominences, and plages. Additionally, the temperature rises due to the absorption of solar activity sound waves originating from the photosphere below [8]. Filaments are dark streams of plasma that occur due to strong magnetic fields. The magnetic fields serve to inhibit convention and therefore appear dark against the chromosphere. Plages are similar in that regard, however, are bright regions against the chromosphere and are typically associated with radio wave emission [8]. A prominence occurs when two sunspots of strong magnetic strength interact with each other. These sunspots form a magnetic connection which causes ropes of the magnetic field to travel from one sunspot to the other. These magnetic flux ropes begin to rise and form bright arcs over the surface of the sun. The three occurrences in the chromosphere are illustrated in [14, Fig. 6]

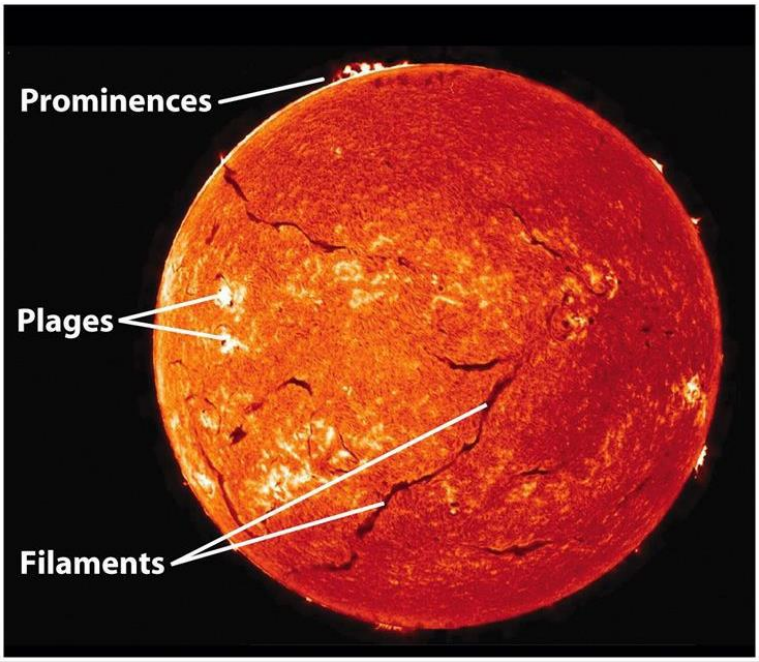

Fig 6: Chromosphere Solar Activity [8]

The third region up in the solar atmosphere is known as the transition region. This region is still very much an active area of study due to the temperature changes that occur here. The transition region is mainly characterized by the factor of 100 temperature increase from the chromosphere to the outer most region of the solar atmosphere [8]. The corona is the last layer of the solar atmosphere and is subsequently the hottest region as illustrated in [15, Fig. 7].

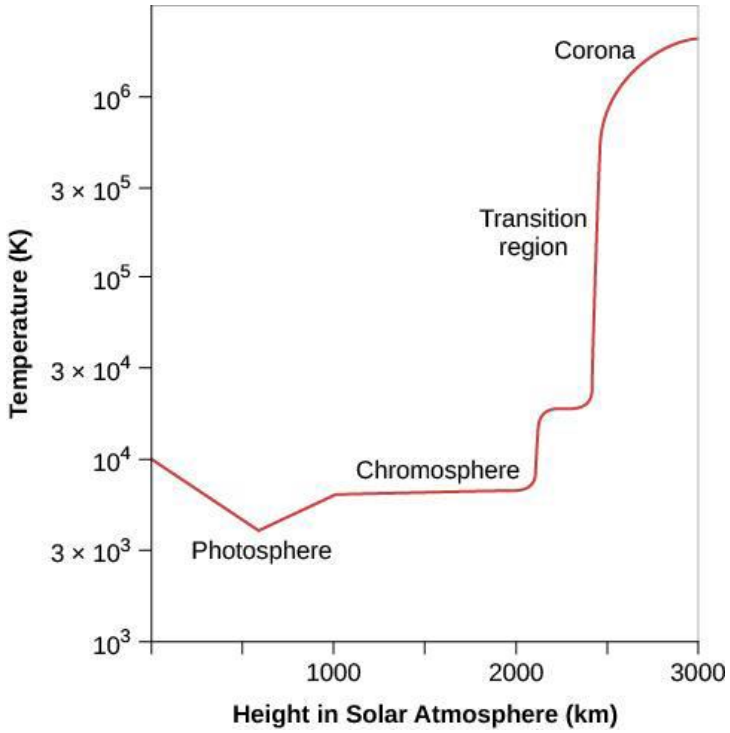

Fig 7: Solar Atmosphere Temperature Changes [8]

Invisible to the eye, the corona is best observed through the lens of x-rays or a total solar eclipse. Apart from the temperature, the occurrence of magnetic activity in the form of coronal holes and streamers help characterize the corona. The magnetic field of the sun is much larger than the one on earth in addition to having to cover more stellar mass. The variation in the sun's magnetic field can cause apparent regions where magnetic field lines are considered open can the plasma can escape [8]. These are known

as coronal holes. Coronal streamers are the opposite. The strong magnetic fields are considered closed and prevent plasma from escaping out.

It is apparent that solar activity is heavily based on how the magnetic field of the sun is interacting. Many features of the sun are either caused by or influenced by the strong magnetic field. This can lead to variation in the solar output of plasma and radiation towards the earth. This subsequently causes variation in the ionosphere which is affected by the plasma and radiation from the sun. The variation the sun experiences with its output is known as the solar cycle. The solar cycle occurs in 11-year periods alternating between solar maximum and solar minimum. The reason behind why the sun experiences a cycle due to its magnetic field and rotation. Unlike the earth where the magnetic field is formed due to an iron core, the magnetic field on the sun is generated in the convection zone. A surface magnetic field can be subjected to alteration more easily than an internal magnetic field. This magnetic field alteration is extremely apparent due to the differential rotational of the sun. The sun is a celestial body composed of plasma which is not the best state to prevent the centripetal force from changing the surface. The sheer size and subsequent gravitational force from the sun induce varying rotational rates based on latitude as illustrated in [17, Fig. 8]. The surface magnetic field and the differential rotation of the sun create what is known as the solar dynamo [8]. The magnetic field lines oriented poleward are slowly dragged perpendicular as the equator of the sun rotates faster than the poles. This eventually causes the magnetic field lines to orient themselves horizontally across the surface of the sun. The mix of north and south field lines can cross paths and form localized closed loops. Eventually, the field lines will untangle and reform but in reverse polarity. If the sun started off with the north pole and

south pole both having their respective polarity, after the 11-year solar cycle, the magnetic north pole will be located in the geographic south pole of the sun and the magnetic south pole will be located in the geographic north pole of the sun.

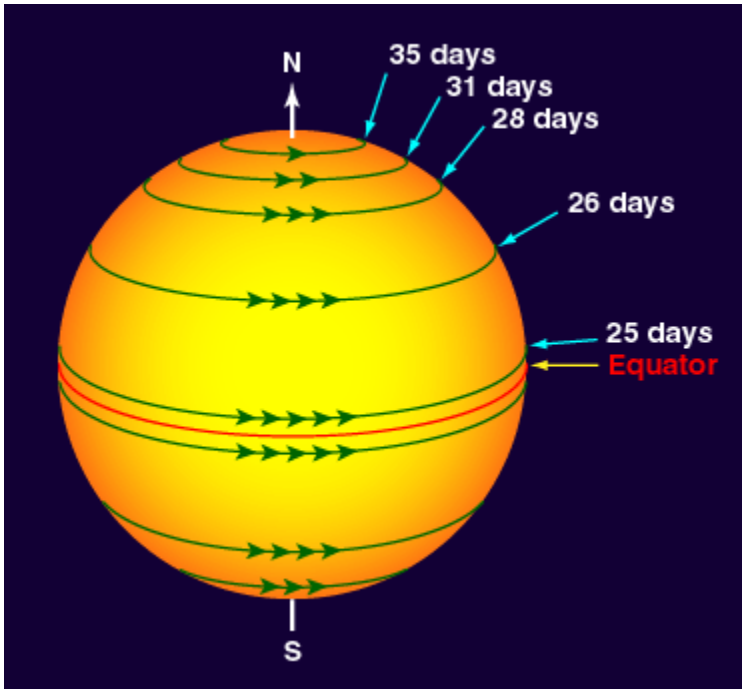

Fig 8: Differential Rotation [8]

Many characteristics of the sun are influenced, in one way or another, by the magnetic field. When the field lines are dragged across the equation by differential rotation and subsequently tangled, the magnetic influences maximize and transitions into what is known as solar maximum. Likewise, when the field lines are oriented poleward and are semi-dipolar, the magnetic influences minimize and the sun is in what is known as solar minimum. The changes in solar activity during solar maximum can affect the earth's ionosphere more spontaneously than if the sun was in solar minimum. While in solar maximum, sunspots are more likely to occur due to the tangled magnetic field lines. Sunspots serve to limit the solar output of the sun. As the ionosphere is in relative equilibrium between the decreasing sunlight going down and the decreasing atmospheric

pressure going up, the multitudes of sunspots during solar maximum can affect how the plasma in the ionosphere will react to traveling radio waves. In addition, solar weather effects will be more pronounced. Solar flares and coronal mass ejections, should they hit the earth, can cause huge disruptions in the plasma frequency and density of the ionosphere. The ionosphere on earth can change somewhat arbitrarily from outside forces and the current state of the sun. Solar minimum presents a more stable output from the sun and consequently a more stable ionosphere. Solar maximum presents a more volatile output and therefore a more varying ionosphere.

The abundance of plasma in the upper atmosphere will cause a radio wave traveling through the ionosphere to experience varying effects based on its own frequency and the surrounding electron density of the plasma [4]. One of the more notable characteristics of the plasma in the ionosphere is the plasma frequency. The plasma frequency, caused by oscillations in the plasma due to electric fields, changes daily due to fluctuations that occur in the number density of particles.

$$\omega_p = \sqrt{\frac{ne^2}{m\epsilon_0}} \quad \text{eq. (16)}$$

The equation for plasma frequency, as seen in [18, eq. (16)], is a function of the number density n, the charge of an electron e, the mass of an electron m, and the permittivity of free space ϵ_0. This equation assumes that the plasma is a cold plasma. That is, the thermal motion of the electrons is insignificant. For warmer plasmas, the thermal motion must be taken into consideration. This frequency is important as it will determine how a radio wave of a given frequency would interact with it.

$$n_r = \sqrt{1 - \left(\frac{\omega_p}{\omega}\right)^2} \quad \text{eq. (17)}$$

The plasma frequency can be compared to the frequency of a given radio wave to determine what is known as the refractive index, as seen in [18, eq. (17)]. The refractive index n_r is a function of the plasma frequency ω_p and the wave frequency ω. The refractive index function give three distinct outcomes based on the plasma and wave frequency. If the refractive index n_r is imaginary, the radio wave will not propagate through the plasma and be absorbed. Mathematically, this means that the plasma frequency ω_p was greater than the radio wave frequency ω. If the refractive index n_r is zero, the radio wave will resonate with the plasma and refract. This will occur when the plasma frequency ω_p is equal to the radio wave frequency ω. The last case is when the refractive index n_r is positive. A positive refractive index n_r means that the radio wave frequency ω is greater than the plasma frequency ω_p and therefore continues to propagate through the plasma. The ability to reflect a radio wave off the ionosphere allows signals to travel over the horizon, what Marconi was able to demonstrate back in 1901. If the radio wave is directed vertically, the reflection off the ionosphere plasma can give a measure of height.

Ionosondes

The foundational idea of an ionosonde is that of a vertically oriented high frequency (HF) radar transmitting radio waves into the ionosphere and subsequently receiving them moments later to determine height measurements. First developed by Briet and Tuve in 1926, the ionosonde has been a staple instrument in making height measurements of the ionosphere. The height measurements derived from radio wave return times are not actual height measurements, but rather, as seen in [20, eq. (18)], a measure of virtual height.

$$h_{virtual} = \frac{c\Delta t}{2} \qquad \text{eq. (18)}$$

When the radio wave frequency interacts with a plasma of similar frequency, the plasma begins to act as a refracting medium. This will cause a group delay effect on the radio wave resulting in a longer return time. The height measurement as a result of this longer return time causes the measured height to be greater than what it is in actuality. Nevertheless, virtual height measurements are useful in determining ionosphere plasma structure and response to plasma changes. The various layers of the ionosphere are made up of different ion concentrations as well as electron density. This makes each layer unique in terms of plasma frequency. Shifting the ionosonde output frequency allows for a height map of the ionosphere called an ionogram, illustrated in [20, Fig. 9].

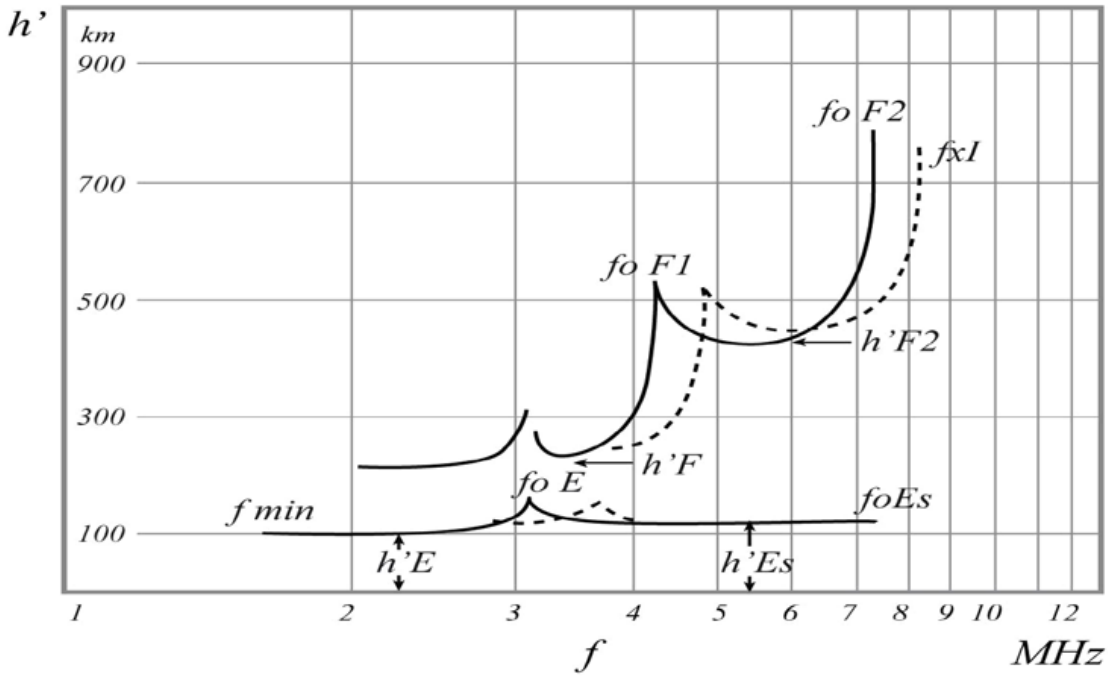

Fig 9: Ionogram [9]

There are several key features to the ionogram that allow for ascertaining specific ionosphere characteristics. The "J" line shapes are indicative of an ionosphere layer. The lowest point of the "J" reveals the virtual height of the layer. The vertical asymptotic line

that could be drawn for the “J” lines indicates the critical frequency of the layer. The critical frequency is the frequency needed to propagate through a layer of the ionosphere without significant absorption and refraction. The ionogram illustrated in [20, Fig. 9] displays three layers, the E, F1, and F2 layers, as well as an Es and dotted line. Es stands for sporadic E. Sporadic E are clouds of ionization that form just above the E layer and most commonly occur during summer months. These clouds can form from meteors that burn up in the ionosphere are spread due to wind shear, but also from electrical storms which indicate solar activity being a component of sporadic E. The dotted lines in the example ionogram indicate the different wave modes an ionosonde can take. There are two modes for radio propagation through a magnetized plasma. The O-mode waves, which are the solid lines, are linearly polarized with the electric field being parallel to the magnetic field. The X-mode waves, seen as the dotted lines, are linearly polarized with the electric field being oblique to the magnetic field [9].

Antennas

Antennas have allowed communication across vast distances for many years. One of the first antenna experiments was Hertz in 1887 [10]. The system he built served to produce radio waves to investigate electromagnetic radiation. The next big advancement in antennas came from Marconi 1901 where his experiment proved transatlantic communication was possible and paved the way towards interest in potential conducting layers in the atmosphere [2]. Antenna design has come a long way since its inception and proves to be essential in everyday communication.

The antenna is a core part of a radio system, or rather an ionosonde. While a core part, an antenna can vary heavily based on design, operational needs, and structure.

However, the defining basis for something to be considered an antenna is that it can radiate and receive electromagnetic waves. Typically constructed with metal, ceramics can be used to make, what is known as, dielectric resonator antennas [10]. The versatility of antennas gives distinct advantages over wired systems with regards to mobility, wide coverage, and low loss. Loss is frequency dependent and actually favors waveguides over antennas in short distances. As the distance increases, the amount of loss generated by the waveguide starts to overtake the loss from an antenna [11]. Not only does this indicate that antennas are the ideal method for long distance communication, but the notion of wire installation over various terrain is not feasible.

The design of an antenna will vary depending on its use. A defining feature of an antenna is the radio band which defines a frequency range it will transmit and/or receive on. A radio band is a section of frequency grouped by powers of 10 as presented in [22, Tab. 1].

Frequency	Band	Wavelength
30 – 300 kHz	Low Frequency (LF)	10 – 1 km
0.3 – 3 MHz	Medium Frequency (MF)	1 – 0.1 km
3 – 30 MHz	High Frequency (HF)	100 – 10 m
30 – 300 MHz	Very High Frequency (VHF)	10 – 1 m
0.3 – 3 GHz	Ultra High Frequency (UHF)	1 – 0.1 m

Table 1: Radio Bands [10]

Each radio band has its own use cases and a distinct frequency range. These frequency ranges give associated wavelength measurements that can become significant in antenna design and structure. Antenna structure can come in many forms. The most notable antenna types are wire antennas, aperture antennas, and antenna arrays [10]. Each of these antenna types has its own unique design, structure, and use cases. Wire antennas are the simplest to construct out of the three and are some of the most commonly used ones with a wide range of application. Various types of wire antennas include dipoles, monopoles, loops, helices, Yagi-Uda, and log periodic [10]. Dipole antennas are the most structurally simple out of all the wire antennas. They consist of two wires connected at a feed point with the other points as two open ends as illustrated in [23, Fig. 10].

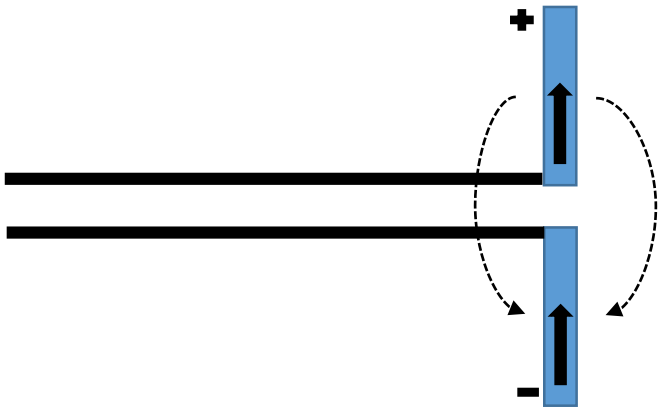

Fig 10: Dipole Antenna [10]

Further understanding of the dynamics of antennas requires solving the Maxwell equations. One of the simplest methods to solve the Maxwell equations is to represent them as a sum of plane waves [12]. The electric and magnetic fields propagate perpendicular to each other along some path. The direction vector that both these fields propagate along is called the Poynting vector. Because both the electric and magnetic fields are in phase with each other, their wavefronts form a plane wave [12]. The

interchange between the two waves and their properties allow for information to be stored. With this information, the electric field and magnetic field can be written as seen in [24, eq. (19 – 20)] respectively.

$$\boldsymbol{E} = E_0 \cos[\mathrm{j}(\omega t - kz)] \qquad \text{eq. (19) [12]}$$

$$\boldsymbol{H} = H_0 \cos[\mathrm{j}(\omega t - kz)] \qquad \text{eq. (20) [12]}$$

$$k = \frac{2\pi}{\lambda} \qquad \text{eq. (21) [12]}$$

The electric field and magnetic field are similar in structure as they both depend on their field magnitude, the angular frequency ω, the wavenumber k, and distance metric z. The wavenumber k, as seen in [24, eq. (21)] is a function of the wavelength λ. The Poynting vector can be equated from these two equations as seen in [24, eq. (22)].

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}^* \qquad \text{eq. (22) [12]}$$

One of the wave properties that define how a wave propagates is its polarization. The polarization of a wave is how the plane wave is positioned relative to the direction of propagation [12]. The polarization of a wave could be linear, circular, or elliptical. Linear polarization occurs when the wave is propagating only in a single plane. Circular polarization occurs when there are multiple linear components are perpendicular to each other in addition to being 90 degrees of phase difference. This configuration can take on a right-handed orientation or left-hand orientation depending on how the wave was generated. Elliptical polarization is essentially the same thing as circular polarization; however, the amplitudes of the wave components can differ as well as the phase difference being something other than 90 degrees.

The radiation pattern of an antenna is the region where the radiated power from the antenna is propagating in. There are two regions to the radiation pattern of an

antenna. The near field pattern and the far field pattern make up the radiation pattern. The boundaries between the two are determined by the directivity of the antenna and the wavelength of the signal being transmitted [10]. The directivity of an antenna is simply how much power is directed in a particular direction. Depending on the location of the receiver, the near field radiation can give a differing output to the far field radiation. With regards to long-distance communication, the far field pattern allows for approximations on how the signal will propagate and appear. The radiation pattern can change depending on the ratio between the dipole length and the wavelength of the signal. Out of all the ratios, the half-wavelength dipole is the most commonly used configuration [10]. Some of the reasons for this is that the directivity of the antenna configuration is relatively good, at about 2.15 dBi, the forward gain of the antenna. Additionally, the input impedance is immune to the changes in the radius of the antenna [10]. This is significant as the constant input impedance allows for ease of input matches with transmission lines.

Despite half-wavelength dipole antennas being the most common, other antenna configurations allow for varying radiation pattern and polarization. Loop antennas are another type of wire antennas that are similarly simple to construct. While generally constructed as circular, loop antennas can take various other shapes, such as ovals and rectangles, so long as the path closes at a point. The size of a loop antenna determines how various parameters can be solved for. Generally speaking, a loop antenna is considered small if the circumference is less than one third the wavelength of the signal [10]. Additionally, the current distribution can be considered constant within a small loop which indicates that the radiated fields will be identical. The radiation pattern for a small loop antenna is similar to that of a dipole antenna, as illustrated in [26, Fig. 11]. If the

loop length cannot be considered small, the current distribution within the loop subsequently cannot be considered constant [10]. A non-constant current distribution within the loop will cause a change in the directivity of the radiation pattern from parallel to the plane of the loop to perpendicular to the plane of the loop. Additionally, many antenna parameters will be changed as small loop simplifications will not work. Loops where the circumference is equal to the wavelength of a signal are called resonant loops. Resonant loops can be approximated as two half wave dipoles separated by twice the radius to simplify structure [10]. In essence, loop antennas can give a variety of radiation patterns depending on if the antenna length is considered small or not compared to the wavelength of the signal.

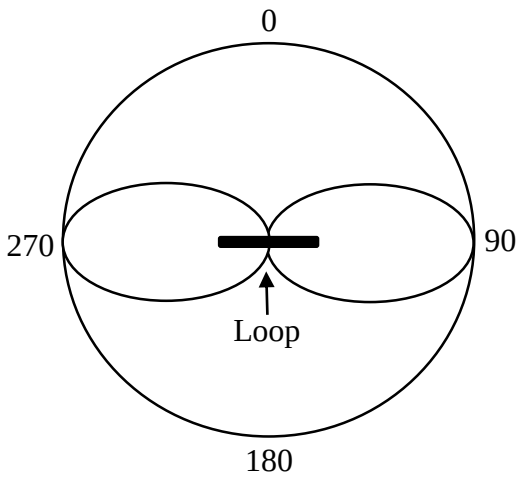

Figure 11: Small Loop Radiation Pattern

Helical antennas are another configuration of the wire antenna family. As the name suggests, the structure of a helical antenna takes the form of a spiral as illustrated in [27, Fig. 12]. The main characteristics of a helical antenna are the diameter, the spacing between the turns, the number of turns, the circumference of the spiral, the pitch angle of

each spiral, and the length of a single spiral. Based on how the antenna is configured, there are two modes of operation the helical antenna can perform at. The first operational mode is called the normal mode helix where the diameter is much smaller than the wavelength of the signal. In this mode, the radiation pattern is similar to that of the small loop and dipole antennas. The radiated field aligns itself perpendicular to the direction of the spiral. Under these conditions, the helix can be considered as a combination of a small loop antenna and a dipole antenna for each turn [10].

Fig 12: Helical Antenna Structure

The second operation mode is called the axial mode helix. This mode of operation is characterized by the circumference of the helix being equal to the wavelength of the signal. This changes how the radiation pattern appears significantly. Instead of radiating perpendicular to the direction of the antenna with two lobes, the radiation pattern in the axial mode radiates in a single main lobe in the direction of the spiral, as illustrated in

[28, Fig. 13]. The axial mode for the helical antenna allows for directionality in operation. A metal plate is typically placed at the base of the antenna to reflect any waves forward as directional antennas only want radiation in a single direction. The ease of construction and radiation pattern of the axial mode makes the helical antenna a popular choice for circularly polarized broadband antennas in the VHF and UHF bands [10].

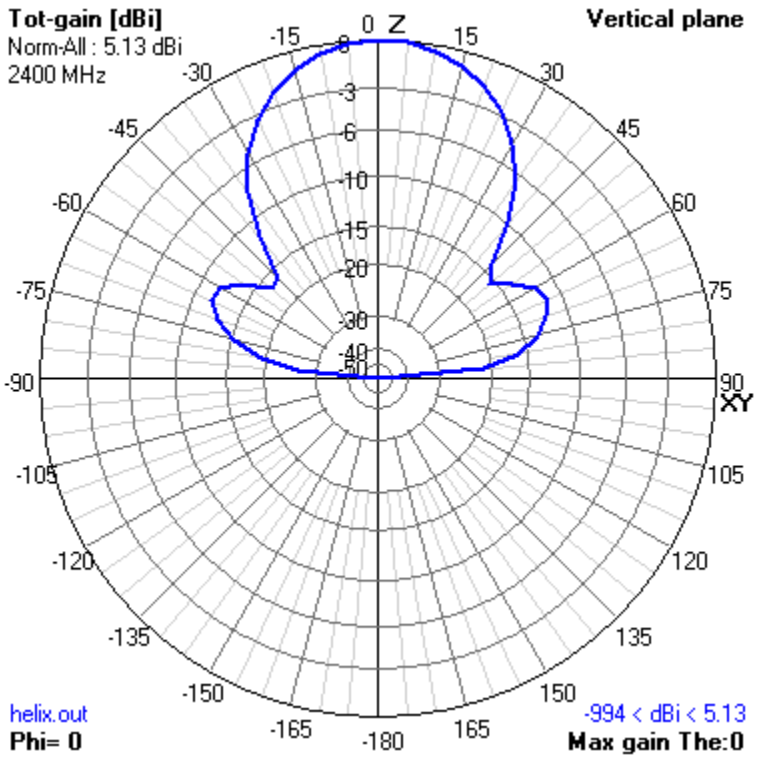

Fig 13: Axial Mode Helix Radiation Pattern

Yagi-Uda Antennas are another type of wire antenna that operates as an end-fire antenna. The main characteristic of an end-fire antenna is that the maximum radiation field is in the direction of the antenna [10]. Similar to the axial-mode helical antenna, the Yagi-Uda antenna radiates parallel to the direction of the antenna array. The Yagi-Uda antenna configuration is illustrated in [29, Fig. 14]. This antenna was first developed and studied in the 1920s by professors Yagi and Uda [10]. Significant work has been done in

the further development of the antenna and has since become popular for the VHF and UHF bands due to relative simplicity and high gain [10].

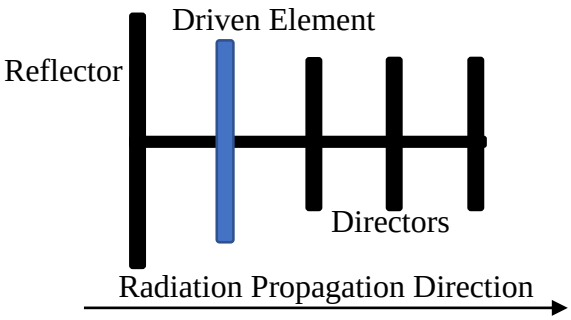

Fig 14: Yagi-Uda Antenna Configuration

There are three main components to the Yagi-Uda antenna. The reflector, which is typically the longest in length compared to the other two components, acts to direct any radiation forward along the antenna axis. The driven element is where the applied power is radiated outwards from a source. This dipole is normally a half wave dipole as the radiation pattern is in line with the direction of the antenna. The directors are parallel parasitic elements that, as the name suggests, direct the signal forward [10]. These elements act to reradiate the signal from the driven element forwards while adding some amount of gain per element. The resultant output of the Yagi-Uda antenna is an end-fire radiation pattern.

While wire antennas make up a good percentage of all antennas, they are not the only type out there. Aperture antennas are another antenna group characterized by the metal plates that form the configurations on how these antennas radiate power [10]. One of the most common aperture antennas are the horn antennas. The shape of the horn can take many forms that influence the antenna gain and radiation pattern. More significantly, the shape of the horn can also dictate the polarization of the traveling wave. Pyramidal

horns, in trapezoidal shapes, open up in both the E and H planes. This allows for variation in both the E and H plane radiation patterns. Furthermore, the pyramidal flared design, usually paired with rectangular waveguides, allows for linear polarization in some direction depending on the sides of the horn. The wave will be linearly polarized in the direction of the E-field should the horn be more vertical than horizontal. Likewise, the wave will be linearly polarized in the direction of the H-field should the horn be more horizontal than vertical [10]. The horn antenna can also be shaped like a cone. Used with cylindrical waveguides, these cone horns produce circularly polarized waves. All horn antennas are directed antennas as the radiation pattern produced shows a main lobe in the direction of the horn.

2017 Total Solar Eclipse

A solar eclipse is a natural phenomenon caused by the moon blocking the sun's path and subsequently enshrouding a path along the earth in shadow. This alignment of the moon, sun, and earth does not happen in each moon cycle. The orbit of the moon around the earth is not perfectly circular nor equatorial. The elliptical orbit the moon has is responsible for the change in tides and the angled orbit causes solar eclipses to not a be a common occurrence. The deviation is around five degrees and significant enough to give rarity to solar eclipses. When a solar eclipse does occur, a rare glimpse of the sun's corona can be seen with protective eyewear, as illustrated in [31, Fig. 15].

Fig 15: Total Solar Eclipse, photo credit: NASA/Aubrey Gemignani from https://www.nasa.gov/image-feature/2017-total-solar-eclipse/

The solar eclipse in August 2017 was a total solar eclipse. A total solar eclipse is a consequence of the elliptical orbit of the moon around the earth. The distance the moon is from the earth can vary which can, on occasion, enter a sweet spot where the moon fully blocks out the sun in the sky. Should the moon be far from the earth, the moon will only partially block the sun in what is called an annular solar eclipse. The shadow of the moon on the earth has two parts. The umbra is the region where the sun is completely blocked by the sun and is known as totality. The penumbra is the region where the moon does not fully overlap the sun and a partial eclipse can be seen. The path of totality for the August 2017 total solar eclipse is illustrated in [32, Fig. 16]. The path of totality begins in Oregon and finish in South Carolina. The convenience of the path provides a great opportunity for viewing and scientific work.

Fig 16: Total Solar Eclipse Path, image credit: Eclipse map/figure/table/predictions courtesy of Fred Espenak, NASA/Goddard Space Flight Center, from eclipse.gsfc.nasa.gov.

Experiment Design

The ionosonde project aimed to investigate how the height and composition of the ionosphere would change during the August 2017 total solar eclipse. The most notable changes that occur in the ionosphere is the day-night cycle. The dayside ionosphere is subjected to EUV causing photoionization that builds up the ion and electron densities. These densities and production rates are significant enough to cause four distinct layers, the D, E, F1, and F2 layers. The night side ionosphere is no longer bombarded from solar radiation. As such, the ions and free electrons begin to form to create neutral atoms and molecules at a comparatively higher rate than on the day side. The D layer dissipates due to particle recombination and the F1 and F2 layers merge leaving on the E layer and the newly formed F layer. A total solar eclipse allows for a night side ionospheric condition during the day side. The umbra of the solar eclipse will block sunlight from interacting with the particles in the ionosphere. This should cause the ionosphere to behave as though it was the night. In order to measure the height of the ionosphere in the path of the eclipse, a method of signal generation, an antenna to transmit and receive such signal and is portable, and a method by which to evaluate the return signal to derive the height must all be designed.

The changing height of the ionosphere during an eclipse has not been the subject of much experimentation nor research. It is well documented that signals of a given frequency will interact with the ionosphere differently depending on the current plasma frequency. The plasma frequency is directly related to the particle density of the plasma which in turn gives relation to sources of production and loss of ionization. A radio wave will either refract, propagate through, or be absorbed depending on the plasma frequency

in relation to the wave frequency. The change in plasma frequency from day to night, while still subject to constant change, is consistent enough in that of being low during the night and high during the day to allow for predictability. Surely enough though, outside forces such as cosmic rays, coronal mass ejections (CME), or geomagnetic storms can change the ionosphere in unpredictable ways. However, such events are outside the realm of control and reasonable predictability. Solar eclipses, on the other hand, are very predictable in their arrival, path, and type. The changes that the ionosphere will experience during an eclipse is something that can be looked into and become potentially predictable.

Design of the ionosonde is key to the project. The software used was GNU Radio as this SDR is extremely versatile in signal design as well as capable of integrating with various radios to act as transmitters and receivers. A USRP N210 was used to act as the transmitter and receiver for the project. The reason for having both transmitter and receiver on the same device is to avoid synchronization issues that can occur with multiple devices. Additionally, the overall design of the ionosonde is less complex with only a single device. GNU Radio has USRP specific blocks that allow for connections between the two over Ethernet cable. The process of transmitting and receiving in GNU Radio is relatively straightforward as the USRP sink and source blocks do most of the work. The transmitting portion of the GNU Radio flowgraph is essentially a signal source connected to a throttle block. The throttle block limits the data rate to the set sample rate and not the computer hardware capabilities, and then connected to the USRP sink block. A mere three blocks are what the transmission portion of the ionosonde flowgraph is made of as illustrated in [35, Fig. 17].

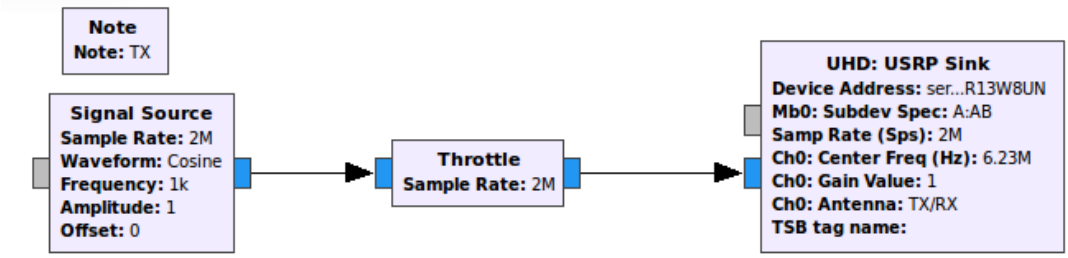

Fig 17: GNU Radio Transmitter

The receiving end of the flowgraph is even simpler. The USRP source block is effectively the whole receiving portion as illustrated in [35, Fig. 18]. That block is connected to a file sink block which saves the output for later use. Of course, the received return signal will need to have additional processes done to it, and that is discussed later on.

Fig 18: GNU Radio Reciever

Typically seen in ionograms are the wide range of frequencies to cover the HF spectrum. These give a good snapshot of the ionosphere. However, this project aims to see how the ionospheric height will change with time during the eclipse. It should be noted that in order to send signals into the sky, an FCC radio license is needed. The project group obtained one which allowed for transmissions within the HF band but with restrictions on the power level. The ionogram of interest will have time on the x-axis with

height still on the y-axis. The transmitted signal to be designed will need to reflect this. The decision was made to have a short frequency sweep on four discrete frequencies. This allows for the different height parameters to be seen as the frequency of each signal will penetrate to different layers. The basis for making height measurements on the ionosphere is to have a radio wave capable of reflecting off the various layers. The design of a unique signal that satisfies these parameters must be done.

The signal design depends on how the transmitting and receiving functions will operate. Based on what the goal of this project is, there are several aspects to the ionosonde that are known. Height measurements of the ionosphere are done by calculating from the return time of the signal which is inferred by the distance from the transmission peak to the return peak. In order to have the transmitted signal appear in the return file, both the transmitter and the receiver would stay on throughout the eclipse. The idea is that when the transmitter transmits a radio wave, the receiver will immediately pick that signal up as a large pulse of energy. Sometime later, the receiver will pick up the return time, which will be smaller in amplitude due to absorption and energy loss in transit. With both the transmitted signal and the return signal in the same output file, the return time of the signal becomes trivial to identify. With all this in mind, the transmitted signal must be designed around this whole process.

The current signal sources in native GNU Radio repository are not easily capable of producing the signal pulse desired. A user-supplied module was introduced to assist with that. The block used is called Signal Generator FMCW (Frequency Modulated Continuous-Wave). While the decision to have the transmission on four discrete frequencies, the notion to have a frequency sweep was still desired. The new signal

source block assists with that. The use of a continuous wave signal ensures that there is a constant energy output, which is necessary to compensate for the absorption the ionosphere will have on the signal [13]. One of the main disadvantages of using CW is that the signal will only provide information about the target in one direction [14]. However, this is not a problem for this application due to the nature of an ionosonde which only sends a signal vertically upwards. The discrete frequencies to transmit at were decided to be 3, 4, 5, and 6 MHz and have a frequency sweep of 100 kHz at each location. The other important factor of the FMCW block is the samples up-chirp parameter. The FMCW block will output a signal that will currently sweep in frequency 100 kHz. The samples up-chirp parameter gives control on how many samples are used to complete that sweep. A per sample basis operation allows for control over timing based on the overall flowgraph sample rate. Determining how many samples to up-chirp on comes down to a simple equation of the sample rate multiplied by the desired up-chirp time. The chirp time of the signal was determined to be 500 μs.

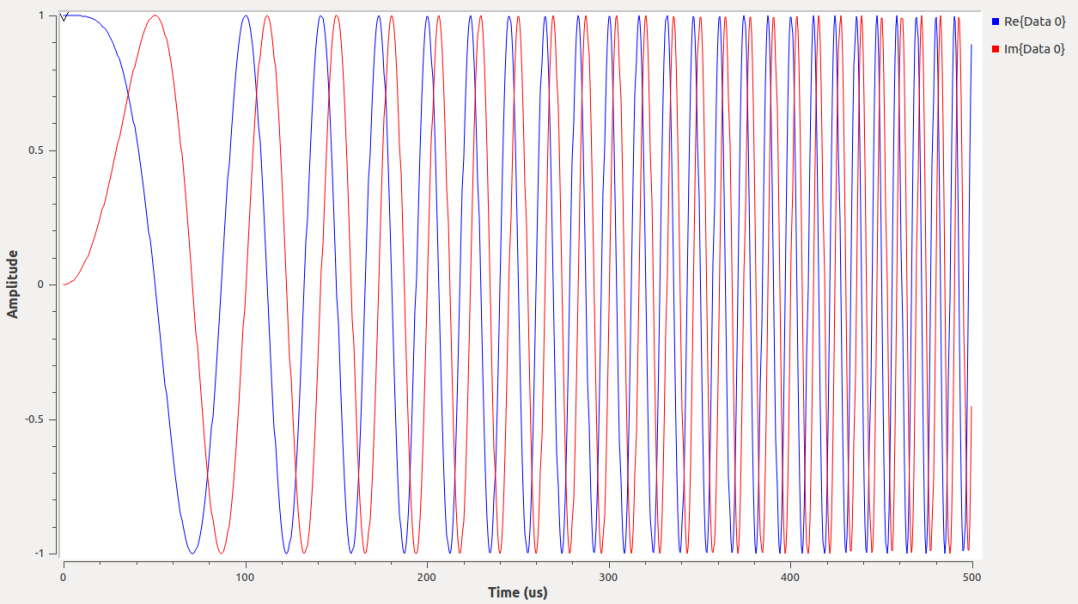

Fig 19: FMCW Signal Pulse

This translates to a short up-chirp of 1 k samples at the flowgraph sample rate of 2 M samples per second as illustrated in [37, Fig. 19].

It is important to recall that the transmitter and receiver will both be continuously on. Transmitting the signal pulse as it is would cause the receiver to only see the transmitted pulse and drown out any potential return signals. Therefore, a duty cycle to the signal pulse needs to be implemented. Based on the height of the ionosphere and the speed of the transmitted pulse, a ratio of on time to off time was implemented. The on and off times were kept in multiples of the up-chirp time in terms of samples. The on-time was selected to be 1 k samples and the off time was selected to be 5 k samples. This ratio comes out to be a 16.67 % duty cycle as illustrated in [38, Fig. 20].

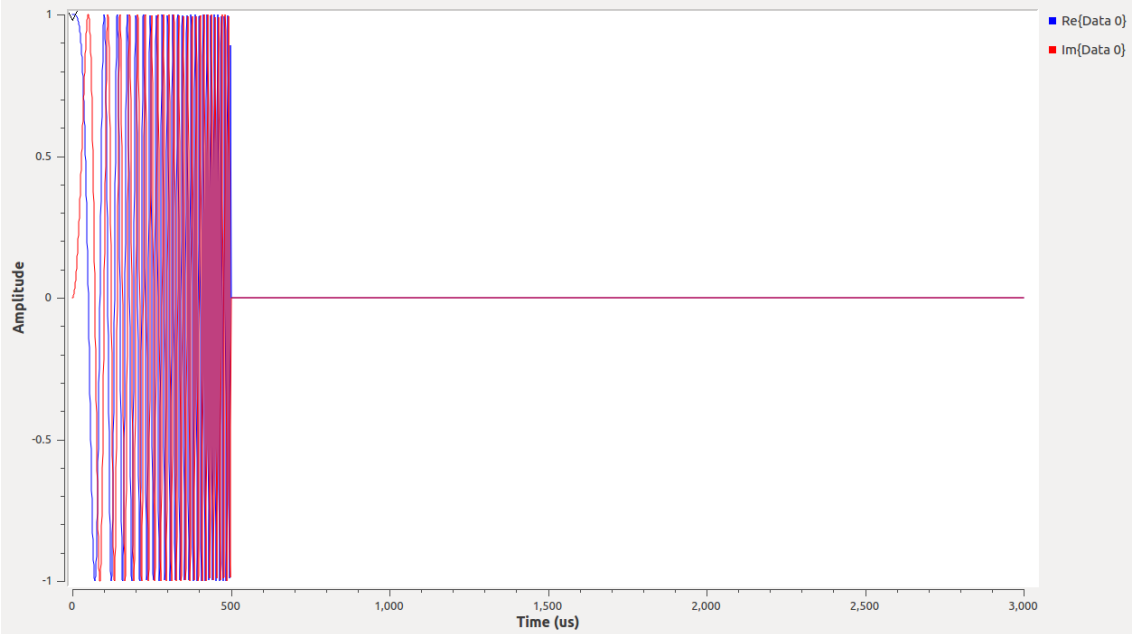

Fig 20: Total Signal Pulse

This total signal was saved to a file for later transmission. The total signal, however, was only saved as a single period. The single pulse period is read in and

repeated over to ensure signal fidelity as well as a later point of exact comparison. The signal would be transmitted at 1 watt due to the restrictions on transmission.

The end goal of the ionosonde project has been to measure the height of the ionosphere during the eclipse. In the ionosonde design process, a sub-goal of designing a more portable and cost-saving ionosonde was put into place. This involved the use of an alternative radio to the USRP N210. This can in the form of the Red Pitaya. The Red Pitaya offered a similar performance characteristic to the USRP N210, but at a much lower price and smaller size. Additionally, the frequency ranges at which the Red Pitaya could operate was not limited by hardware. The USRP N210 required a specific daughter card in order to operate in the HF region. However, the Red Pitaya was not as straightforward to operate with as the essentially plug-and-use USRP. The Red Pitaya did contain an Ethernet port for connection to a host machine, much like the USRP. A common issue that SDR devices have is that the throughput of each device is limited by the data rate on the connection between the two as well as the processing power of the host machine [15]. Data rate issues are most notably seen when connections are made using a universal serial bus connector (USB). The Ethernet port allows for a better connection than USB and is fast enough to alleviate most data rate issues.

The Red Pitaya board make use of a micro-SD card in order to image software on to use. This was necessary as compatibility with GNU Radio was not trivial. This process allowed the Red Pitaya to load GNU Radio on the board and perform the ionosonde flowgraph inside, rather than act as a peripheral. Another approach would be to use the OOT blocks that were designed to operate specifically with the Red Pitaya. The Red Pitaya source and sink blocks allowed functionality with the board, however, this did not

allow for signal testing. The USRP has been widely used and the operational characteristics, such as the noise floor level, are well known. The Red Pitaya, on the other hand, is relatively new and not as known as the USRP. The reason for the inability to test operational parameters on the Red Pitaya using GNU Radio is that the measurement ticks for signal levels do not mean anything as illustrated in [40, Fig. 21]. The gain measurements for signals in GNU Radio are only useful in seeing the relative difference between a peak and the apparent noise floor. A function generator must be used, where the gain of the signal can be manually set, to give a point of reference. The noise floor for the Red Pitaya turned out to be extremely poor. The noise floor was about roughly 40 dBm greater than the USRP N210. The noise figure for the Red Pitaya would not allow for returns to be seen due to the attenuation the ionosphere will apply as well as the power restriction from the FCC license.

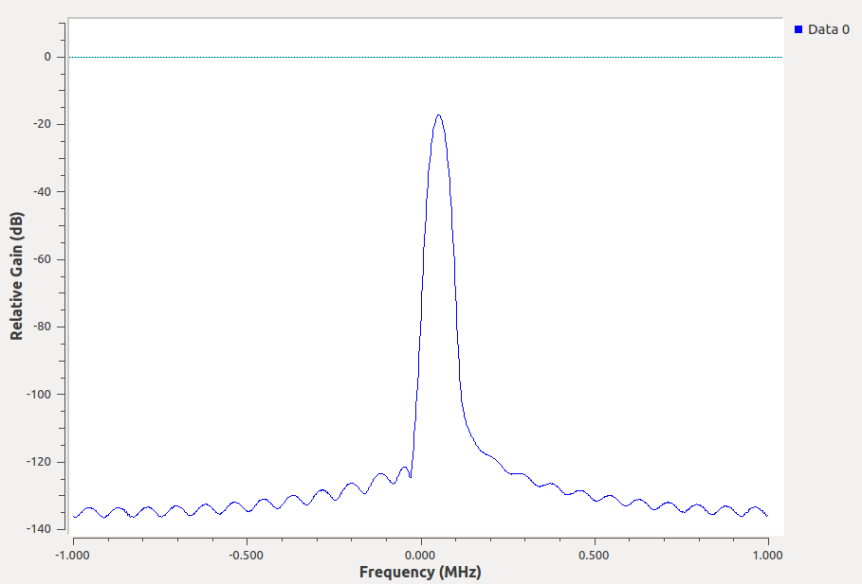

Fig 21: GNU Radio Gain

The outcome here was simply to forgo the more portable and affordable ionosonde route and focus on integrating the USRP N210 into the system.

The current ionosonde takes the form of a GNU Radio flowgraph connected to a USRP N210 external radio. Completion of the ionosonde requires the design and construction of antennas. Antenna design can take numerous forms as there are many configurations. However, the hard deadline of August 21, 2017, reinforced the notion towards simple antenna design. Additionally, the frequency band that transmission would take place on would be the HF band. With this in mind, the goal is to direct a signal skyward and subsequently receive the return. Dipole antennas have great directivity in their radiation pattern, in addition to being structurally simple to construct, which is crucial for ionospheric reflection. This lead to the design of a half-wave dipole antenna.

The transmit frequencies to use were 3, 4, 5, and 6 MHz. The dipole antenna would have four parallel wires of the corresponding length separated with spacers and hoisted in an inverted vee configuration. The inverted vee antenna is simply a dipole antenna but with the arms directed diagonally downwards in an upside-down "V" shape. Not only does this configuration relieve the structural burden by not requiring similarly tall side masts as the center pole with a taut rope, but the upwards directivity is increased in addition to improved HF band performance. While the frequencies to use were set to be 3, 4, 5, and 6 MHz, these frequencies are more akin to guidelines and approximations. The most important aspect of the antennas was that each was built uniformly. The length of each wire for each transmit frequency could be approximated in feet using the velocity factor as seen in [41, eq. (23)].

$$Length = \frac{468}{f(MHz)} \qquad \text{eq. (23)}$$

Spools of wire were measured out individually and cut accordingly. As there were to be both a transmit and receive antenna, two sets of each frequency were cut. The actual

length that was cut was several inches longer as cutting wire is easier than adding wire and to help compensate for later wire maneuvers. The wire lengths were very long as the frequency range in use is on the low side of the HF band. Spacers in the form of polyvinyl chloride (PVC) pipes were implemented to prevent wire tangling. Each wire was wrapped once around each of the PVC spacers in order to prevent slipping and tangling. This subsequently makes each wire length electrically shorter by the circumference of each pipe, however, the increased length of the original cut helps to alleviate this shortening. Each cut wire length was cut in half for each side of the inverted vee antenna arm. The antenna wire configuration is illustrated in [42, Fig. 22].

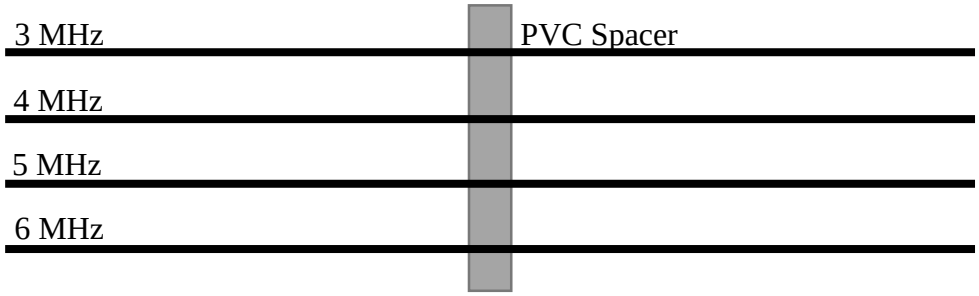

Fig 22: Inverted Vee Antenna Wire Configuration with Spacer

The arms of the antenna were joined together at a center feed point which ran a coaxial cable from the antenna to the USRP N210. This completed the structural component of the transmitter side of the ionosonde.

The wires were loosely cut in order to finely tune the antenna upon completion. A voltage standing wave ratio (VSWR) meter allows for the measurement on how well the antenna input impendence is matched to the wires it is connected to. The value the meter reads is based on the equation seen in [42, eq. (24)].

$$VSWR = \frac{1+|\Gamma|}{1-|\Gamma|} \qquad \text{eq. (24)}$$

VSWR is a ratio of the reflection coefficient Γ in the transmission wires. Values close to 1 indicate that the lines are well matched and that the reflected power back from the lines is minimal. Each line is cut evenly on both sides until the VSWR meter shows a value close to the target frequency. At this point, the VSWR meter is now used to find where around the target frequency does the best transmission come from. The final frequency values are presented in [43, Tab. 2].

Target Frequency (MHz)	**Actual Frequency (MHz)**
3	3.15
4	3.93
5	5.18
6	6.28

Table 2: Antenna Frequencies

The lengths of each wire were noted and subsequent antennas were constructed. A receiver antenna with the same wire lengths was constructed and thus the ionosonde has been fully constructed.

The transmitter and receiver antennas were placed close to each other as the indicator on the receiver side for the transmitted signal required instantly receiving the signal at the start. The transmitted signals were linearly polarized as working with the constructed antennas allowed for that. Additionally, a known polarization, as well as signal structure, ensures that in post-processing the detected signal on the receive side is, in fact, the same transmitted signal and not noise or some stray signal. However, it was well known that linear polarization was not the optimal choice regarding signal polarization. Circular polarization would have been the better choice simply due to the

nature of the environment the transmitted signals propagating towards. It is well known that when a signal refracts off the ionosphere, at a given frequency, the signal will be attenuated by some amount and continue attenuation while propagating. Linear polarization is not robust enough to resist significant attenuation from the ionosphere. Circular polarization is capable of resisting significant signal degradation from the ionosphere. The reason is that the ionosphere, as well as other atmospheric conditions, can rotate the signal which will cause attenuation on the receive antenna both are not aligned. This affects linear polarization more than circular polarization. Additionally, HF attenuation is also more pronounced in the presence of rain. Nevertheless, linear polarization was used not only because there was high confidence a return signal would be detected at the power output in use, but also due to the hard time limit of the total solar eclipse which strained further antenna design and construction. The transmit and receive antennas were placed in a bi-static configuration as well as cross-polarized from each other to minimize coupling that can occur. The complete antenna configuration is illustrated in [44, Fig. 23].

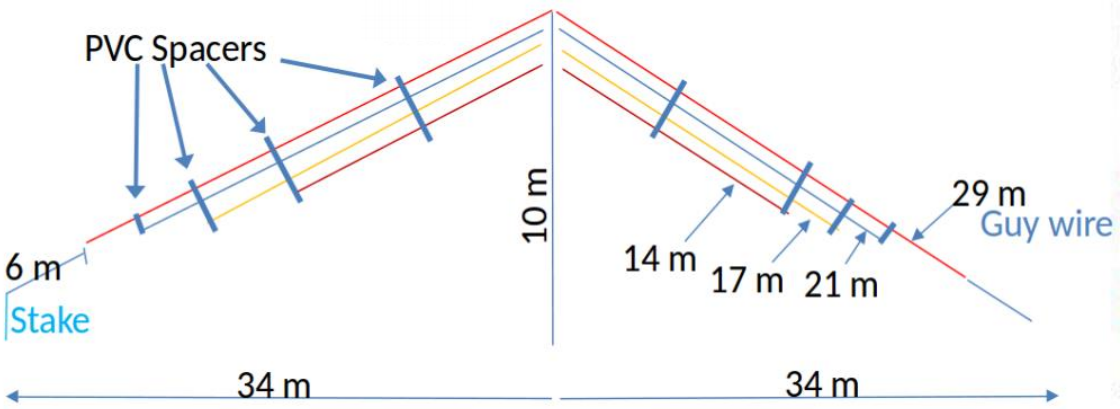

Fig 23: Inverted Vee Antenna Configuration [16].

With both a transmit and receive antenna properly configured and constructed, the ability to test the capabilities of the ionosonde is available. At the time of testing, the sun was in the declining phase of solar maximum as well as it being summer. The total solar eclipse would present a night side condition in the ionosphere so testing was typically done at dusk to emulate this change from a day side condition to a night side condition. It quickly became apparent that the return signals were simply too attenuated from the reflections to be seen on the receiver. In order to compensate for the lack of a return signal, external devices were added in series to the ionosonde transmit and receive channels as well as an improved software processing method. On the transmit side, a power amplifier and a low pass filter were added in series between the transmit antenna and the USRP N210 transmit output. The power amplifier was a TAPR PennyWhistle amplifier. This provided the transmit signal a 17-19 dB gain [16] which would increase the likelihood that the attenuation from the ionosphere would not completely reduce the signal. The low pass filter acts to suppress signal harmonics in order to prevent waveform distortion. Limiting waveform distortion will increase the likelihood of successful detection in post-processing. On the receive side, a low noise amplifier and a band pass filter are used in series between the receive antenna and the receive port of the USRP N210. The low noise amplifier serves to increase the signal gain by about 18 dB [16] to help detection. The band pass filter in conjunction with the low noise amplifier serves to suppress the unwanted noise around the return signal.

The new hardware implementations to the ionosonde will make the return signal much more detectable. In order to translate the peak-to-peak time difference between the transmitted and return signal, a receiving side software method implementation is needed.

Cross-correlation will simultaneously bring the return signal out of the noise and lower the noise. From signal generation, it is already known what the transmitted signal is structured as and is already saved with GNU Radio. The software processing boils down to using the generated transmit signal and cross-correlating it with the return stream. This process is illustrated in [46, Fig. 24].

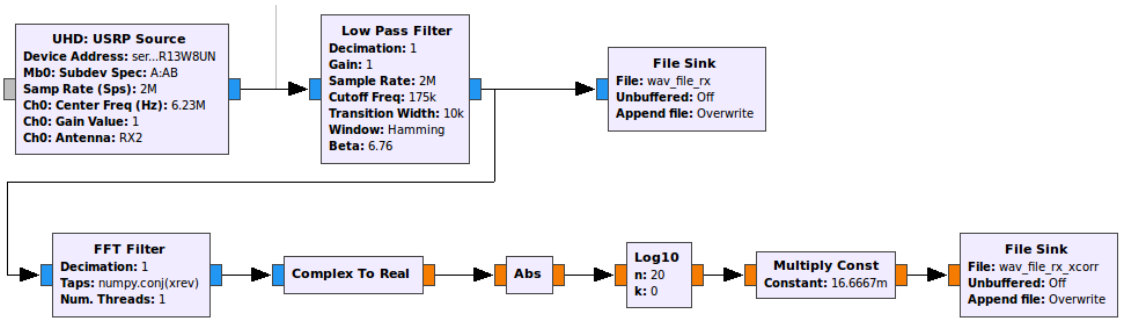

Fig 24: Software Cross-Correlation

The implementation of a cross-correlation method in GNU Radio involved filtering the originally transmitted signal against the return stream. The USRP source block outputs an IQ stream of return data from the receive antenna. A low pass filter block helps further suppress noise that could affect the cross-correlation process. The data stream is filtered against a reversed and complex conjugated transmitted signal to lift the return signal out of the noise. The rest of the blocks serve to transition the complex signal to a real signal and display power. This allows for the transmit peak and return peak to be visible for timing measurements.

Results

Preparation for the total solar eclipse had individuals set up the ionosondes at various points along the eclipse path. Locations in Oregon, Kansas, and South Carolina were selected along the eclipse path to serve set up the ionosondes. In all, three sets of transmitting and receiving antennas were constructed. Additionally, the ground station at

Blacksburg, Virginia was used for gathering more data and serve as a point of comparison, despite not being along the eclipse path. Ionograms of the total solar eclipse are illustrated in [47, Fig. 25]. The vertical dotted lines on the left side indicate the start, maximum point, and end of the eclipse. The horizontal dotted lines on the left indicate an estimation on the ionosphere altitude would normally be. The rate at which each the ionosonde transmitted signals gave a tremendous amount of data points for the virtual height of the ionosphere. The produced ionograms did not use all of the raw data points collected during the total solar eclipse. Rather, averages were made over many data points to eliminate any outlying data points as well as to lower the noise in the resulting ionograms. This post-processing method can be done as the strong law of large numbers dictate that over many iterations, the virtual height data points, a convergence will appear with a large enough sample size. With enough data points averaged over for each increment of time, a convergence of a virtual height is achieved with confidence.

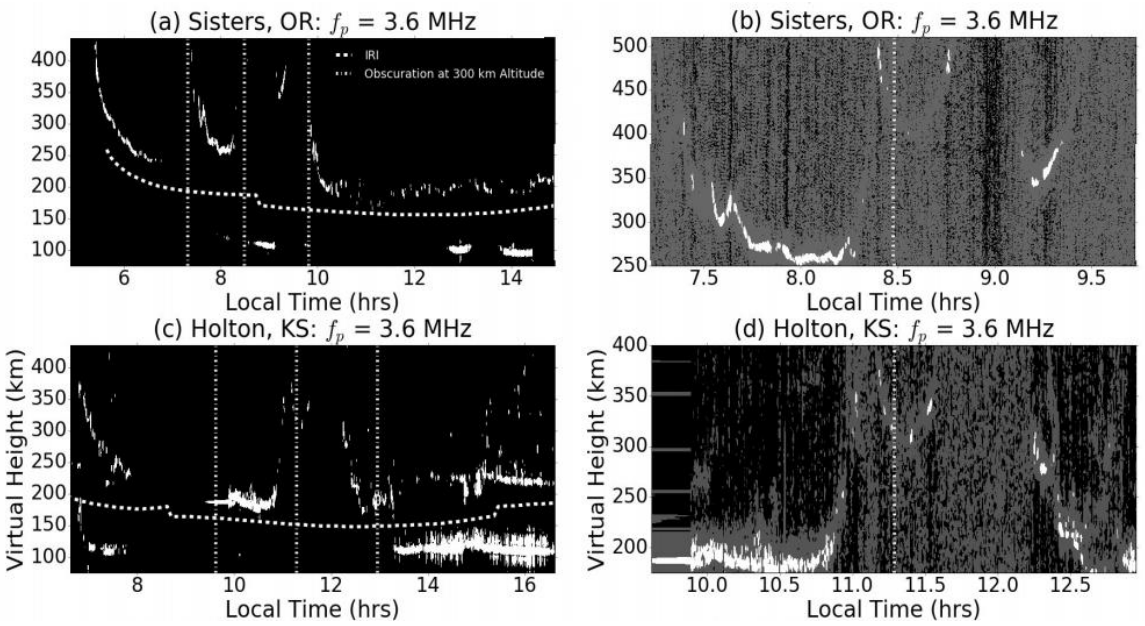

Figure 25: Ionograms of Eclipse [16]

The ionograms on the left are the whole ionosonde capture while the ionograms on the right are focused in on the maximum point of the eclipse. The F region plasma of the ionosphere is greatly affected by the total solar eclipse as seen on the right side ionograms. The fixed frequency shows that the plasma frequency of the ionosphere abruptly decreases, allowing the radio wave to penetrate further upwards. The virtual height of the ionosphere during the onset of the eclipse in both the Kansas and Oregon locations indicate different upward slopes. This could possibly indicate that as the total solar eclipse was forming, plasma from nearby regions was funneling in to recover the night side condition. There are additionally several patches of low altitude lines that could be indicative of sporadic E.

This project, which began and ended in the months of summer, was the first time where the design of something was realized and produced actual results from the field. Past projects were simply to design something and present its capabilities. It was not until now that a project design was actually implemented with physical components, in the form of antennas to transmit and receive, and used out in the field for real-life purposes. This whole process truly left a big impression on just how satisfying building something up to test a real-world event could be like. Despite the process of building an ionosonde and sending signals up into the ionosphere being relatively straightforward on the surface, the actual implementation and design considerations that went into making an ionosonde were beyond my expectations. Each time a version of the ionosonde came out and was discussed in the project meetings, new considerations and angles of approach were pointed out which caused many iterations of design and discuss to occur. This proved were rewarding as the final iteration of the ionosonde was a culmination of many

viewpoints and trains of thought. The ionosonde project, while straightforward in approach, proved to be more complex as time passed. Design choices and considerations based on implementation throughout of the project proved to further the understanding and appreciation of the design process that goes towards turning a scientific proposal into something physical and usable in the real world [17].

Future Work

The ionosondes developed for this project were aimed at uncovering how the ionosphere would change during a total solar eclipse. This imposed a hard time limit that influenced design choices. This is to say that further optimizations and changes in design could be done to significantly improve upon the current ionosonde design for future work.

One of the main points of constraint during the project was the antenna design. Half-wave dipoles were used with discrete frequencies as wire lengths along the antenna. Additionally, there were only a single transmit and receive antenna as part of the ionosonde. As a result, linear polarization was used, despite it not being optimal. Stated previously, linear polarization was easy to implement, especially with the limited antennas constructed, as well as having confidence that a return would happen. Improvements upon the antenna design would be regarding the polarization used. Rather than a single receiver and a single transmitter antenna, pairs of half-wave dipole antennas would be used as a transmitting and receiving antenna or other more complex designs. These would allow for a wider frequency coverage rather than a few discrete frequencies. This would further the study of ionospheric dynamics and frequencies where major changes occur. The antennas would be perpendicular to each other forming an “X” shape,

which would allow for circular and other types of polarization. Additionally, polarization configuration would also be done in software to further the study of the dynamics of polarization in the ionosphere.

The antenna itself would be a source of change for future ionosonde experiments. The transmit and receive antennas were only capable of operating on four distinct frequencies. These four frequencies were based on VSWR measurements and are approximately the original target frequencies. The transmitted signal is swept up by 100 kHz, which is important for ionospheric measurements, but will decrease the VSWR on the line it is transmitting from. A better solution to this would be the implementation of a broadband vertical incident antenna so that the transmit and receive capabilities are not limited to discrete frequencies. Furthermore, the use of non-resonant crossed receivers with active amplifications on each would not only be simpler to construct and transport, but also give the same gain as the original antennas. This would be significant as, despite the rarity of total solar eclipses, the accessibility of the eclipse path is not always easy to get to. The ability to improve construction and transportation of the antennas of the ionosonde will allow for greater opportunity to arrive at such eclipse locations.

Conclusion

The results of the ionograms gave insight into how the dayside ionosphere would behave when suddenly introduced to a night side ionosphere by a total solar eclipse. Given the several ionosonde locations and transmitting frequencies, the dynamics of the ionosphere absorption and reflection along the eclipse path are observed. Plasma transport in the upper F region of the ionosphere seemed to play a significant role in how the ionosphere reacted to the oncoming umbra of the total solar eclipse. The ionosonde

constructed was capable of transmitting a signal of various frequencies through the design of an inverted vee antenna. The inverted vee antenna allowed for the simplest construction for the desired transmit and receive frequencies in the time span prior to the total solar eclipse. The antennas were certainly not the best design choice for ionospheric measurements and study, especially with regards to signal polarization, however, the reliability in the antenna performance allowed for subsequent measurements to be taken with confidence. The whole design of the ionosonde signal output was flexible enough to allow for variation in wave characteristics as the cross-correlation method in combination with strong law of large numbers allowed for good detection on returns even if the returning signal is mismatched with the return antenna with regards to its polarization. Overall, the ionosonde design produced ionograms of the total solar eclipse which gave further knowledge into the behavior of the ionosphere and how the multiple layers respond to change.

Summary

The ionosphere is a constantly changing medium. A most notable change is the day to night cycle the ionosphere undergoes. A solar eclipse produces a nighttime condition during the day which will cause the ionosphere to behave in a potentially unusual manner. In anticipation for the August 2017 total solar eclipse, a team of people set out to construct an ionosonde to measure how the ionosphere would react. The ionosonde was based in GNU Radio and utilized custom made antennas to transmit and receive HF signals. Data collected from the total solar eclipse was converted into ionograms which gave a visual representation of how the ionosphere reacted.

References

[1] "Welcome to GNU Radio!," *GNU Radio Manual and C API Reference: Main Page*. [Online]. Available: https://www.gnuradio.org/doc/doxygen/index.html. [Accessed: 21-Mar-2019].

[2] J. B. H. Baker, "Lecture 2 History: Ionosphere Radiowaves," in *TS: Ionospheric Radar Tech*, 23-Aug-2018.

[3] T. F. Tascione, *Introduction to the space environment*. Malabar, Fa.: Orbit Book Co., 1988.

[4] J. B. H. Baker, "Lecture 8 The Ionosphere: Altitude Structure," in *TS: Ionospheric Radar Tech*, 25-Sep-2018.

[5] J. K. Hargreaves, *The Upper atmosphere and solar-terrestrial relations: an introduction to the aerospace environment*. New York: Van Nostrand Reinhold Co., 1979.

[6] J. B. H. Baker, "Lecture 4 Introduction: Basic Plasma Physics," in *TS: Ionospheric Radar Tech*, 4-Sep-2018.

[7] J. B. H. Baker, "Lecture 7 The Sun: Basic Structure," in *ECE-5146 Space Science I*, 13-Sep-2017.

[8] J. B. H. Baker, "Lecture 8: Solar Activity," in *ECE-5146 Space Science I*, 15-Sep-2017.

[9] J. B. H. Baker, "Lecture 11 Ionosondes: Basic Principles," in *TS: Ionospheric Radar Tech*, 9-Oct-2018.

[10] Y. Huang and K. Boyler, *Antennas: from theory to practice*. Chichester, U.K.: Wiley, 2008.

[11] T. A. Milligan, *Modern antenna design*. Hoboken, NJ: IEEE Press, 2005.

[12] S. R. Saunders and Aragón-Zavala Alejandro, *Antennas and propagation for wireless communication systems*. Chichester, England: J. Wiley & Sons, 2007.

[13] E. J. Amin, A. B. Suksmono, and A. Munir, "Accuracy analysis of FM chirp in GNU radio-based FMCW radar for multiple target detection," *2014 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 2014.

[14] Y. P. Saputera, D. Herdiana, H. Madinawati, A. B. Suksmono, and A. Munir, "Linear frequency modulated continuous wave radar using GNU radio and USRP," *2015 1st International Conference on Wireless and Telematics (ICWT)*, 2015.

[15] W. Song, "Configure Cognitive Radio using GNU Radio and USRP," *2009 3rd IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, 2009.

[16] L. J. Kordella, G. D. Earle, M. Moses, X. Han, D. Sardana, D. Sweeney, R. McGwier, and W. Lloyd, "Distributed High-Frequency Soudings of the Ionosphere During the 2017 Solar Eclipse." CEDAR, CEDAR, Jun-2018.

[17] R. McGwier and W. Lloyd, "Using GNURadio and Red Pitaya for Citizen Science," GRCON, tech., 2018.

Further Readings

1. Mahafza, Bassem R. “Radar systems analysis and design using Matlab”, © 2000 by Chapman & Hall/CRC
2. A. Wojtkiewicz, J. Misiurewicz, M. Nalecz, K. Jedrzejewski and K. Kulpa, “Two-dimensional signal processing in FMCW radars,” Politechnika Warszawska, Warszawa, 1996.
3. R. L. Smith, Micro Syntethic Aperture Radar Using FM/CW Technology, Brigham Young University, Master Thesis 2002.
4. A. Prabaswara, A. Munir, A. B. Suksmono, “GNU radio based software defined FMCW radar for weather sruveilance application”, in Proc. of 6th International Conf. on Telecommunication System Services and Application (TSSA), Denpasar, Bali, Oct. 2011
5. L. K. Patton, A GNU radio based software defined radar. Dept. of Elecrical Engineering, Wright State University, Thesis, 2007
6. M. Dillinger, K. Madani and N. Alonistioti, Software defined radio: architectures, System and Functions, Wiley and Sons, 2003
7. L. Patton, “A GNU Radio Based Software Defined Radar”. Master’s Thesis. Wright State University, 2007.
8. B. R. Mahafza, Radar Systems Analysis and Design Using MATLAB. CRC Press. 2009.
9. L. K. Patton, A GNU Radio Based SoftwareDefined Radar, Wright State University, Thesis, 2007

10. Bu-Chin Wang, “Digital Signal processing Techniques and Applications in Radar Imaging Processing”. Published by John Wiley & Sons, Inc., Hokoken, New Jersey, 2008.
11. M. Pilgrim, “Dive Into Python: Python from Novice to Pro”, May 4, 2004
12. Zhi Yan ,Zhangchao,Ma,etc,”Spectrum Sensing,Access and Coexistence Testbed for Cognitive Radio Using USRP”, Circuits and Systems for Communications, 2008. ICCSC 2008. 4th IEEE International Conference on 26-28 May 2008 Page(s):270 – 274