UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Faculty of Science and Technology
Department of Physics and Technology

Design and Implementation of a Software Defined Ionosonde

A contribution to the development of distributed arrays of small instruments

—

Markus Floer

FYS-3931: Master's thesis in space physics, June 2020

Design and Implementation of a Software Defined Ionosonde

Markus Floer

A thesis for the degree of Magister Scientiae

Department of Physics and Technology,
Faculty of Science and Technology,
University of Tromsø – The Arctic University of Norway

in cooperation with

Department of Arctic Geophysics,
The University Centre in Svalbard

June 2020

Abstract

In order to make advances in studies of mesoscale ionospheric phenomena, a new type of ionosonde is needed. This ionosonde should be relatively inexpensive and small form factor. It should also be well suited for operation in a network of transmit and receiver sites that are operated cooperatively in order to measure vertical and oblique paths between multiple transmitters and receivers in the network. No such ionosonde implementation currently exists. This thesis describes the design and implementation of a coded continuous wave ionosonde, which utilizes long pseudo-random transmit waveforms. Such radar waveforms have several advantages: they can be used at low peak power, they can be used in multi-static cooperative radar networks, they can be used to measure range-Doppler overspread targets, they are relatively robust against external interference, and they produce relatively low interference to other users that share the same portion of the electromagnetic spectrum. The new ionosonde design is thus well suited for use in ionosonde networks. The technical design relies on the software defined radio paradigm and the hardware design is based on commercially available inexpensive hardware. The hardware and software implementation is shown to meet the technical and scientific requirements that were set for the instrument. The operation of the instrument is demonstrated in practice in Longyearbyen, Svalbard. With this new ionosonde design and proof of concept implementation, it has been possible to re-establish routine ionospheric soundings at Longyearbyen, Svalbard; to replace the Dynasonde instrument that was decommissioned several years ago. It is also possible to use this new design as a basis for larger networks of ionosondes. The software and hardware design is made publicly available as open source, so that anyone interested can reproduce the instrument and also contribute to the project in the future.

Acknowledgements

I would like to thank my brilliant supervisors; Juha Vierinen, Lisa Baddeley and Mikko Syrjäsuo for guiding me through this project. Thank you, Juha, for sharing your vast knowledge with me and including me in your projects. Your enthusiasm and skill has inspired me to endeavour further into the world of radar science. Thank you, Lisa, for your sage advice, for believing in me, and for your top-notch quality British witticisms. Thank you, Mikko, for your excellent input and all the help and support in completing this thesis.

I would like to thank The University Centre in Svalbard for providing the coolest working environment a student could ask for. Thank you to all my friends that I've had the pleasure of sharing this great experience with. Thank you to Rasmus and Eike for trusting me with your dogs, and Anna for showing me the ropes of dog sledding. A big thank you to my peers on the mainland who I spent the previous years studying with.

Finally, a huge thank you to my family, for all the support and for believing in me all this time.

Contents

List of Figures

1

Introduction

As of spring 2020 there are no ionosondes operating on Svalbard. An ionosonde is a radar instrument used in near-earth space research activities. It mainly focuses on obtaining vertical electron density profiles in the ionosphere. The need for an ionosonde on Svalbard is further underlined by the highly dynamic ionosphere at high latitudes where effects of ionizing solar radiation absorbed in the upper atmosphere and the interaction between Earth's magnetic field and the solar wind are combined. Signatures in High-Frequency (HF) radar measurements at high latitudes are associated with phenomena such as polar cap absorption, polar cap patches and plasma convection. Observations of the spectral properties and frequency dependence of these signatures would yield information about the structure and dynamics of the phenomena they originate from.

An ionosonde can also be used in the general study of propagation of HF radio waves. Trans-polar propagation of radio waves is particularly important for navigation and communication, as satellite coverage is scarce on polar latitudes. Therefore, aircraft often rely on HF radio for telecommunication when the aircraft is not near air traffic control ground stations. HF radio waves propagating in the ionosphere are refracted, which enables long distance communication across the polar caps where ground stations are few and far in between. However, HF radio waves may also be completely absorbed in the ionosphere, rendering HF communication useless. Additionally, highly variable ionospheric plasma poses a challenge for HF telecommunications. It is therefore imperative to study and understand various physical phenomena

that contribute to HF radio propagation in the polar cap.

It has recently been shown that the coded continuous-wave technique is well suited for meteor radar networks [Vierinen et al., 2016]. This technique is also used for single frequency HF range-Doppler sounding [Hysell et al., 2016] and oblique HF sounding [Chartier et al., 2020] with good results. It is therefore expected that the same radar technique would be well suited for a vertical incidence ionosonde. Coded long pulses require relatively low peak power to acquire a sufficient signal to noise ratio. In addition, long transmit pulses modulated with a pseudo-random code is resilient against interference that may contaminate radar measurements. Using long pulses will also allow the use of wide bandwidth range-Doppler processing of radar echoes, enabling studies of ionospheric irregularities with wider range-Doppler extent than pulsed systems would be able to observe.

Using different pseudo-random coded long pulses make it possible to have a network of ionosondes for both vertical and oblique sounding of the ionosphere using simultaneous transmit signals with different codes to distinguish them from one another. A similar approach is used by the global positioning system (GPS). To more easily realize such a network, emphasis should be put on having low-cost, commercially available hardware, as well as open source software in its design.

While there exist high quality science grade commercial ionosonde solutions already (e.g., Digisonde, VIPIR, or CADI, discussed in Ch. 2), none of these are based on open source software. These instruments are also somewhat expensive. Additionally, all of these commercial ionosondes use short pulses, either uncoded or complementary coded and are therefore not designed to be used in a dense network of ionospheric sounders.

The United States National Academy of Sciences decadal survey; Solar and Space physics, has identified Distributed Arrays of Small Instruments (DASI) as one of the instrumentation goals for the next decade [Baker et al., 2013]. Such instrumentation networks can go after mesoscale phenomena, such as gravity waves and travelling ionospheric disturbances. Therefore, the goal of this project is to develop an ionosonde that would meet the need for a DASI-style network of ionosondes in the future. The open source nature of the design will allow the broader community to participate in improving the software in the future.

As this thesis is for the degree of Master of Science (sivilingeniør) with a specialization in space physics, the goal is to implement an ionosonde that can be used as a blueprint for a future network of low-powered, low-cost ionosondes. While the main focus is engineering, the work is interdisciplinary

due to the necessary knowledge about plasma- and ionospheric physics.

The thesis is divided into 6 chapters. The introduction (1) outlines the main scientific goal and scope of the thesis. The background (2) chapter outlines the characteristics of the ionosphere, how high-frequency electromagnetic waves propagate in the upper atmosphere and techniques and concepts related to radar probing of the ionosphere.

Chapter 3 presents the design and implementation of the new Longyearbyen ionosonde. Chapter 4 contains the testing and verification of the implemented design. Chapter 5 presents the resulting system and some of the data obtained during the first month of operating the Longyearbyen ionosonde. Chapter 6 summarizes the work conducted and the result, and provides some suggestions for further work.

In addition, an appendix is added containing derivations of some of the key physical relationships the work is based upon and installation instructions of the software used to operate the ionosonde. Some of the source code is also presented while the entire software is provided with a separate link.

2 Background

The primary goal of this thesis is to implement a software defined radio ionosonde. This chapter lays out the relevant concepts related to this topic area. The chapter is divided into six sections. The first section covers what an ionosonde is used to measure; the ionosphere. The second section deals with propagation of electromagnetic waves in ionospheric plasma, which the ionosonde relies on. The third section provides an overview of the ionosonde, which is a radar that is used to measure properties of ionospheric plasma. The fourth section discusses the radar equation, which allows us to determine the amount of electromagnetic wave power received by a radar receiver. The fifth section discusses coherent radar signal processing, which is relevant to ionosonde measurements. The final section introduces the software defined radio paradigm, which we will be using in the implementation of the Longyearbyen ionosonde.

2.1 The Ionosphere

The atmosphere of our planet is divided into several layers. The ionosphere is the region in which energetic radiation from the Sun is absorbed by the atmospheric constituents which partially ionizes the otherwise neutral atmosphere. The ionosphere can thus be said to be the conducting part of our atmosphere. It was already postulated by Gauss in the first half of the 19th century that variations seen in the Earth's magnetic field could be accounted for by currents

flowing in the atmosphere. In the beginning of the 20th century the idea of a conducting layer in the atmosphere became more relevant when Guglielmo Marconi managed to transmit a wireless signal across the Atlantic ocean, a feat only possible if the electromagnetic signal would have bounced off the upper atmosphere, which lead Oliver Heaviside in 1902 to propose the existence of a conducting region in the atmosphere. Following the Norwegian aurora polaris expedition of 1902-1903, Norwegian explorer and physicist Kristian Birkeland estimated electrical currents flowing in the atmosphere up to hundreds of kilometers in altitude, with strengths reaching up to a million Amperes [Birkeland, 1908]. The existence of the ionosphere was proven by Edward V. Appleton in 1924 [Appleton and Barnett, 1925], for which he received the Nobel prize in 1947.

2.1.1 Key physical characteristics of the ionosphere

As previously mentioned, the ionosphere is the part of our atmosphere that forms the interface between the Sun's ionized atmosphere and the Earth's neutral atmosphere. The energetic radiation propagates through Earth's atmospheric density gradient and ionizes the constituents along its path. The altitude that this occurs at extends from roughly 60 km to 1000 km above sea-level. Above this threshold the density of Earth atmosphere is too low and any constituents are already ionized. Below the 60-90 km threshold the atmospheric density is too high for most of the ionizing radiation to propagate further since the energy of the radiation is already spent ionizing the atmosphere above.

The ionization transforms the otherwise neutral gas into a plasma - a gas of positively charged ions and free electrons. To a first-approximation this plasma can be considered quasi-neutral with the number density of free electrons equaling that of the positive ions ($n_e \approx n_i$). A basic measurement of the electron density, n_e, is thus a useful tool when studying the Earth's ionosphere.

The electrons within ionospheric plasma oscillate with a natural frequency known as the plasma frequency. This frequency is expressed as

$$\omega_{pe} = \sqrt{\frac{n_e q^2}{m_e \epsilon_0}} \tag{2.1}$$

where the angular plasma frequency ω_{pe} is a function of the electron density n_e and the constants q, m_e and ϵ_0, representing the elementary charge, electron mass and vacuum permittivity respectively. By inserting the physical constants into equation 2.1, an approximate version of the equation is $f_p = 8.98\sqrt{n_e}$ where f_p is in [Hz] and n_e is in [m^{-3}]. As we shall find in later in this chapter,

this relationship between electron density n_e, the plasma frequency, ω_p and the propagation path of a wave transmitted from an antenna with frequency ω provides the physical explanation behind an ionosondes ability to obtain ionospheric density profiles. It essentially exploits the fact that transmitted waves with frequencies that matches the plasma frequency or some certain frequencies close to it will reflect off the plasma. Receiving the reflected waves will contain information about the height and density of the reflecting plasma.

Although the density of neutral particles in the ionosphere is far less than in the lower atmosphere, the region is highly dynamic because of the constant absorption of radiation which can be highly variable. The effects causing these dynamics can be attributed to two processes; ionization and recombination. Ionization is the driving effect which additionally may induce aurora and ionospheric currents. Recombination counteracts the effects of ionization, whereby particles which are active in driving the dynamics of the ionosphere depletes their energy by recombining into neutral particles. Recombination is a collisional process which is mostly dependent on the density of neutral particles. It is most prevalent in the lower part of the ionosphere (60 - 100 km) where the neutral density is higher. Ionization is a more complex process in that it varies more with altitude, and also depends on several different factors apart from the neutral density. This balance between ionization and recombination - between a conducting atmosphere and a neutral one - shifts as these factors change. Figure 2.1 shows a mathematical model which considers the balance between the incoming solar radiation intensity $I(z)$ and the background neutral atmospheric density $n(z)$ to calculate the ion-production rate $q(z)$ as a function of altitude.

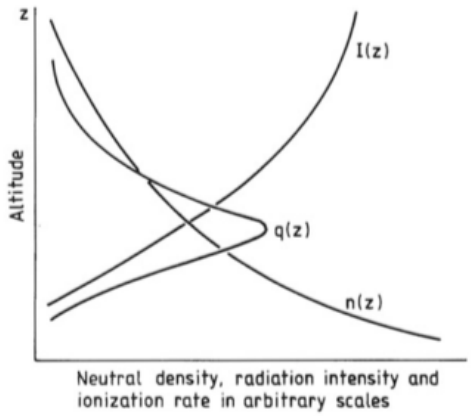


Figure 2.1: Plot showing the maximum ionization rate q occurring at the intersection between radiation intensity I and neutral density n, all functions of altitude. Figure adapted from Brekke [2012].

Through this model the ionosphere can to a first order approximation be considered as stratified layers of plasma where each layer is defined by the electron density within. The model was developed by Chapman [1931], and we refer to such layers as Chapman-layers. The expression for the ion-production rate $q(z)$ per unit volume at any given altitude is the Chapman production function:

$$q_v(z) = \kappa_v \sigma_v n_0 I_\infty \exp\left[-\frac{z}{H} - \frac{\sigma_v n_0 H}{\cos\chi_v} \exp\left(-\frac{z}{H}\right)\right] \tag{2.2}$$

where κ_v is the photo-ionization frequency, σ_v is the radiation absorption cross section, n_0 is the neutral particle density, I_∞ is the solar radiation flux at the top of the atmosphere, z is the altitude and χ_v is the solar zenith angle. H is the atmosphere scale height given by $H = k_b T_n / m_n g$ where k_b is the Boltzmann constant, T_n is the neutral particle temperature, m_n is the neutral particle mass and g is the gravitational acceleration constant of Earth.

The main source of ionization is our sun, which constantly emits radiation across the electromagnetic spectrum and ejects high-energetic particles in the form of solar-wind. Consequently, the level of ionization depends on the solar cycle, diurnally, seasonally and also latitudinally. Cosmic rays also ionize the planet and occurs rather sporadically. At the high latitudes where incoming particle precipitation also acts as a source of ionization, the Chapman-model for ion-production rate must also take this source into account. The ion-production rate due to particle collisions per unit volume is given by the Chapman production function:

$$q_e(z) = \kappa_e F_e W_{ion} \sigma_n n_0 \exp\left(-\frac{z}{H}\right) \tag{2.3}$$

Here κ_e is the collisional ionization efficiency, F_e is the flux of precipitating electrons, W_{ion} is the ionization energy and σ_n is the molecular cross section.

The effects of radiation and particle precipitation yields two ion production rates; $q_v(z)$ and $q_e(z)$. These are balanced out by processes which decreases the ion-production rate; recombination of free electrons with ions that form neutral particles, and attachments of free electrons to neutral particles which reduces the overall free electron density. By assuming a horizontally uniform quasi-neutral ionospheric plasma, meaning that the electron density matches that of the ion density ($n_e = n_i$) and ignoring diffusion and advection, the continuity equation for the electron density can be considered as the balance between ionization and recombination effects:

$$\frac{dn_e}{dt} = q_{v,e} - \alpha_r n_e^2 - \beta_r n_e \tag{2.4}$$

Here α_r and β_r represent the effects of recombination and reattachment respectively.

In a steady-state condition of the ionosphere the electron density can be considered for two cases, the first one being the case where $\alpha_r \gg \beta_r$ (90-130 km). Solving for the electron density yields:

$$n_e = \left(\frac{q_{v,e}}{\alpha_r} \right)^{\frac{1}{2}} \tag{2.5}$$

With increasing altitude recombination becomes less prevalent and we have the case where $\alpha_r \ll \beta_r$ which yields the following electron density:

$$n_e = \frac{q_{v,e}}{\beta_r} \tag{2.6}$$

In the lower altitude range where recombination α_r is the dominant loss process the electron density is proportional to the square-root of the ion-production rate. Going up in altitude where β_r is the dominant loss process the electron density is directly proportional to the ion-production rate [Baumjohann and Treumann, 1997].

An important factor to consider when dealing with quasi-neutrality and conductivity of the ionosphere is the gyrofrequency ω_g. Electrically charged particles contained in a magnetic field will gyrate around the magnetic field lines with a frequency that depends on the mass of the particle and the strength of the magnetic field as described by equation 2.7:

$$\omega_g = \frac{qB}{m} \tag{2.7}$$

where the gyrofrequency ω_g increases with the strength of the magnetic field B and the charge q, and decreases with the mass of the particle m. Generally we only consider singly charged electrons and ions, so that q only differs by sign between the two species. Consequently, the mass difference between electrons and ions ($m_e \ll m_i$) implies that the gyrofrequency of electrons is much higher than the gyrofrequency of ions when exposed to the same magnetic field ($\omega_{g,e} \gg \omega_{g,i}$).

It follows that the level of ionization is most prevalent during daytime due to photo-ionization from direct sunlight. At higher latitudes the seasons play a larger role as the tilt of Earth's rotational axis either results in constant exposure to sunlight during summer or the complete absence of it during the dark winter months. The solar cycle plays more of a role when it comes to the intensity - or amount of ionization. During solar maximum the increase in radiation intensity, sunspot number and ejection of solar-wind particles raises the average level of ionization. The solar wind particles are also the dominant source of ionization on the high latitude nightside through processes known as geomagnetic storms and substorms. Therefore, during solar minimum the

average level of ionization is lowered and there is less activity on the nightside as well.

The approach of stratifying the layers within the ionosphere in terms of the electron density has resulted in 3 main regions within it; the D-, E- and F-region. Figure 2.2 displays the electron density for night and daytime conditions during both solar maximum and solar minimum.

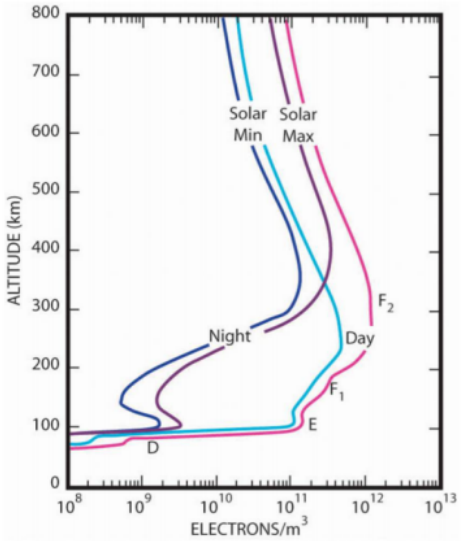


Figure 2.2: Logarithmic scale plot of the electron density between 0 - 800 km altitude. Figure adapted from Baddeley [2019].

In the figure we see the different regions labeled along the electron density curve. Starting from the bottom we have the D-region defined by a typical daytime electron density of $n_e = 10^8 - 10^9\, m^{-3}$. The D-region is between $\sim$ 60 - 90 km. The atmosphere here is defined as weakly ionized, meaning that the electron-neutral collision frequency ν_{en} is higher than the electron-ion collision frequency ν_{ei}. The dynamics observed here are mostly attributed to neutrals since the ion-neutral collision frequency is greater than the ion-gyrofrequency ($\nu_{in} > \omega_{g,i}$).

The E-region is the partially ionized ($\nu_{en} > \nu_{ei}$) region stretching from about $90 - 130$ km with electron densities ranging from $10^9 - 10^{11}\, m^{-3}$. The E-region is where we observe the horizontal current systems due to an emerging relative motion between ions and electrons. The main ionization sources of the D- and E-region is solar EUV during daytime and particle precipitation during geomagnetic storms and substorms.

Continuing upwards from 130 – 600 km we have the F-region. Characterized by an electron density between $10^{11} - 10^{12}\ m^{-3}$ and being fully ionized ($\nu_{en} < \nu_{ei}$). There are no horizontal currents in this region due to the reduced effects from collisions with neutrals that induce relative motion between electrons and ions ($\nu_{en,in} < \omega_{g,(e,i)}$). In the lower part of the F-region is also an additional sub-region called the F1-region which is an enhanced electron density caused by incoming solar UV-radiation during daytime. Therefore the F-region is divided into the F1- and the F2-region, with F2 being the only region present during day and night, while the F1-region dissipates during the night due to the higher recombination rate of the ions in this region.

To understand the phenomena within the plasma and how it affects the propagation of electromagnetic waves it is necessary to know the different parameters of it, mainly density, temperature and altitude distribution. In addition, the background magnetic field is also affecting EM-waves so it should be included, but it is not an intrinsic property of the plasma as it originates from the Earth's core. These parameters are directly influential in deciding cut-off and resonance frequencies of the plasma, these are frequencies at which waves are reflected or absorbed respectively. They are also necessary when discussing refraction and polarization of the waves, as was investigated by Appleton [1932]. The plasma density is also affected by neutral dynamics such as gravity waves and advection, the effects of which is an ongoing research topic.

2.1.2 High latitude ionosphere and phenomena

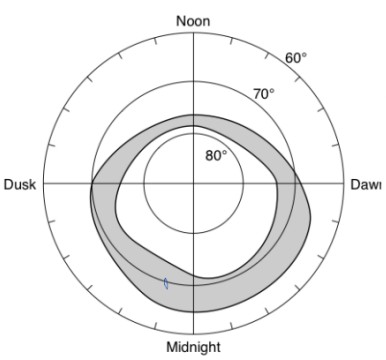
A complete overview of winds, waves, convection cells and other dynamic patterns in the upper atmosphere is beyond the scope of this thesis, we will however attempt to shed some more light onto an important aspect of ionospheric dynamics, namely the electron density, with a particular interest in the high-latitude polar region where magnetic field lines converge and form the polar cap which is the defined as the area enclosed by the interface between the closed magnetic field lines of the Earth to the south and open magnetic field lines to the north. The high-latitude region can typically be regarded as anything above 65° latitude[1].

As discussed in section 2.1.1, the electron density is an important parameter in describing the various ionospheric phenomena and structure. The polar cap is a particular region of interest since it is here that the effects of the solar wind and the solar magnetic field embedded within it interacts with the Earth's

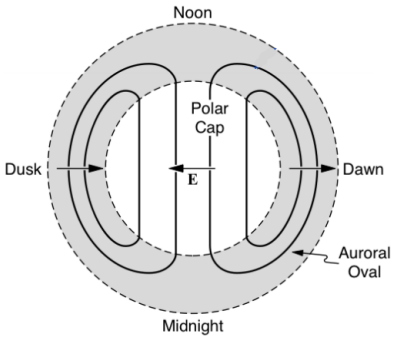
1. There is a distinction to be made between magnetic latitude and geographic latitude, however the location of the magnetic north pole relative to the geographic north pole is not of crucial importance within the scope of this thesis.

magnetic field (magnetosphere) which can result in numerous dynamics and phenomena, some of which are listed here: 1) The opening and closing of magnetic field lines which expands and contracts the polar cap region [Wang et al., 2016]. 2) The horizontal convection of plasma across the magnetic poles in the F-region and the large scale electric fields and currents in the polar cap [Cowley and Lockwood, 1992]. 3) The precipitation of high-energetic particles during geomagnetic storms and substorms resulting in highly dynamic plasma parameters and the subsequent turbulence and instabilities that follow [Perreault and Akasofu, 1978; Akasofu, 1964, 2013, 1981]. 4) The absorption of radio waves in the polar cap during heavy ionization in the D-region [Bailey, 1964].

Figure 2.3 shows some of the main characterizations of the auroral oval. The plots are polar projection plots looking down onto the north polar region from above the ionosphere with latitude marked by concentric rings going from 60° to 80° in figure 2.3a. The statistical location of the auroral oval, where particles precipitate into the atmosphere from the solar wind is shown in grey. The times are marked as magnetic local time which is a co-ordinate system widely used in ionospheric physics as it relates phenomena with respect to the location of the sun. Magnetic noon points towards the sun, with magnetic midnight pointing away from the sun. The polar cap region is defined as the area encompassed inside the poleward boundary of the oval. Figure 2.3b shows the large scale horizontal electric fields and plasma velocity contour inside this region.



(a) A typical auroral oval within the polar cap.



(b) Model of induced electric fields and velocity contour for horizontal drift in the F-region polar cap.

Figure 2.3: The auroral oval and the polar cap. Both plots adapted from Baumjohann and Treumann [1997].

Phenomena

Phenomena of particular interest related to the ionosonde discussed in this thesis are presented here. In general, ionospheric electron density profiles are useful in terms of just characterizing the general state of the ionosphere. The polar region is a region of high ionospheric activity, yet elusive due to the remoteness and lack of measuring infrastructure. Density profiles are also useful for describing the phenomena known as sporadic E. Sporadic E are thin, sudden patches of high electron density that occur as the name suggest sporadically in the E-region [Kirkwood and Nilsson, 2000; Turunen et al., 1993; Lanchester et al., 1991; Nygren et al., 1984; Nygrén et al., 1984].

Another phenomena is high latitude spread F which refers to the F-region appearing diffuse due to density irregularities which scatter the probing radio wave. The received signal that measures spread F is then a superposition of several waves that are reflected from multiple heights and locations within the F-region. Spread F is often associated with geomagnetic storms and substorms that generate these density irregularities [Perkins, 1973].

Polar cap absorption is an effect whereby high-frequency radio waves are absorbed completely in the ionosphere due to energy loss via electron-neutral collisions as described in equation 2.15 of section 2.2.1. Absorption happens primarily during events of ionization at low altitudes where the neutral density is high.

The study of basic plasma physic instabilities and waves can also utilize measurements obtained by an ionosonde. The waves and instabilities share common traits in that they all depend on the scale size of interest and the plasma parameters; namely electric fields and density gradients. For a study of gravity waves in this context, see Crowley and Rodrigues [2012]. The instabilities of interest is the Farley-Buneman instability, other two stream instabilities and gradient drift instabilities presented further by Milan et al. [2003].

Ultra Low Frequency waves (ULF waves) are pulsations of the Earth's magnetic field that have a frequency range between ~1 mHz - 100 Hz [Pilipenko, 1990]. The ULF wave energy flux that maps to the ionosphere is associated with many ionospheric processes such as the formation of auroral arcs. The measurements of ULF waves can also be used to estimate joule heating rates, a parameter which is associated with currents and conductivities in the ionosphere [Baddeley et al., 2005; Hartinger et al., 2015].

2.1.3 Typical ionospheric plasma parameters over Svalbard

Longyearbyen, situated at the Svalbard archipelago, lies at 78°N where large parts of the year either consists of perpetual darkness, during the polar night, or perpetual sunlight, during polar day. This location is statistically often also located inside the auroral oval and polar cap, where the primary source of dynamics is ionospheric convection. During these periods of the year the ionospheric conditions between day and night should be similar to a first approximation.

This section presents data from the International Reference Ionosphere (IRI) and the International Geomagnetic Reference Field (IGRF) models to get a brief description of the typical day and nightside ionospheric plasma parameters during polar day and polar night. The data is shown in table 2.1, table 2.2 and figure 2.4.

The latest IRI model is developed by COSPAR and URSI [2019] and is an empirical standard model based on all available ionospheric data, primarily from a worldwide network of ionosondes and the few but powerful incoherent scatter radars. Satellites, rockets and other types of radars also contribute to this model. Because of the relative scarcity of these measurements, empirical data is lacking.

The IGRF model is a mathematical model of Earth's magnetic field, developed by IAGA [2019]. It contains various coefficients used to describe the different aspects of the field. Magnetic surveys from all over the world contributes to the determination of these coefficients and they are re-calculated every 5th year to keep up with the changing magnetic field of Earth. Because of the relatively stable magnetic field which changes on a significantly larger timescale than the ionosphere, empirical data obtained for the magnetic field is sufficient to produce a robust and precise model.

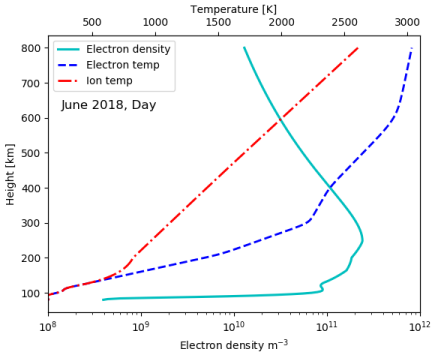
Even though extensive efforts are taken to measure the ionosphere for the IRI model, the data shown in figure 2.4 are heavily model based, particularly the electron and ion temperatures. In addition, at higher latitudes the highly dynamic ionosphere, coupled with the fact that ground based measurements are even more sparse (due to lack of accessible locations) means the modelled data fails to capture the true variability and dynamics of the system. More measurements are thus needed to describe the dynamics with more precision.

Table 2.1: Model data of plasma parameters over Longyearbyen for June 2018 00 UT. 1st column contains altitudes in intervals of 100 km. 2nd column contains the electron density, 3rd and 4th column are electron and ion temperatures respectively and 5th column contains the background magnetic field strength.

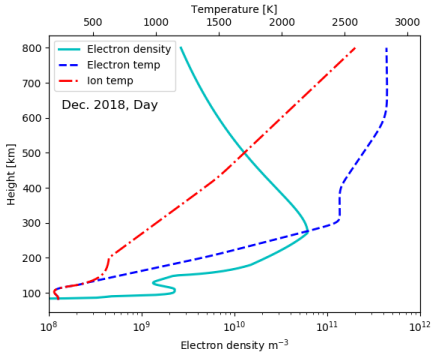
alt [km]	n_e/m^3	T_e [K]	T_i [K]	B [nT]
800	1.0×10^{10}	2848	2607	39800
700	1.5×10^{10}	2868	2307	41380
600	2.6×10^{10}	2898	2007	43000
500	4.8×10^{10}	2984	1707	44760
400	1.0×10^{11}	2995	1407	46570
300	1.8×10^{11}	1970	1102	48470
200	7.7×10^{10}	1396	799	50470
100	4.2×10^{10}	221	221	52500

Table 2.2: Model data of plasma parameters over Longyearbyen for December 2018 00 UT. Comparing parameters from December with June we see that the electron densities have decreased by roughly one order of magnitude and electron and ion temperatures have decreased by 10 - 25%.

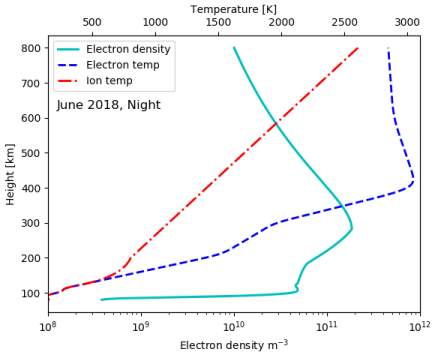
alt [km]	n_e/m^3	T_e [K]	T_i [K]	B [nT]
800	1.8×10^{9}	2105	1501	39800
700	2.6×10^{9}	2075	1500	41380
600	3.9×10^{9}	2011	1499	43000
500	6.7×10^{9}	1760	1498	44760
400	1.3×10^{10}	1471	1395	46570
300	2.6×10^{10}	1127	1060	48470
200	4.4×10^{10}	798	724	50470
100	1.6×10^{9}	190	190	52500



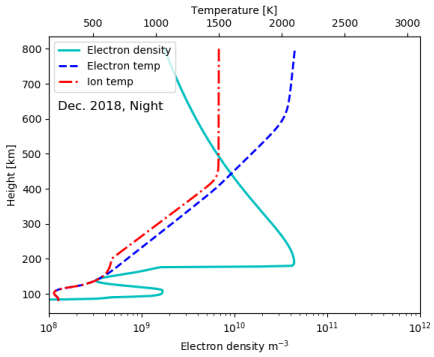
(a) June 2018 dayside. Continuous photoionization leads to high electron density.



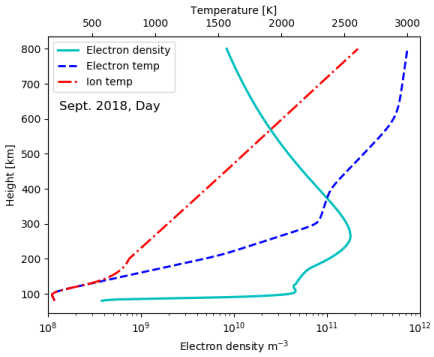
(b) Dec. 2018 dayside. Low rate of photoionization during polar night leads to a decrease in electron density.



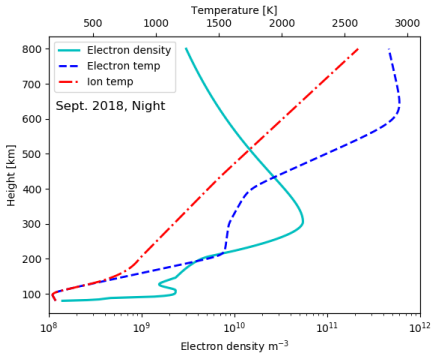
(c) June 2018 nightside. Not much change relative to dayside since the nightside is still exposed to sunlight during polar day.



(d) Dec. 2018 nightside. Electron density in the E-region decreased compared to dayside. Temperatures have decreased but data-quality is poor.



(e) Sept. 2018 dayside. Roughly identical to dayside conditions in June.



(f) Sept. 2018 nightside. No more polar day and thus a clearer distinction between day and nightside conditions.

Figure 2.4: Data from the IRI model showing electron density, electron temperature and ion temperature in the altitude range 80-800 km over Longyearbyen, Svalbard. The 1st column shows dayside conditions, the 2nd column shows nightside conditions. Temperature along upper x-axes, electron density along lower x-axes. Altitude along y-axes.

2.2 Propagation of Electromagnetic Waves in the Ionosphere

In this section we present some of the mathematical framework used to describe the propagation of electromagnetic waves in the ionosphere. The mathematical descriptions of electromagnetic waves propagating in the ionosphere is based on magneto-ionic theory which was developed alongside the development of radio communications and radar in the decades around World War II.

The ionosphere is a birefringent medium, this means that the refractive index of it depends on the polarization and propagation direction of the incident radio wave. The birefringent property mainly stems from the additional effects from the Earth's magnetic field on the ionospheric plasma. When a radio wave is transmitted it is thus split into 2 different waves. These two waves are called the Ordinary wave and the Extraordinary wave[2]. These two waves represent two different "modes" of propagation that the incident wave can divide itself into and they are usually referred to as the "O-mode" and the "X-mode" of the incident wave respectively. To determine to what extent the incident wave divides itself into each of the two modes of propagation one has to resort to complicated parts of magneto-ionic theory provided by Davies [1990] and references therein.

The motivation for exploring these two modes of propagation is because they take different paths when travelling through the ionosphere due to dispersion. This means that an ionogram can have two different echo traces of the same transmitted signal. The propagation modes are described using dispersion relations which relates a wave's frequency ω with its corresponding wave number k. The dispersion relation for the O- and X-mode has been derived in appendix A for propagation along the magnetic field with the included effects of particle collisions.

In the case of propagation parallel to the magnetic field, both O- and X-modes become circularly polarized around the magnetic field, but with the E-field rotating in opposite direction for each mode. The O-mode will rotate its electric field around the background magnetic field using a Left hand scheme (clockwise if magnetic fields points up) and the X-mode will rotate about the magnetic field using a Right hand scheme (anti-clockwise if background magnetic field points up).

This special case for propagation parallel to the magnetic field also incorporates the effect of the gyrofrequency of electrons around the magnetic field, which

2. Another mode of propagation is the Z-mode but it is out of the scope of this thesis.

is why for this case, the O- and X- mode are referred to as the L- and R-mode respectively[3].

If an O-mode wave vector turns parallel to the magnetic field it will transition smoothly to an L-mode propagation and if an X-mode wave vector turns parallel to the magnetic field it will transition into an R-mode propagation (chapter 9.5 of Baumjohann and Treumann [1997]). Their dispersion relations (when including collisional effects) are as follows:

$$\frac{c^2k^2}{\omega^2} = 1 - \frac{\omega_p^2}{1 - \frac{\omega_c}{(\omega - i\nu_e)}} \qquad \text{(R-wave)} \tag{2.8}$$

$$\frac{c^2k^2}{\omega^2} = 1 - \frac{\omega_p^2}{1 + \frac{\omega_c}{(\omega - i\nu_e)}} \qquad \text{(L-wave)} \tag{2.9}$$

The derivation of the R- and L-mode dispersion relation can be found in appendix A.

2.2.1 The Appleton-Hartree equation

The Appleton Hartree equation describes the refractive index n of Earth's ionospheric plasma. As an electromagnetic wave travels through the ionosphere, its ray-path will be bent (refract) and its intensity will decrease as a result of absorption. The refractive index (Appleton Hartree equation) yields how much the ray will be bent and by how much it will be absorbed in the plasma as a function that ultimately altitude and frequency of the wave itself. This is because the constituents of the equation (plasma frequency, collision frequency ν_e and gyrofrequency ω_g) are themselves functions that depend on altitude.

$$n(\omega, \omega_g(z), \omega_p(z), \nu_e(z)) = n(\omega, z) \tag{2.10}$$

Here ω is the angular frequency of the incident wave and z is the altitude. There is however, a more complicating factor to consider, which is the angle between the propagation path and the background magnetic field which results in different wave-modes of the incident wave occurring inside the plasma (O- and X-mode polarization). We then have

$$n = n(\omega, z, \theta) \tag{2.11}$$

where θ is the angle between the incident wave path and background magnetic field(which itself also depends on altitude). The complete Appleton Hartree

3. The X-mode takes into account the effects of the gyrofrequency regardless of propagation parallel or perpendicular to the magnetic field, the O-mode however does not.

equation is given as follows

$$n^2 = 1 - \frac{X}{1 - iZ - \left(\frac{Y_T^2}{2(1-X-iZ)}\right) \pm \left(\frac{Y_T^4}{4(1-X-iZ)^2} + Y_L^2\right)^{\frac{1}{2}}} \tag{2.12}$$

where X, Y and Z are defined as:

$$\begin{aligned} X &= \frac{\omega_p^2}{\omega^2} \\ Y_T &= Y \sin\theta \\ Y_L &= Y \cos\theta \\ Y &= \frac{\omega_g}{\omega} \\ Z &= \frac{\nu_e}{\omega} \end{aligned} \tag{2.13}$$

Here Y_T and Y_L refer to the transverse and longitudinal component of the wave with respect to the background magnetic field. For propagation along the magnetic field, $Y_T \approx 0$. For propagation perpendicular to the magnetic field, $Y_L \approx 0$. ω_p and ω_g refer to the plasma and electron-gyrofrequency respectively. ν_e is the electron-neutral collision frequency.

The derivation of The Appleton Hartree equation can be found in appendix B. The refractive index is complex due to the fact that electrons that are put in motion by the incident wave loses energy due to collisions - thereby absorbing energy off the wave. This can be seen by inserting the refractive index into the expression for the electric field of a plane electromagnetic wave traveling in the z-direction. By relating the complex refractive index $n^2 = (\mu + i\chi)^2$ to the complex wave number $k = 2\pi n/\lambda$ through the relations

$$k = \frac{2\pi}{\lambda}, \qquad n = \frac{\lambda_0}{\lambda} \tag{2.14}$$

where λ_0 is the wavelength of the wave in vacuum, we see that when inserting for k into the electric field

$$\begin{aligned} \mathrm{E}(z,t) &= \mathrm{Re}\left[\mathrm{E}_0 e^{i(kz-\omega t)}\right] \\ &= \mathrm{Re}\left[\mathrm{E}_0 e^{i(2\pi(\mu+i\chi)z/\lambda_0)-\omega t}\right] \\ &= e^{-2\pi\chi z/\lambda_0}\mathrm{Re}\left[\mathrm{E}_0 e^{i(kz-\omega t)}\right] \end{aligned} \tag{2.15}$$

χ yields a factor of exponential decay as a function of range due to the collisional friction acting upon individual electrons. The real part of n deals with the refraction (bending) of the wave.

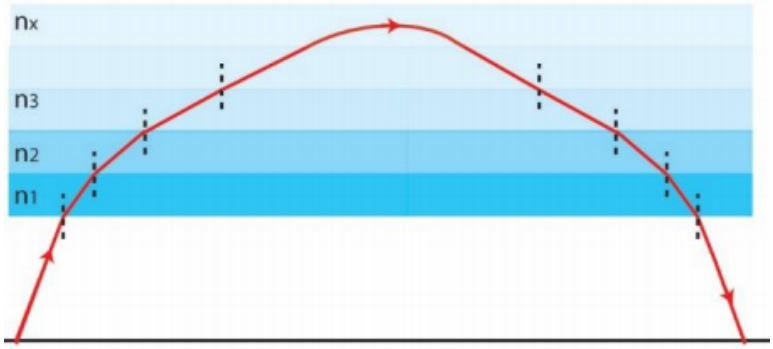


Figure 2.5: Example of how refraction of an electromagnetic wave might look like in the ionosphere. The refractive index changes with altitude because of the changing plasma parameters, magnetic field and collisional frequency. The point at which the wave refracts down towards earth again is the point where the refractive index is below unity and the group velocity reverses. Plot adapted from Baddeley [2019].

Using equation 2.12 we will now explore the refractive index for 2 cases. The simplest case is when we neglect collisions and magnetic field, i.e. ω_g, $\nu_e \sim 0$, the refractive index then simplifies to:

$$n^2 = 1 - X = 1 - \frac{\omega_p^2}{\omega^2} \tag{2.16}$$

The refractive index relates to the wave number k and the frequency ω through the following relation:

$$n^2 = \frac{k^2 c^2}{\omega^2} = 1 - \frac{\omega_p^2}{\omega^2} \tag{2.17}$$

Assuming that $\omega > \omega_p$, as the wave propagates upwards, the electron density increases and thus the plasma frequency ω_p increases through the relation given in equation 2.1. As ω_p increases, n and k decreases to the point where eventually $\omega_p = \omega$, the point at which $n = k = 0$. This is the point of total internal reflection and the point at which the direction of propagation of the incident wave reverses and is reflected back. The frequency ω where this occurs at is called the cut-off frequency. If the frequency of the incident wave is higher than the maximum plasma frequency the wave will penetrate through the ionosphere without being reflected.

The next case we will discuss is one that includes the magnetic field. We still neglect collisions with $\nu_e \sim 0$ and set the refractive index to $n = 0$ to obtain the conditions for reflection. In this case, the Appleton Hartree equation 2.12

evaluates to 3 possible solutions:

$$1 - X = 0 \tag{2.18}$$
$$1 - X - Y = 0 \tag{2.19}$$
$$1 - X + Y = 0 \tag{2.20}$$

The wave ω corresponding to 2.18 is referred to as the "ordinary" wave (O-mode) since it is the same reflecting wave for the case of no magnetic field. The other two possible waves in 2.19 and 2.20 are referred to as "extraordinary" waves. A distinct feature of the extraordinary waves is that they only reflect at the points where $\omega = \omega_p + \frac{\omega_g}{2}$ (X-mode) and $\omega = \omega_p - \frac{\omega_g}{2}$ (Z-mode) regardless of the angle θ between the incident wave and the magnetic field.

2.3 The Ionosonde

2.3.1 The Breit & Tuve experiment

Breit and Tuve [1926] detailed the first radar experiment designed to test the existence of this conducting layer first theorized by Heaviside in 1902. In their experiment they transmit what they refer to as "interrupted trains of waves". Using receiver antennas at different locations they detect, amplify and plot the transmitted signal in an oscillogram where a single "hump" corresponds to the transmitted signal, but if the oscillogram shows two or more "humps" corresponding to the transmitted signal then the additional humps refer to "echoes" of the transmit signal. These echoes are the signatures of reflections of the signal coming from the conducting layer in the upper part of our atmosphere. Using the displacement between the echo and the original transmit signal in the oscillogram they obtain the time-delay between transmission and echo received, thereby revealing the "virtual reflection height". The experiments of Breit and Tuve can be seen as the first ionospheric radar measurement, and they mark the beginning of ionospheric sounding using electromagnetic waves (radar).

The technique utilized by Breit and Tuve was that of pulsed radar where a short radio burst is transmitted with a certain frequency and is then silenced for the receiver to listen for echoes. The pulse repetition interval was about 1 ms where the pulse varied between 1/3 and 1/5 of each interval which corresponds to a 33% and 20% duty cycle respectively. Different wavelengths of the pulse were tested, varying between some 675 to 20 meters, but echoes were mostly prominent around ~70 m wavelength, corresponding to a frequency of ~4.3 MHz. Some distorted echo was also observed on 40 m (7.5 MHz). Experiments were conducted with a 5 kW power in the transmission signals,

but there were also tests where 10 kW power were used.

The most pronounced echoes were observed 12 km south of the transmitter in Bellevue, Washington D.C. Between 1030 - 1050 A.M on July 28, 1925 double humps in the oscillograms of the receiving end could be seen with pronounced fading in one of the humps. The transmitted signal had a wavelength of 71.3 m, corresponding to a 4.2 MHz frequency. The first hump was concluded to be the direct overground transmission propagating horizontally between the antennas. It showed a nearly constant amplitude and was received best on a nearly vertical antenna. The second hump had its largest amplitude when the receiving antenna was horizontal, i.e. the effective area of the antenna was greatest for rays coming down vertically.

The fading observed in the second hump was hypothesized to be due to interference between reflections of the ionosphere and reflections coming up from the ground, forming a node at the antenna location. There was however very marked variability in the amplitude of this fading signal, suggesting that the reflection itself is "governed by interference phenomena or that it is caused by sudden changes in the reflecting layer more or less as the flickering of light on a wavy surface of water" Breit and Tuve [1926].

By assuming the echo signal as a pure reflection from zenith one can express the time of flight of the signal as

$$\Delta t = \frac{2h}{c}$$

By doing measurements of the time delay between echo and transmission they acquire the height of the reflecting layer. Two experiments were conducted using this method, one in July and one in September. The height of the reflecting layer was found to be higher in September than in July and in both experiments the layer seems to be higher in the afternoon than in the morning.

The height they deduced was not in reality the height of the reflecting layer. In the paper the authors suggests that the wave they transmit is more refracted than it is reflected, i.e. the ray-path is gradually bent as the wave propagates upwards until it reaches the critical frequency and starts to propagate downwards again. The time of flight is thus longer than the case for just pure reflection because the signal spends a portion of the time propagating horizontally. The result is that the height of the reflecting layer appears to be higher up than it actually is. The height they deduced is then in modern terms described as the virtual reflection height which was about 220 km in daytime with some reflections from about 90 km [Hulburt, 1974].

The paper further discusses other sources for errors in the calculation of h,

primarily the changing group velocity of the wave in the refracting medium and the possible oblique directions of the rays they observe (neither vertical nor horizontal). In regards to the group velocity, they also discuss the different polarization modes yielding two different rays that arrives at different point in time due to different group velocity. The presence of the different polarization modes supported the claim that the electron density was different from one that is proportional to the square of the altitude, and that changes in density can be rather sudden and discontinuous. The regions of enhanced electron density became known as the F- and E-regions shortly after the publication [Hulburt, 1974].

2.3.2 Modern ionosondes

The ionosonde is a type of radar designed to probe the ionosphere for its virtual height. Modern, more sophisticated versions of the ionosonde are also able to measure echo parameters such as amplitude, polarization, direction of arrival, phase and Doppler shifts. In general however, the basic ionosonde just conducts a pulsed frequency sweep across the HF-spectrum and listens for echoes when it is not transmitting. The basic ionosonde consists of a High-Frequency transmitter, typically ranging from 0.1 MHz to 30 MHz, a receiver able to track the transmitted pulse, an antenna with a vertical beam pattern efficient over the HF-spectrum and some digital data analysis circuitry. Figure 2.6 shows a map with existing ionosonde sites as of 1994. The ionosonde in Longyearbyen shown on the map is as of June 2015 no longer operational.

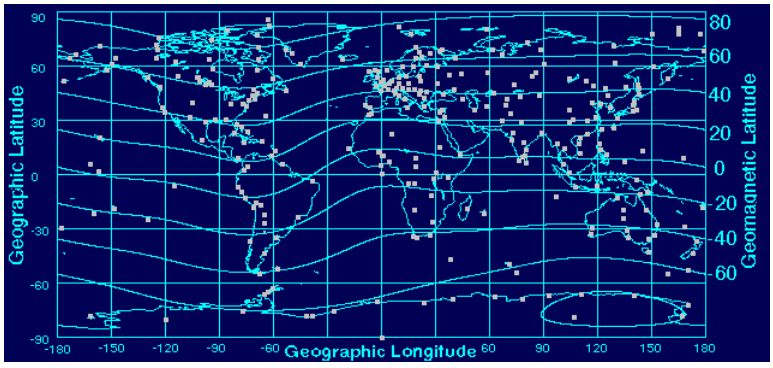


Figure 2.6: Map highlighting existing ionosonde sites with geographic latitude and longitude as well as geomagnetic (outdated) latitude. Image obtained from ngdc.noaa.gov.

The Dynasonde

The Dynasonde distinguishes itself from other ionosondes in that it uses the raw data of the echo to conduct real-time analysis without resorting to Fourier transforms, pulse coding or coherent summations. The radio echo for the Dynasonde is defined by the signal's complex representation to enable measurement of amplitude, phase, pulse group delay, Doppler shift, direction of arrival and wave polarization.

To do these measurements, the transmissions of a Dynasonde are spectrally "pure" radio pulses, where at the receiving end the signal's complex representation is digitized every 10 microsecond during time of arrival within a 467-5333 microsecond (70-800 km) listening window. A sample of the background noise outside the time of arrival window is also measured to subtract from the incoming signal to distinguish the echoes more and helps discard false echoes from real ones. Pulses are transmitted in groups called "pulsets" to make the transmissions more recognizable from possible false echoes.

The collection of complex amplitudes are what makes up the Dynasonde echo. A result of this is that the Dynasonde provides good statistical data for what constitutes a real echo by having a large set of raw data that are recognized echoes. This raw data provides an autonomous classification of the echoes and the subsequent parameter estimates, each with their own error estimate. One of the recent incarnations of the dynasonde system is the VIPIR (Vertical Incidence Pulsed Ionosphere Radar). A description of the system can be found in Bullett et al. [2014].

The Digisonde

The Digisonde is another type of ionosonde that measures virtual height, echo amplitude, phase shift, Doppler shift, angle of arrival, wave polarization and also wave-front curvature [Davies, 1990]. Instead of basing its measurements on the raw IQ echo data like the Dynasonde, the Digisonde uses pulse amplitude modulation and pulse compression to correlate the received signal with the transmitted pulse. By alternating the amount of frequency bins used in the transmission, the Digisonde can operate in different modes depending on if the user wants high temporal resolution (few frequency bins) or high frequency (range) resolution (many frequency bins). A recent implementation of the digisonde is the Digisonde-4D system, described in Reinisch et al. [2009].

2.3.3 Ionosondes in the study of the ionosphere

Because the physical parameters of the ionospheric plasma affect how the radio waves propagate it is possible to use the parameters of the measured radio wave itself to determine unknown ionospheric parameters. We have a known signal that we transmit into the unknown ionosphere and we measure the returning transmit signal(s). How the returning signals look compared to how they were transmitted gives us the necessary data to work out exactly what the ionospheric conditions must have been like in order to affect the transmit signal the way they did. Among the ionospheric parameters which affect the propagating wave, we mainly focus on the critical plasma frequency (when the frequency of the transmission matches that of the ambient plasma frequency) and also to some extent the electron gyro-frequency and collision frequency, referring to the Appleton Hartree equation given in 2.12. These parameters are the ones that modify the echo properties at an increasing rate as the wave reaches the point of total reflection.

By analyzing the information obtained from the echo (its travel time, phase, polarization, angle of arrival, amplitude and Doppler shift) one starts to form an image of the spatial irregularities, its motions and time-evolution of the reflecting medium [Reinisch, 1986]. In particular, using the images obtained by ionosondes, one can utilize other sources of data to look at how varying solar activity, aurora, meteors, other ionization sources and winds affect the spatial and temporal evolution of the ionosphere. Figure 2.7 shows an ionogram obtained by the Gakona digisonde in Alaska with added markers to explain the different data-points of the ionogram.

What is not explained in figure 2.7 but also essential are the letters and numbers listed on the left of the figure. These letters and numbers refers to the peak-frequency, and thereby also peak density within the various regions. The uppermost of these shows the f_oF2 corresponding to a frequency of 4.838 MHz. Referring back to equation 2.18; f_oF2 refers to the critical frequency of the O-mode wave in the F2-region, similarly, f_oF1 refers to the critical frequency for the O-mode in the F1-region and so on.

As a consequence of the physics behind reflection of HF-waves in the ionosphere, ionosondes are only ever able to measure electron-density up until the peak electron-density height, and not the structures beyond in altitude since the waves will already have reflected off the structures beneath. It is then only the bottom side of the peak-density of figure 2.2 that is visible to an ionosonde, while the topside remains invisible.

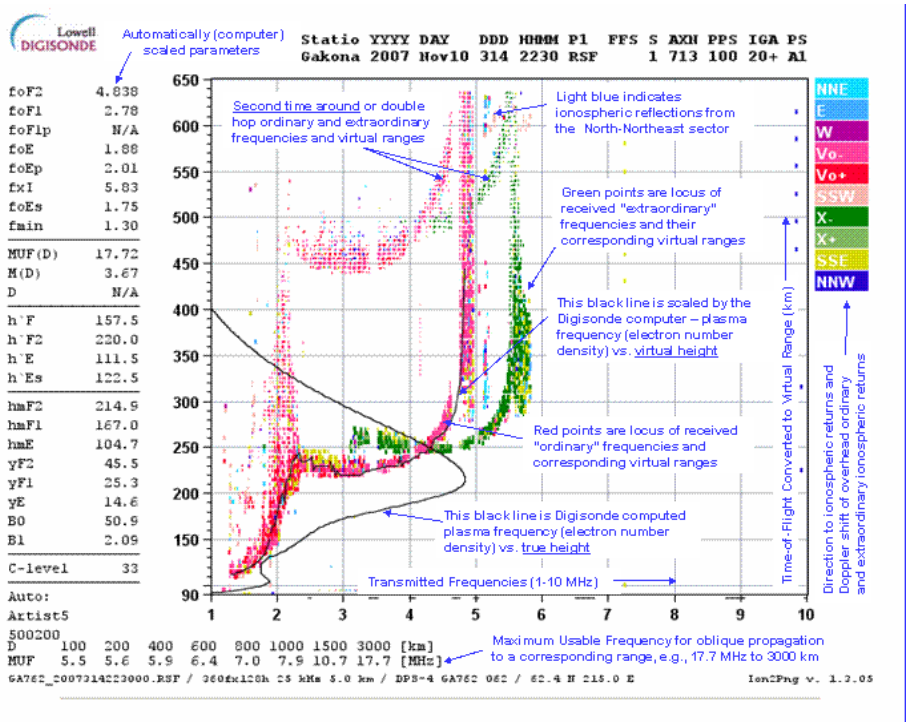


Figure 2.7: Sample ionogram from the Gakona digisonde in Alaska, November 10th, 2007 at 2230 UT. Data points are coloured to highlight which polarization the echo has and from which direction it came from. The y-axis displays the virtual range and the x-axis displays the frequency of the echo. Downloaded from HAARP: `http://www.haarp.alaska.edu/haarp/dsonde.html`

2.3.4 Ionosonde applications in a larger network

The ionosonde is analogous to the fish at the bottom of the ocean trying to observe the waves on the surface overhead. In our case, the surface is hundreds of kilometres away and the surface itself is not a clear transition between ocean and air, it is a gradually changing environment transitioning from our relatively stratified atmosphere into a highly irregular environment where the densities and structures propagates as waves and are exposed to the current systems of our atmosphere and the external, ionizing radiation.

As mentioned earlier, an ionosonde would provide snap-shots of this environment and over time be able to reveal some low-resolution, general picture of the structure of the ionosphere. On its own it is a novel and effective instrument in characterizing HF-propagation in support of communications operations. In order to characterize the highly dynamic ionosphere (in particular in the polar regions) several ionosondes working together would be more benefi-

ciary to obtain higher-resolution imaging. Through the use of triangulation one would be more able to ascertain the motion of the reflecting plasma by looking at echoes that have a known location due to the known time-of-flight between all the stations and the known distance between them. Additionally, with a network of ionosondes one could have a portion providing high time-resolution ionograms using for instance longer pulse widths whilst other could provide good range-resolution ionograms for the same reflecting area by using shorter pulse widths (Pulse widths will be discussed later in section 2.5.3). This would enable researchers with novel data to look further into the dynamics of the ionosphere by studying irregularities such as gravity waves, travelling ionospheric disturbances (TIDS) and Ultra Low Frequency (ULF) waves [Hunsucker, 1990].

2.4 Radar

An ionosonde is in essence, a simple radar (RAdio Detection And Ranging) system. It transmits a pulse of known power and duration, into the ionosphere,and then detects the Radio Frequency (RF) power that has been backscattered by a target.

2.4.1 Radar equation

The radar equation is a formula that considers the effects of the various elements of a radar on the signal power. Assuming that we have a lossless transmission medium and no effects of multi-path interference, that is, the transmit signal interfering with itself, the radar equation can be formulated as follows:

$$P_R = \frac{P_T G_T A_R \sigma}{(4\pi)^2 R_t^2 R_r^2} \tag{2.21}$$

The power received P_R is here a function of the transmitted power P_T, the transmitter gain G_T, the effective aperture of the receiver antenna A_R, the reflecting characteristics of the target i.e. the radar cross-section σ, the distance between target and transmitter R_t and the distance between target and receiver R_r. For a monostatic radar we have $R_t = R_r$ and thus:

$$P_R = \frac{P_T G_T A_R \sigma}{(4\pi)^2 R^4} \tag{2.22}$$

Using the radar equation one can assess the performance of the radar under various circumstances by varying the different factors. For instance, if we want to know what the requirements on the transmitter power are for the radar

to be able to receive 1 W of power from a target 300 km away with a known cross-section we can insert the known quantities into equation 2.22 for a 1st order approximation.

2.4.2 Gain

One of the more esoteric aspects of radars and antenna theory is the antenna gain, which in equation 2.22 is the factor G_T. The gain of an antenna is a dimensionless number that describes its performance. It is a combination of the directivity D and electrical efficiency ϵ:

$$G = \epsilon D \tag{2.23}$$

The efficiency ϵ is a measure of how well the antenna transforms incoming electrical power to radio waves. It is defined as the ratio between the radiated power and input power to the antenna ($\epsilon = P_{rad}/P_{in}$) and is thus a number between 1 and 0 since the output power cannot be higher than total input power. An efficiency of 1 is the ideal case with zero energy loss in the antenna. The case of zero efficiency means that the antenna would quench the incoming power completely and not radiate anything (at which point it would no longer be considered an antenna). The directivity D is a measure of how well the antenna transmits radio waves in a certain direction. It is usually defined in spherical coordinates:

$$D(\theta, \phi) = \frac{U(\theta, \phi)}{P_{tot}/(4\pi)} \tag{2.24}$$

where the directivity D is the ratio between the radiation intensity (power per unit solid angle) generated in a particular direction U and the averaged radiation intensity across every direction $P_{tot}/(4\pi)$.

The ideal isotropic antenna radiates uniformly in all directions and has a directivity of 1, it often serves as a reference when considering the directivity of other types of antennas. The perfect half-wave dipole is also a known reference antenna, it has a directivity of 1.64 in the direction perpendicular to itself. The short dipole antenna (short = less than 1/10 of the wavelength) has a directivity of 1.5[4].

Higher directivity means that the radiated power gets more concentrated. A dish antenna can have a directivity of 10 000, which implies that the radiated power is highly concentrated within a narrow region of space. Radiated power that is concentrated is referred to as a beam. The beam-width of a directive antenna

4. Values for directivity obtained from Wikipedia: `https://en.wikipedia.org/wiki/Dipole_antenna`

is defined as the angle between the axis of highest directivity (boresight) and the axis at which the directivity is halved.

It should also be noted that the gain of a receiving antenna can be defined in the same way. In equation 2.22 the receiver gain is incorporated into the effective aperture A_R which is related to gain in the following way:

$$A_R = \frac{G}{\lambda^2 4\pi} \tag{2.25}$$

where G is the receiver antenna gain and λ is the wavelength of the radio waves.

2.4.3 Noise and signal to noise ratio

One of the main things to consider in radar applications is noise effects on the signal. If we tried to measure our signal with a receiver without transmitting anything, we would still pick up a signal in our receive antenna due to ambient presence of random electrical signals. These ambient electrical signals consequently induces randomly fluctuating voltages in radar antennas that we see as clutter, i.e. unwanted signals that decreases the quality of what we try to observe. The ratio between preferable and unwanted signal is called Signal to Noise Ratio (SNR):

$$SNR = \frac{P_R}{P_N} \tag{2.26}$$

Where P_R is given by the radar equation in 2.22 and the noise power P_N is defined as:

$$P_N = k_b T B \tag{2.27}$$

Where k_b is the Boltzmann constant, T is the system noise temperature and B is the bandwidth of the signal the antenna is picking up. The source of the noise varies greatly, as it can be anything from celestial objects in the galaxy, to lightning storms or even tiny man-made circuits. Most noise sources have a white spectrum identical to that of thermal noise, this means that all the noise sources can be represented simply as thermal noise T. The work of Lott et al. [2006] recommends various levels of noise temperature T for different operating frequencies. Figure 2.8 shows the various recommended noise temperatures based on the frequencies being used.

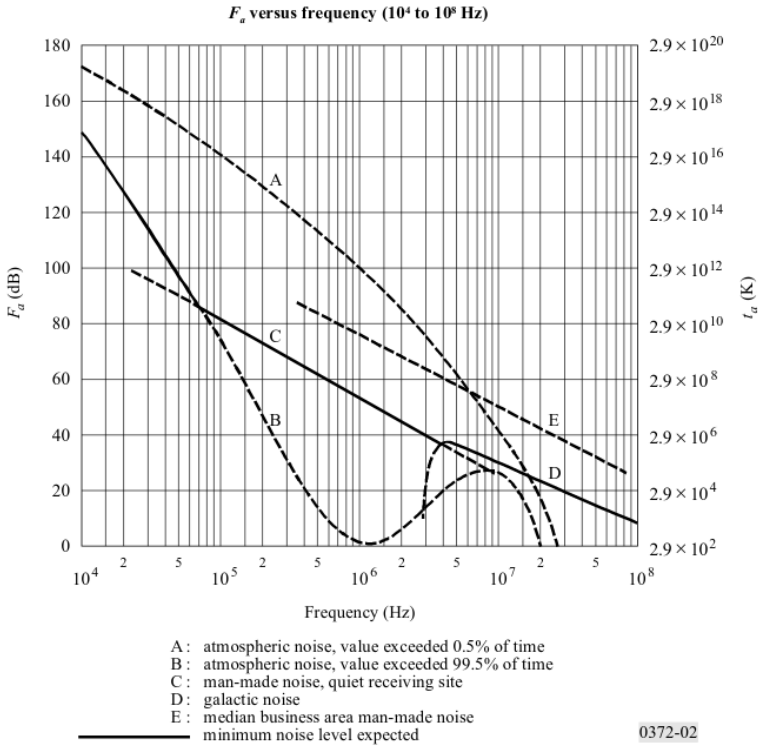


Figure 2.8: Sky noise temperatures from different sources based on the ITU P.372-8 recommendation. Plot adapted from Lott et al. [2006].

2.5 Radar Signal Processing

This section presents some concepts and techniques used in the radar signal processing of the new Longyearbyen ionosonde.

2.5.1 Fourier transform

The Fourier transform is important because a radar measures a signal in time, but the data a radar produces is often a measured power as a function of frequency. The Fourier transform is a function that can transform a signal in time domain into frequency domain and is what is being used by the new Longyearbyen ionosonde in the signal processing involved in making ionograms.

One of the many ways a Fourier transform can be formulated is this:

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi f t} dt \tag{2.28}$$

The Fourier transform multiplies a time-domain signal $x(t)$ with an analyzing function $e^{-i2\pi f t}$ that represents any arbitrary sinusoid over the integral through Euler's relation. If the function and analyzing function are similar they will multiply and sum to a large coefficient, and if they differ, the coefficient will be small.

Since any signal $x(t)$ is sampled over a finite amount of time, the signal is represented by a discrete set of points, extending from the first sample at t_0 to the N'th sample at t_{N-1}. The Fourier transform on a discrete set of samples is the discrete Fourier transform (DFT):

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi k n}{N}} \tag{2.29}$$

Instead of computing an integral from $-\infty$ to $+\infty$ the discrete signal $x[n]$ is instead evaluated as a summation from the sample at $n = 0$ to the sample at $N - 1$ to determine the discrete frequency coefficients X_k. Since the signal is no longer continuous, it can no longer be evaluated at any frequency and we are restricted to a set of k frequency bins. There are a number of ways to compute the DFT, one of which is the Fast Fourier Transform algorithm (FFT) described by Cooley et al. [1967].

2.5.2 Complex baseband representation

A complex baseband signal is a complex representation of a real-valued time-domain signal, for instance the electric field of an electromagnetic wave or the voltage induced in an antenna due to incoming electromagnetic radiation. The complex baseband signal is a useful concept because it allows us to sample a high-frequency signal with a low sample rate without the loss of information, such an operation is known as Digital Down Conversion (DDC). The opposite of DDC is Digital Up Conversion (DUC), which is used to convert a low-frequency signal to a high-frequency signal.

The process of converting a signal into its complex baseband representation can be understood as shifting the whole spectrum of the signal in negative direction. Consider a signal with a known frequency ω_0 and some arbitrary phase, ϕ and amplitude, A:

$$x(t) = A\cos(2\pi\omega_0 t + \phi) \tag{2.30}$$

By multiplying $x(t)$ with the function $f(t) = e^{-2\pi i\omega_0 t} = \cos(2\pi\omega_0 t) - i\sin(2\pi\omega_0 t)$ we get a complex function

$$y(t) = Ae^{-2\pi i\omega_0 t} \cdot \cos(2\pi\omega_0 t + \phi) \tag{2.31}$$
$$= Ae^{i\phi} \cdot e^{-i(2\pi\omega_0 t+\phi)} \cdot \cos(2\pi\omega_0 t + \phi) \tag{2.32}$$
$$= Ae^{i\phi}[\cos^2(2\pi\omega_0 t + \phi) - i\sin(2\pi\omega_0 t + \phi) \cdot \cos(2\pi\omega_0 t + \phi)] \tag{2.33}$$
$$= \frac{A}{2}e^{i\phi}[1 + \cos(4\pi\omega_0 t + 2\phi) - i\sin(4\pi\omega_0 t + 2\phi)] \tag{2.34}$$

Noting that $x(t)$ and $f(t)$ have the same frequency, we see that by multiplying x and f we get a complex function $y(t)$ that consists of an offset factor $\frac{A}{2}e^{i\phi}$ and an oscillation factor with frequency $2\omega_0$. When filtering this complex signal with a low-pass filter that has a cut-off frequency less than $2\omega_0$ we are left with the offset (the complex baseband signal)

$$z(t) = \frac{A}{2}e^{i\phi} \tag{2.35}$$

The signal, in terms of its amplitude and phase, has been converted to a zero-frequency (DC) component while still retaining the original signal information. The amplitude

$$|A|^2 = 2|zz^*| \tag{2.36}$$

and phase

$$\phi = \arctan\left(\frac{\mathrm{Im}(z)}{\mathrm{Re}(z)}\right) \tag{2.37}$$

remains the same.

When considering a signal with a certain bandwidth we can refer to that signal as a band limited signal. A band limited signal consists of spectral components within a finite band as described by the integrals in equation 2.38:

$$x(t) = \int_{f_0-\frac{B}{2}}^{f_0+\frac{B}{2}} X(f)e^{i2\pi f t}df + \int_{-f_0-\frac{B}{2}}^{-f_0+\frac{B}{2}} X(f)e^{i2\pi f t}df \tag{2.38}$$

The procedure of shifting frequency and low pass filtering can be done on all frequency components of the band limited signal, and the resulting complex baseband signal will become shifted around zero as described by equation 2.39.

$$x_{bb}(t) = \int_{-\frac{B}{2}}^{\frac{B}{2}} X(f - f_0)e^{i2\pi f t}df \tag{2.39}$$

A diagram of a signal processing system that conducts up- and down-conversion to translate signals between real and complex baseband representations is shown in Figure 2.9.

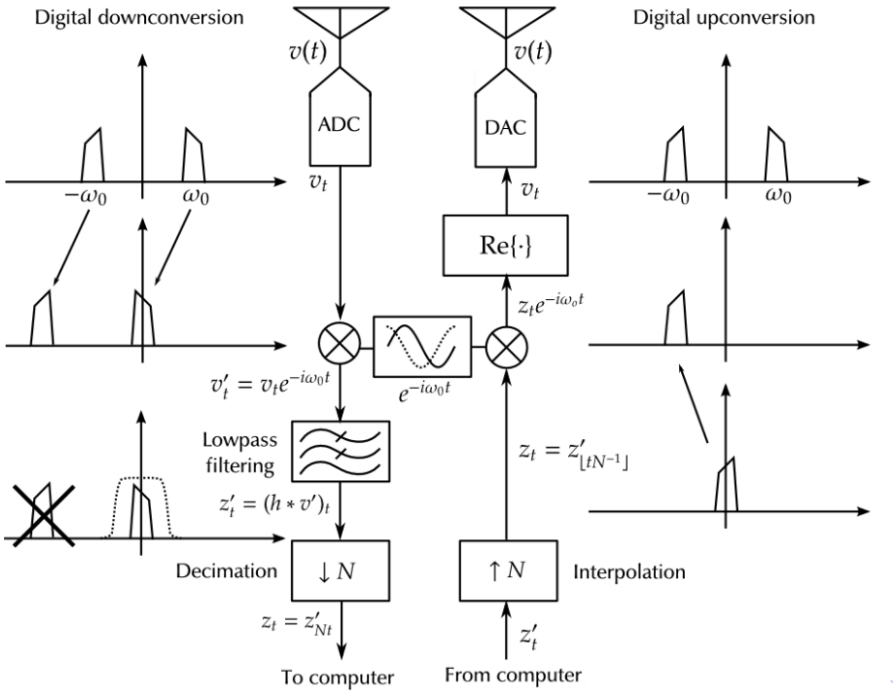


Figure 2.9: Digital down conversion and up conversion used to transform real signals to and from the complex baseband representation (Vierinen, personal communications).

2.5.3 Radar range resolution and pulse compression

The ability to distinguish different reflected signals from the ionosphere is determined by the range resolution. The range resolution depends on the pulse width of the transmit signal:

$$R_r = \frac{cT_p}{2} \tag{2.40}$$

where T_p = pulse length [s] and c is the speed of light [m/s]. The range-time diagram shown in figure 2.10 shows the linear relation between range and time. For the range resolution the factor of 2 arises because the signal needs to travel to the target and back before detection. If the pulse width is reduced, consequently R_r will become smaller and the range-resolution will increase. There is a limit however to how small width a pulse can have. Assuming there is a limit to how much power one can transmit, the shorter pulse one transmits the less power is transmitted and the signal becomes harder to detect. To mitigate this one can transmit several short pulses, but that will have implications for the ambiguous range of the radar. If the pulse is repeated several times before a target backscatter is received, telling which of

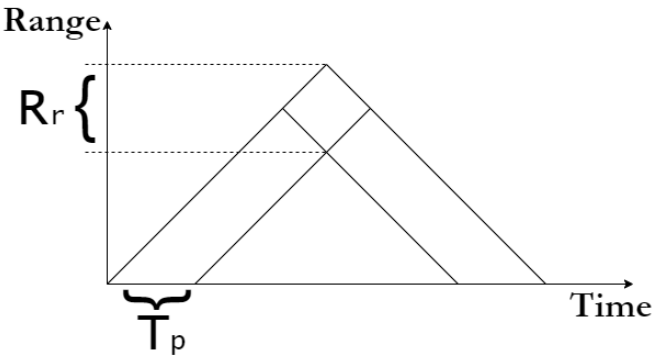


Figure 2.10: The spatial extent of the pulse determines the range resolution.

the repeated pulses got reflected becomes complicated due to aliasing, hence range information becomes ambiguous. There is thus a maximum possible pulse repetition frequency (PRF) in order to measure the ionosphere without the range becoming ambiguous.

Pulse modulation is a technique that is often used to mitigate the issue of ambiguous range while retaining enough power within the pulse. It is a technique where longer pulses are used, but the effects of using shorter pulses are still obtained, this is why the technique is also referred to as pulse compression. The two main categories of compressing a pulsed signal is that of frequency modulation and phase modulation. In phase modulation, the phase of the signal is varied over the duration of the pulse. The technique is implemented by dividing the long pulse into a series of equally long sub-pulses, with each sub-pulse being assigned a phase-shift using a predetermined code. Figure 2.11 shows a binary phase modulation scheme that utilizes two phases to separate the sub-pulses. By convention, + represents a 0° phase shift while a $-$ sign

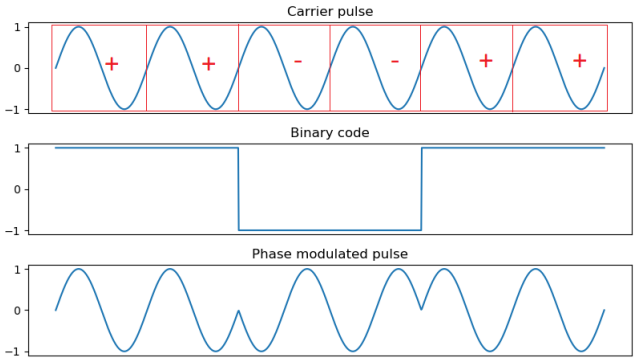


Figure 2.11: Example of a pulse compression scheme that uses two phase offsets to separate the sub-pulses from eachother to achieve higher range-resolution.

represents a 180° phase shift which is the equivalent of multiplying the signal with −1. There are other phase-modulation schemes that uses more than just two phase-offsets to divide the pulse, such modulation schemes are called poly-phase modulation schemes[5]. The effect of dividing the original pulse into separate sub-pulses can be seen in figure 2.12.

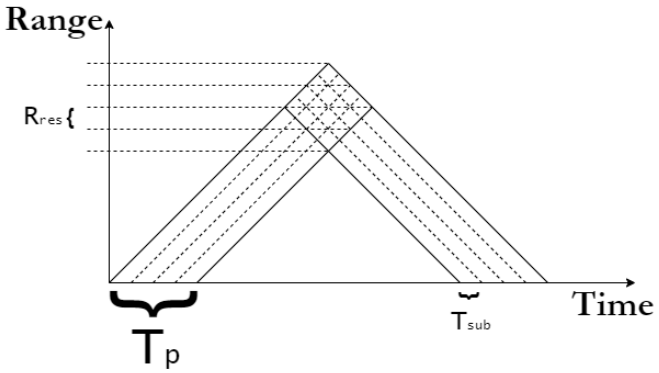


Figure 2.12: Range-time diagram for a transmission with pulse compression.

In a pulse-compressed signal, the range-resolution no longer depends on the pulse-width. It now depends on the bandwidth of the transmitted pulse since the bandwidth determines how many sub-pulses we are able to compress into the original pulse T_p, hence the bandwidth also determines range-resolution. The range-resolution described in terms of the bandwidth is given by:

$$R_{res} = \frac{c}{2B_{tx}} = \frac{cT_{sub}}{2} \tag{2.41}$$

Where c is the speed of light and B_{tx} is the bandwidth of the transmit signal and T_{sub} is the duration of each sub-pulse. For a transmit pulse with a 100 kHz bandwidth, the range resolution would be 1.5 km. In this case, the width of each sub-pulse is 10 μs.

There is however also a limit to the range resolution when using pulse compression. The limitation is determined by the refractive index n given in equation 2.12 where there is a requirement for the signal to retain the same refractive index throughout its spectrum. If the transmit pulse has a bandwidth that exceeds the maximum bandwidth limit then the lower end of its spectrum will undergo a different kind of refraction than the upper part of the signal. The consequence is that the radar is no longer coherent.

5. The well known Barker codes are examples of poly-phase codes. They are often used because of the low sidelobes they produce of the signal in the frequency domain.

2.5.4 Coherent radar signal processing

In order to separate the ranges of the different sub-pulses when receiving a pulse-compressed signal, the technique of linear least-squares is utilized. The measurement equation is equivalent to a convolution operation between the transmitted signal and the echo for each range. In order to estimate the echo as a function of range, the transmitted signal needs to be deconvolved. While it is standard practice in radar literature to use slightly different terminology, such as matched filter or inverse filtering to denote this operation, we have chosen to use standard terminology from linear regression of Kaipio and Somersalo [2006], as it is more general.

The received signal can be expressed as a sum of the received sub-pulses, where each received sub-pulse is a product of the transmit pulse E_{t-r} and an amplitude of the scattered signal V from range r at time t as shown in the range-time diagram in figure 2.13.

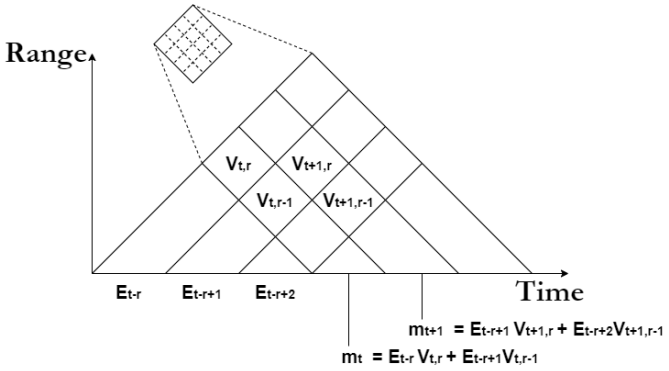


Figure 2.13: Range-time diagram for a pulse compressed signal. The sub-pulses within the transmit pulse are denoted E, while the scattering signal is denoted V.

In general, we can express the received signal as the following sum:

$$m_t = \sum_{r}^{R} E_{t-r} V_{t,r} + n_t \tag{2.42}$$

The received signal m_t is a vector where each element is composed of the product between the transmit pulse E and the amplitude of the scattered signal V at time t and the corresponding range r. The last term n_t represents the ambient noise. If we assume that the scattered signal amplitude remains constant throughout the period T_p in which the pulse propagates through the scattering volume, we can exclude the time dependence t from V and re-write

the received signal as:

$$m_t = \sum_r^R E_{t-r} V_r + n_t \tag{2.43}$$

Where V now only is a function of range r. Radar targets which have a constant scattering amplitude over a pulse period T_p is by convention considered as a "coherent" target. Radar systems that rely on this constant scattering amplitude are thus referred to as coherent radars.

The validity of the assumption that the scattered amplitude is constant throughout the long pulse-period T_p depends on the Doppler width of the scattered spectrum. If the scattered signal has a broad Doppler width it means that the scattering medium changes rapidly relative to the duration of the pulse and the target is deemed "incoherent".

When the time dependence of the scattered amplitude V is removed, equation 2.43 becomes a convolution equation. In the case of ionosonde measurements of the ionosphere, we can typically make this assumption. Due to the linearity of the convolution equation, the representation of the received signal can be done in the form of matrices:

$$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{v} + \boldsymbol{n} \tag{2.44}$$

Where $\boldsymbol{A}$ is a Toeplitz matrix that relates the transmit signal and the constant scattered signal $\boldsymbol{v}$ [Vierinen et al., 2016]. Because the scattered signal is assumed to be constant over the pulse-duration, the linear-least-squares method can be used for obtaining an estimation of the scattered signal $\boldsymbol{v}$:

$$\hat{\boldsymbol{v}} = (\boldsymbol{A}^H \boldsymbol{A})^{-1} \boldsymbol{A}^H \boldsymbol{m} \tag{2.45}$$

The convolution theorem states that a convolution in time-domain constitutes point-wise multiplication in the frequency-domain. This means that equation 2.43 can also be expressed as the product of the Fourier transforms of each component:

$$m(\omega) = E(\omega)V(\omega) + n(\omega) \tag{2.46}$$

The scattered signal can then be obtained by dividing the received signal with the transmit pulse in the frequency domain:

$$\frac{m(\omega)}{E(\omega)} - \frac{n(\omega)}{E(\omega)} = V(\omega) \tag{2.47}$$

In order for this estimate to work, the discrete Fourier-transform of the transmitted signal $E(\omega)$ cannot have zeros in its frequency domain. This is the case for well designed radar transmit signals.

2.5.5 Spread spectrum sounding

When we think about a radio signal being communicated between different stations, be it signals travelling from a cellphone to a base station or signals transmitted from a radar and reflecting of the ionosphere, we generally think of the signal as fixed in frequency, with a certain bandwidth, that undergoes some arbitrary modulation be it in phase, amplitude, frequency or pulse. The underlying property of these signals, whatever modulation they may undergo, is that most of the power (i.e. amplitude over time) is contained within the bandwidth of the signal. Spread spectrum is a technique, or rather, an umbrella term for a number of techniques that, as the name implies, spreads the power of the signal across the spectrum.

Frequency-hopping is a spread spectrum technique that spreads the power of the signal by sectioning the bandwidth of the original signal into smaller sub-bands and transmits the sub-bands as separate signals. It is a form of coherent radar communication where the receiver must know in advance how to interpret the signals that arrive, this is done by having a code known to both the transmitter and the receiver that determines the order of the smaller sub-bands. Dividing the original signal is done by having the smaller sub-bands occupy a smaller bandwidth within the original bandwidth, where each sub-band is separated by having a different carrier frequency. Signals that undergo frequency-hopping are less susceptible to interference because they only occupy a narrow band over a small amount of time, which means they also cause less noise to other signals. Different frequency-hopping signals can also share the same band as long as their codes differ from each other, enabling the use of multiplexing which is a modulation technique that allows several users on the same frequency band.

In the case of the new ionosonde, the need for a wide band of available frequencies is necessary because of the changing plasma parameters in the ionosphere that in turn changes the resonance and cut-off frequencies. The frequency band we therefore want to use contains most of the HF-spectrum, ranging from ~2 MHz - 30 MHz. By probing the ionosphere using small segments of 100 kHz wide signals we are effectively performing frequency-hopping spread spectrum within the HF-band. This means that our system will have little effect on other users within the HF-band, and makes it more resilient to other noise sources.

2.6 Software Defined Radio

A radio is a piece of technology that is characterized by the ability to transmit or receive (or both) radio waves. Radio waves is an umbrella term for electromagnetic waves ranging from frequencies between 30 Hz and 300 GHz. Whenever radio waves are used in communication, information is embedded into the radio waves by changing some parameter of the waves, meaning that either phase, frequency or amplitude of the waves are changed. This controlled change of the radio wave is referred to as *modulation*.

Artificially made radio waves originates in most cases from an antenna where the power from an alternating current is converted into loops of electric field lines which propagates away from the antenna close to the speed of light. These loops of electric fields are what makes up the radio waves. A radio wave is thus at its core no more than an alternating current where the power is propagating away from the conductor in the form of closed electric field line loops. The process of modulating radio waves is then nothing else than changing the properties of an alternating current, which as is already known, also has a waveform-representation consisting of phase, frequency and amplitude.

In the early days of radio technology, modulation of radio waves was a relatively tedious process where each step of the signal processing involved in changing the waveform to a specific state required another specific physical hardware component. This meant that for each desired change in either frequency, phase or amplitude, it usually required the change of some hardware component in order to achieve this. Components such as amplifiers, mixers, switches and filters had to be physically implemented through hardware and there were specific amplifiers, filters and mixers one had to use for such and so type of modulation one wanted to perform. All in all a very resource demanding process.

With the evolution of the digital age, these hardware components are becoming more and more obsolete. Hardware is replaced by software, and where one previously had to spend time and space connecting various hardware components, one can nowadays use a simple base-portion of hardware, capable of changing its properties through software in order to perform various modulation techniques. Such a system is called a Software Defined Radio. A software defined radio is a system where radio components such as converters, mixers, filters, modulators, and detectors are implemented via software, most signal processing is thereby performed by the central processing unit of a personal computer. The main benefit of SDR is that radio communications systems are able to implement more features using only a single set of hardware, with all the complicated application specific logic implemented using easily transformable software.

3

Longyearbyen Ionosonde Design & Implementation

This section presents the design of the Longyearbyen ionosonde, the requirements that the system has to fulfill, and the implementation of the design. In the design we define the high-level constraints on the system in order to fulfill certain aspects we want to achieve. In the requirements we present the implications the design brings that need to be adhered to in order for the design to be implemented. The implementation presents how the design was realized while fulfilling the requirements.

3.1 Design

A design is often something that provides some predetermined descriptions or conditions of an end-product. In many cases however, constructing a system without a prior design may also be considered a design activity. This project is a combination of both those types of design. On one hand we have a set of predetermined design choices or conditions that we already impose on the system before development. On the other hand, some requirements or design choices come as a result of the implementation or developing process. It is difficult to make any precise distinction between the two, but I will attempt to separate them by providing, in the following section, a description of our

predetermined design choices as opposed to design choices that have emerged as a consequence of the development process.

3.1.1 Open source code

As a part of our predetermined design, we want the software of the software defined ionosonde to be available to the public through open source. Open source means that the code is available to the public for their own use and modification. There are several reasons for having an open source code, one of which is the benefit of other users being able to implement a system of their own, without having to spend resources on a separate piece of software. The main potential of this system lies in its use in parallel with other similar instruments. Ideally, one wants to implement as many of these instruments in order to unlock the full potential in terms of data acquisition. As an example, if one in the future implements several ionosondes or equivalent instruments to measure Doppler-shifts one can ascertain several velocity components of the reflecting medium and thus be able to describe its motion.

Additionally, having the same software distributed across several instruments conducting the same measurements means that the data will be easier to compare and incorporate with data-sets from other instruments.

Having an open source code also enables further software development by other parties, which means potentially more people working on the same project to achieve a common goal.

3.1.2 Low cost, commercially available hardware

Alongside an open source software, having hardware components that are easy to obtain, economically friendly and simple to use makes the implementation easier and simplifies the realization of additional instruments. We therefore include these characteristics in our design.

3.1.3 Local considerations

Svalbard is a harsh environment with regular blizzards, strong winds and freezing temperatures. Coupled with a low level of infrastructure outside the settlements, this poses some challenges in terms of mechanical design that need to be considered. We need the system to survive mechanically over longer periods of time in these harsh conditions, and also to have the system only depend on mains power. This design choice may not be considered as

something that should be predetermined, but rather emerge as a consequence of the location we are implementing the system at. However, having the predetermined design already able to deal with harsh conditions makes it more applicable to other, potentially even more hazardous locations. We therefore state the need for a resilient system in our predetermined design.

3.1.4 Power consumption

We aim for the system to have as low power consumption as possible, possibly even down to a maximum power usage of $< 100\,\mathrm{W}$, so that the system can potentially be operated with solar or wind power in future versions or extensions of this design. The lower the power consumption of the system, the easier it is to implement alternative energy sources. Having the system self-sufficient would increase the capability and lower the dependence of external sources on the system, potentially enabling the implementation of the instrument in regions were there is little to no infrastructure.

3.1.5 Form factor

We want the ionosonde to have a relatively small form factor, to keep the cost and footprint of the system small. A small form factor will make it possible to use the design for dense networks of ionospheric sounders in the future, make the physical implementation simpler and will allow scientists to set up an ionosonde with relative ease.

Consequently, we want the transmit and receive antennas to be relatively small in size. The computers and radio hardware will be stored inside two intermodal containers already located on site, which have connection to the mains power grid. One for the transmitting site and one for the receiver site.

3.1.6 Software user interface

The set of frequencies used and duration of frequency measurement needs to be user-configurable. This is mainly to provide ease of use for other users not necessarily familiar with the software implementation, but also because it allows users to adapt to various frequency licenses they may be operating under. The frequency license is ultimately what determines the available frequencies for the system and is provided by the governing communications authority.

The operation of the software also need to be fully autonomous once it is employed.

We have in this section presented the design choices we impose prior to the system development stage. We will in the following section go into the further design of this process.

3.2 System Requirements

Following the general system design, we now come to the more hands-on development part of the design that emerges as the development progresses, that is to say; we now come to the part of the design process that is not predetermined.

We refer to these emerging design elements as requirements rather than design choices, mainly because they are not defined in the set of predetermined design choices, but rather emerges as design elements we need to adhere to as developers.

With this in mind, let us first consider the main objective of this thesis. The main objective is to develop an ionosonde. This objective should be seen in the context of implementing a relatively undemanding radar to be utilized in a distributed array(s) of small instruments as a contribution to the ongoing DASI initiative [Foster, 2005; on Distributed Arrays of Small Instruments for Research et al., 2006].

In terms of design and requirements, there are two sides to this objective. From a scientific standpoint, this project entails the construction of an ionosonde so that researchers have access to high latitude ionospheric data. From an engineering standpoint we can also say that the purpose of the project is the development of a novel instrument with a certain design such that it is easy to reproduce, is easy to maintain and operate, and can potentially operate independently in remote regions.

There is therefore a distinction to be made between requirements that emerge from a scientific standpoint and requirements that are solely technical, i.e. requirements that emerge from an engineering standpoint.

In the following paragraph we discuss some high-level scientific requirements on the system. The science requirements does not dictate the implementation of the ionosonde, but provides a guideline for what we want to obtain with the ionosonde from a research perspective.

The main requirement of an ionosonde is the ability to measure electron-densities. Expanding upon that, the required range/altitude over which we

need to be able to measure electron-densities is within the ionosphere, thus we need to be able to measure electron-densities within a range of $\sim$ 80 – 1000 km. Additionally, we want to incorporate the ability for oblique sounding so that when the system is incorporated into a system of several ionosondes, it can be set up in a way to measure travelling ionospheric disturbances (TIDs) [Verhulst et al., 2017]. For this reason we expand the range extent to 1500 km even though the measurable ionosphere from an ionosondes point of view only extends up to the F2-region of approximately 400 km altitude. These requirements define what we want to observe and where we want to observe it. Furthermore, another requirement from a science standpoint is the ability to produce ionograms at-least once every minute in order to observe the time-evolution of the backscattering medium.

By defining these requirements such that from a scientific standpoint the instrument will succeed, we are at the same time introducing certain technical requirements that will have to be dealt with. The following sections presents the technical requirements that our designed system has to fulfill in order for us to achieve our goal of implementing a small form factor, low powered radar system for doing HF-probing of the ionosphere with components that are affordable and available "off the shelf".

3.2.1 Frequency

The peak plasma frequency of the ionosphere is typically below 15 MHz. Using equation 2.1, a 15 MHz plasma frequency would result in backscatter signals from the ionosphere with densities up to $n_e \approx 3 \times 10^{12} m^{-3}$. As shown in figure 2.4, this is more than adequate to probe the high latitude ionosphere.

With oblique propagation and scatter from sporadic E, ionospheric returns can be obtained from significantly higher frequencies than this. However, we will focus on vertical sounding, where the highest operating frequency is determined by the peak F-region plasma frequency. Ideally, we would like to measure radar echoes with as many frequency channels as possible between 0 and 30 MHz.

Before one can implement a transmitter, a frequency license has to be obtained from the Norwegian Communications Authority (NKOM) that distributes the limited frequency channels in the electromagnetic spectrum to various users. In the frequency application we applied for as many channels as possible between 0 and 30 MHz with a width of 100 kHz[1]. The license is required in order to avoid cluttering signals of other users in the spectrum, and likewise avoiding

1. The choice of 100 kHz will be discussed in the subsection on Bandwidth Limitations.

deterioration our own signal from interference with others. NKOM keeps track of all the transmitting users of the electromagnetic communications spectrum and licenses users of the different frequencies to keep the spectrum from becoming overpopulated.

From the frequency license application, we were offered a set of frequency channels between 0 MHz and 30 MHz given in figure 3.1. Each band is 100 kHz wide, with 4 exceptions being 50 kHz wide and 1 being 30 kHz wide. In total 33 frequency channels, with most of the channels being within 2-8 MHz as this is where most of the reflections is expected to occur. The higher frequencies are more useful for observations of sporadic E.

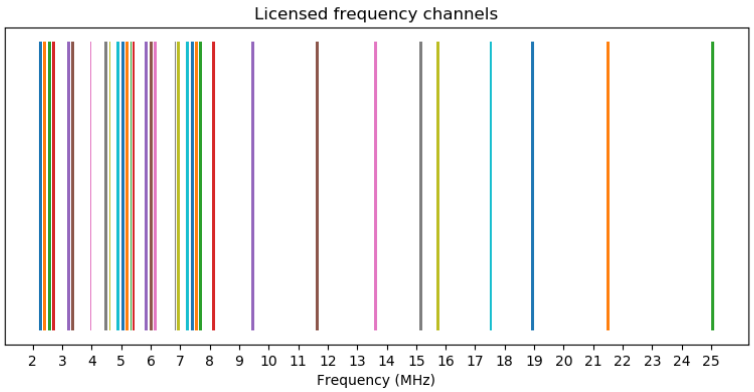


Figure 3.1: Frequency channels granted by NKOM. In total 33 channels with 100 kHz bandwidth, with 5 channels between 3.95 MHz and 6.85 MHz having a bandwidth of 30 kHz and 50 kHz.

The license also states that the effective radiated power (ERP) may not exceed 10 W (40 dBm). The ERP of our system is calculated in section 5.2.2 using measurements obtained in chapter 4.

3.2.2 Bandwidth limitations

There are two reasons for the choice of 100 kHz bandwidth channels. One of them has to do with the dependence of incident wave frequency on the refractive index given in equation 2.12. The refractive index across the bandwidth $f - B/2$ to $f + B/2$, where f is the carrier signal frequency and B is the bandwidth of the signal must be, to a first approximation, constant. If this is not the case, the lower part of the spectrum will have a different refractive index than the upper part of the spectrum which will potentially lead to parts of the same signal travelling in different directions and consequently arrive at the receive at a different time than the upper part of the signal. This would cause

an error in the received echo in the form of blurring in range and consequently decrease the range resolution.

The other reason for the choice of bandwidth stems from the discussion about range resolution for a pulse-compressed signal in section 2.5.3, where we stated that a 100 kHz bandwidth pulse-compressed signal has a 1.5 km range resolution. This is a reasonable range resolution for probing the E-region ionosphere (comparable to the scale height in the E-region).

3.2.3 Antennas

Due to the small form factor requirement in the design, we will not obtain a very directional antenna. However, the antenna needs to be able to perform with some efficiency within the frequency range between 0 and 30 MHz. Additionally, the transmit and receive antennas will be be separated by roughly 100 m. Because of this, the radar is assumed to be mono-static since any difference in range between targets in the ionosphere, the receiver and transmitter are well within the range-resolution of the signals.

Since the ionosonde site already has implemented antennas, the transmit antenna is predefined to be a folded dipole of length 25 m. The antenna is setup as an inverted V, using a vertical mast of about 9 m. We can approximate the antenna as a half-wavelength dipole on the higher end of the frequency range (approximately above 6 MHz) and a short dipole on the lower end of the frequency range.

To estimate the transmit antenna gain we consider the directivity D and efficiency ϵ introduced in section 2.4.1. The directivity of a perfect half-wave dipole is $D = 1.64$, while for a short dipole we have $D = 1.5$. The folded dipole is not much different in terms of directivity compared to a regular dipole, so we assume that $D = 1.64$ at frequencies above 6 MHz and $D = 1.5$ at frequencies below 6 MHz.

The antenna efficiency cannot be properly ascertained unless highly involved testing is conducted. What we do know however is that ϵ is somewhere between 0 and 1. By assuming that the impedance of the antenna is well matched to the impedance of the feedline, and that the antenna is a decent conductor, we make an educated guess of an antenna efficiency of $\epsilon = 0.6$. Assuming more power loss due to impedance mismatching we obtain an efficiency of $\epsilon = 0.1$.

One of the ways to measure how well matched the impedance of the antenna is to the impedance of the feedline is by measuring the standing wave ratio

(SWR) defined by equation 3.1:

$$\text{SWR} = \frac{1 + \sqrt{P_r/P_f}}{1 - \sqrt{P_r/P_f}} \tag{3.1}$$

where P_f is forward power going into antenna and P_r is reflected power from the antenna. When transmitting at the resonant frequency of a halfwave dipole antenna, the SWR is close to 1:1. Since we transmit across a wide spectrum of frequencies, the transmit signal will not always be close to the resonant frequency of the 25 m antenna (~6 MHz), which is why we set the efficiency ϵ to 0.6.

With the assumptions made in the previous paragraphs and using equation 2.23, we estimate the gain G_T to be somewhere between 0.15 and 1, with the lower values of gain achieved on lower frequencies.

On the receiving end is a small loop antenna with 1 m diameter which is considered to be a good receiver in the HF-spectrum while being less susceptible to noise than dipole receivers. The gain of this antenna will be incorporated into the discussion about its effective aperture A_R in the upcoming section; 3.2.4.

3.2.4 Approximating transmit power

In order to determine the minimum amount of necessary transmit power, we need to determine the signal-to-noise ratio of the ionospheric radar echo. For this, we need to use the radar equation to obtain an estimate of necessary received power P_R and a noise temperature factor T to estimate noise power. Considering the radar equation, we assume mono-static radar setup, with a lossless transmission medium for the case where multi-path interference is neglected as described in chapter 2, equation 2.22.

The radar cross-section of the ionosphere is relatively large. For sake of simplicity, the assumption that the ionosphere is a perfect reflector is applied. The size of the illuminated portion of the ionosphere therefore determines the radar cross-section. By assuming that the half-power full width opening angle of the transmit antenna is 45°, we obtain an approximate expression for radar cross-section σ as follows:

$$\sigma = \pi R^2 \sin^2(\pi/4), \tag{3.2}$$

where R is the range to target. Therefore, the radar equation in 2.22 can be

rewritten as:

$$P_R = \frac{P_T G_T A_R}{32\pi R^2} \tag{3.3}$$

which is how we define the received signal power.

To estimate the effective receiver aperture A_R of equation 3.3 we consider the antenna factor of the receiver. The design uses an ALA-1530 magnetic loop antenna, as this specific antenna was available. Similar electrically small active antennas are often used in other ionosonde designs [e.g., Reinisch, 2009]. Based on measurements provided by the antenna manufacturer, the antenna has an antenna factor of 4.5 dB/m at 2 MHz and 2 dB/m at 14 MHz [Wellbrook, 2020a]. When expressed linearly instead of logarithmic, the antenna factor is 1.68 m^{-1} at 2 MHz and 1.2 m^{-1} at 14 MHz. Assuming that the impedance of the antenna is 50 Ohm, the antenna factor A_f can be related to the effective receiver aperture A_R as follows[2]:

$$A_R = \frac{2.75^2}{A_f^2} \tag{3.4}$$

Thus, the effective aperture A_R of the receiver antenna is 2.68 m^2 and 5.25 m^2 when receiving at 2 MHz and 14 MHz respectively.

The receiver noise power is given by equation 2.27. Typical values for system noise temperature T is approximately between 10^6 - 10^7 K based on the ITU P.372-8 recommendation of Lott et al. [2006] seen in figure 2.8. Because the sky noise temperature is so high, we can discount the effects of additive receiver noise due to receiver electronics, as the sky noise power is the dominating effect, we merely need the receiver noise temperature to be significantly less than the sky noise temperature. As described earlier, due to dispersion of radio waves[3], we use $B = 100$ kHz as the upper limit of bandwidth.

In order to be able to successfully measure the ionospheric echo, we require the ratio of signal-to-noise to be above 10. To obtain a minimum requirement on the transmit power, the worst case radar parameters is applied to the radar equation with effective receiver aperture $A_R = 2.68\, m^{-1}$, transmit gain $G_T = 0.15$ and system noise temperature $T = 10^7\, K$.

Assuming that the ionospheric reflecting layer is at an altitude of $R = 200\, km$, the minimum transmit power required to obtain a signal-to-noise ratio of 10 is

2. Relation obtained from Wikipedia: `https://en.wikipedia.org/wiki/Antenna_factor`
3. variation in refractive index as a function of frequency

expressed as:

$$\frac{P_R}{P_N} = \frac{P_T G_T A_R}{32\pi R^2 k_b T B} = 10 \tag{3.5}$$

By applying the different values and solving for the transmit power P_T it follows that:

$$P_T = \frac{320\pi R^2 k_b T B}{G_T A_R} = 1380 \text{ W} \tag{3.6}$$

This assumes that we are using a short pulse of length $T_p = B^{-1} = 1/100\,kHz = 10\,\mu s$. The relation between B and T_p can be seen in equation 2.41. It is possible to build a radar that transmits very high power short transmit pulses. In fact, some ionosondes take this type of an approach. However, building a high power transmitter is expensive. In order to reduce the power requirement we can rely on the fact that the ionospheric echo is a coherent radar target. By coherently integrating the radar echo as described in section 2.5.4, we can effectively increase the amount of transmitted power.

For example, if we use a longer transmit pulse of length τ_p, the total amount of transmitted power becomes $P_{tot} = B\tau_p P_T$, where P_T is peak transmit power, B is bandwidth and τ_p is the length of the transmit pulse. Applying this relation to equation 3.6 results in the following form of the radar equation:

$$P_T = \frac{320\pi R^2 k_b T}{G_T A_R \tau_p} \tag{3.7}$$

The usable pulse length τ_p depends on the Doppler bandwidth of the radar echo. The pulse length τ_p is effectively the sample spacing for the received radar echo.

The Nyquist sampling criteria for complex valued signals states that a signal with a bandwidth of B_d needs to be sampled at minimum with a sample spacing of $\tau_p = B_d^{-1}$. This means that if we assume that the radar echo has a Doppler bandwidth of B_d, we can use a pulse length up to $\tau_p = B_d^{-1}$.

An example of the Doppler bandwidth of an ionospheric echo at HF frequencies is shown in figure 3.2.

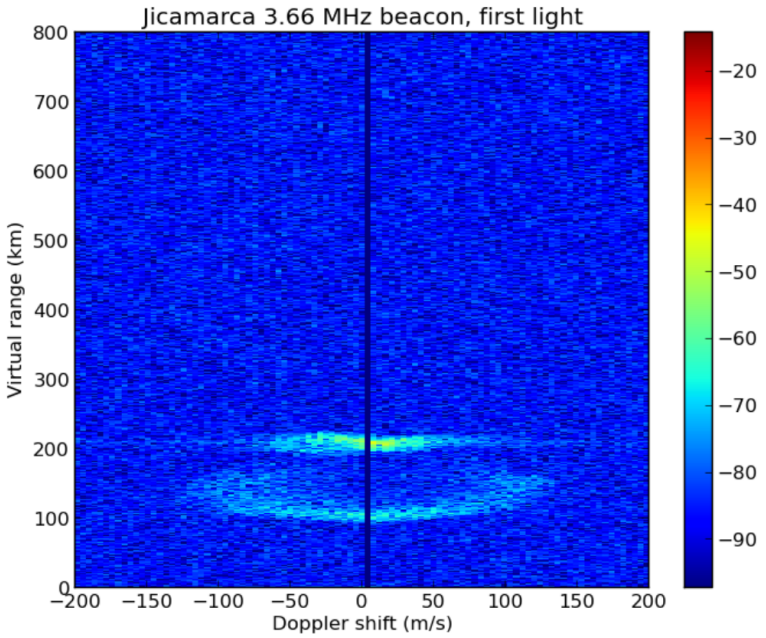


Figure 3.2: An example of the range-Doppler distribution of radar echo power at HF, using an HF radar located at the Jicamarca radio observatory in Peru (Vierinen, personal communications). From the figure, we can estimate that most of the radar echo power at around 200 km altitude is localized in a ≈ 20 m/s Doppler bandwidth.

The echo just above 200 km altitude has a Doppler width of approximately $\Delta v = 20$ m/s and echoes between $\Delta v = 100$ m/s Doppler shifts are observed. Using this figure, we can estimate the upper limit of the Doppler width to be $\Delta v = 100$ m/s. The Doppler bandwidth in Hz for a mono-static radar can be obtained using:

$$\Delta f = \frac{2 f_{\text{radar}} \Delta v}{c} \tag{3.8}$$

For $\Delta v = 100$ m/s one obtains $\Delta f = 2.5$ Hz, as the radar transmit frequency was 3.66 MHz. This means that the maximum coherent integration period is approximately $\tau_p = 1/2.5\,Hz = 0.4\,s$. With a more modest $\Delta v = 20$ m/s, the Doppler bandwidth would be approximately $\Delta f = 0.5$ Hz and the maximum coherent integration period would be 4 s.

To obtain an estimate for a valid pulse width τ_p, the assumption that the ionospheric echo has a Doppler bandwidth less than 2.5 Hz is made. This means that $\tau_p = 1/2.5 = 0.4$ s. In this case using equation 3.7, the minimum required transmit power is:

$$P_T = 30 \text{ mW} \tag{3.9}$$

The value for minimum required transmit power of 30 mW is obtained assuming the worst case scenario, where the estimated values for transmit gain G_T, receiver antenna effective aperture A_R, system noise temperature T, and Doppler bandwidth B_d were for the least favourable. If more favorable values were used, for example $G_T = 1.0$, $A_R = 5.25$, $T = 10^6$, and $B_d = 0.5$ ($\tau_p = 2$), the resulting minimum required transmit power becomes:

$$P_T = 5.3 \ \mu\text{W}. \tag{3.10}$$

This can be thought of as the lower limit for transmit power, where we will still obtain echoes at the more favorable part of the radar performance parameter range.

To conclude, the radar needs to transmit at least 30 mW of power when using coherent signal processing. If we do not use coherent signal processing and rely on 10 μs pulses, we would need to transmit 1.4 kW of power.

3.2.5 Maximum range extent

The F-region of the ionosphere is typically below 500 km. However, the group velocity ($v_g = \frac{d\omega}{dk}$, where k is the wavenumber) of electromagnetic waves in the HF-band can be significantly smaller than the speed of light in vacuum. Additionally, ionosondes often measure multiple hops of the electromagnetic wave between the ground and the ionosphere. An example of a typical high latitude ionogram in shown in Figure 3.3. It shows that most the echoes are obtained at ranges below 1500 km, assuming a group velocity corresponding to the speed of light in vacuum.

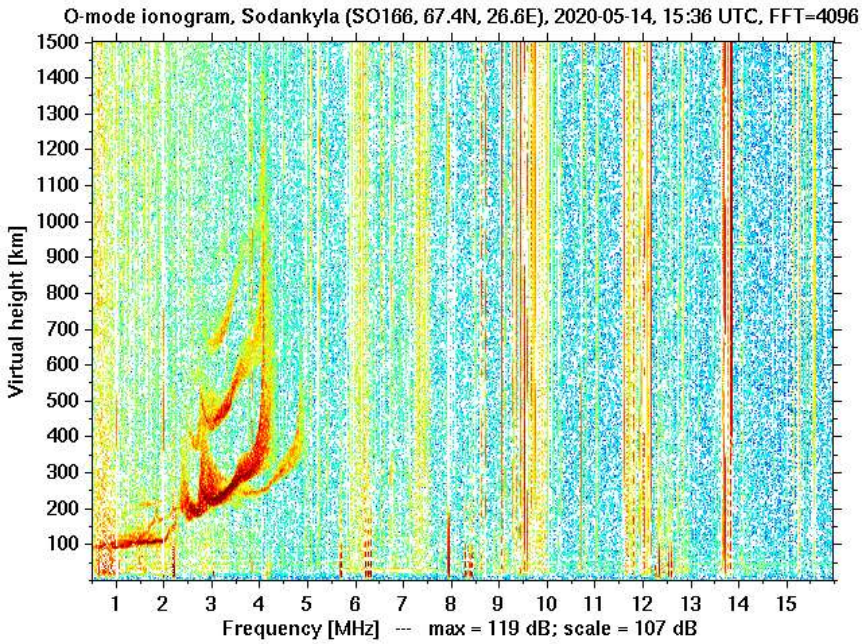


Figure 3.3: Example of an ionogram obtained using the Sodankylä Geophysical Observatory ionosonde. Virtual range along the y-axis and frequency along the x-axis.

At frequencies slightly beyond 4 MHz, the range of the received signal apparently extends to heights up to 1000 km. This is due to the decreasing group velocity of the electromagnetic wave as it approaches the peak-density. Due to this effect, we require the maximum measurable range extent to be at least 1500 km, but preferably higher to accommodate for potential future use of oblique sounding which is based on horizontal propagation.

3.2.6 Maximum Doppler extent

The maximum Doppler extent, or the widest Doppler width as discussed in section 3.2.4, will probably occur when the HF radar measures echoes from ionospheric irregularities in the E-region. Radar echoes from these irregularities are described by the Farley-Buneman theory, and the echo Doppler width in this case is determined by the ion-acoustic velocity, which is approximately ±600 m/s [Milan et al., 2003]. In order to study such irregularities, the measurable Doppler width extent needs to be up to 1200 m/s. Using equation 3.8, this corresponds to a Doppler bandwidth of 29 Hz which means that the maximum coherent integration period for studying such irregularities is $\tau_p = 1/29\,\text{Hz} \approx$ 34 ms.

3.2.7 Timing

The timing accuracy is related to the range measurement error. If we assume that the timing error is $\Delta\tau$, this is translated to a range measurement error of:

$$\Delta R = c\Delta\tau, \tag{3.11}$$

where ΔR is the error in range measurement. Our aim is to measure the range of the ionospheric echo to within 3 km accuracy at minimum, but preferably much better than this. This translates to a upper limit for timing error of $\Delta\tau < 10\mu s$.

3.2.8 Transmit signal

As discussed in chapter 2, the ionosonde will rely on pulse-compression, and we envision that the system can be used in a scenario where multiple independent transmitters can be operating. We need also flexibility in choosing the transmitted waveform to adhere to frequency license requirements.

Requiring that the radar is able to transmit an arbitrary frequency or phase modulated waveform with up to 1 MHz bandwidth at center frequencies anywhere within the frequency band of the ionosonde will enable the ability to perform frequency hopping spread-spectrum sounding[4] while retaining range resolution.

In the scenario where there are multiple transmitters, we need to be able to distinguish between signals transmitted by different transmitters, similar to how global positioning system satellites can be distinguished from one another using the fact that each satellite transmits a different waveform that has a low correlation with the waveforms transmitted by other satellites.

3.3 Implementation

This section presents the implementation of the Longyearbyen Ionosonde. Using the design and adhering to the requirements as formulated in previous sections, a functioning ionosonde prototype was developed using the implementation described here.

As the implementation is done using Software Defined Radio (SDR), an introduction to the SDR hardware is presented here.

4. Spread spectrum is described in section 2.5.5

3.3.1 USRP - Universal Software Radio Peripheral

The Universal Software Radio Peripheral (USRP) is the name of a range of software defined radios (SDRs) developed by the Ettus Research company. Many of their designs are open source hardware, and all of them are supported by the open source USRP Hardware Driver software (UHD), also developed by Ettus Research.

The UHD is the piece of software that provides the interface between the hardware within the USRP and the computer. The UHD software provides the necessary communication and controls to carry out modulation and transmit or receive radio signals. An image of the USRP N200 model is shown in figure 3.4.

Figure 3.4: The USRP N200 software defined radio, developed by Ettus Research. Image downloaded from `https://www.ettus.com/all-products/un200-kit/` in April 2020.

The USRP N200

The USRP N200 is the USRP-model used in the current revision of the Longyearbyen ionosonde. It features a modular design that enables the use of various USRP-daughterboards connected to the main motherboard which gives it a frequency range from DC to 6 GHz.

In the current revision of the Longyearbyen ionosonde, a BasicTX daughterboard is used in the transmitter and an LFRX daughterboard is used in the receiver. With the use of the BasicTX daughterboard in the transmitting N200 and a LFRX daughterboard in the receiving N200, the frequency range of the

Longyearbyen ionosonde ranges from 1 MHz to 33 MHz[5].

Figure 3.5 shows a schematic of the USRP N200, highlighting the main signal paths.

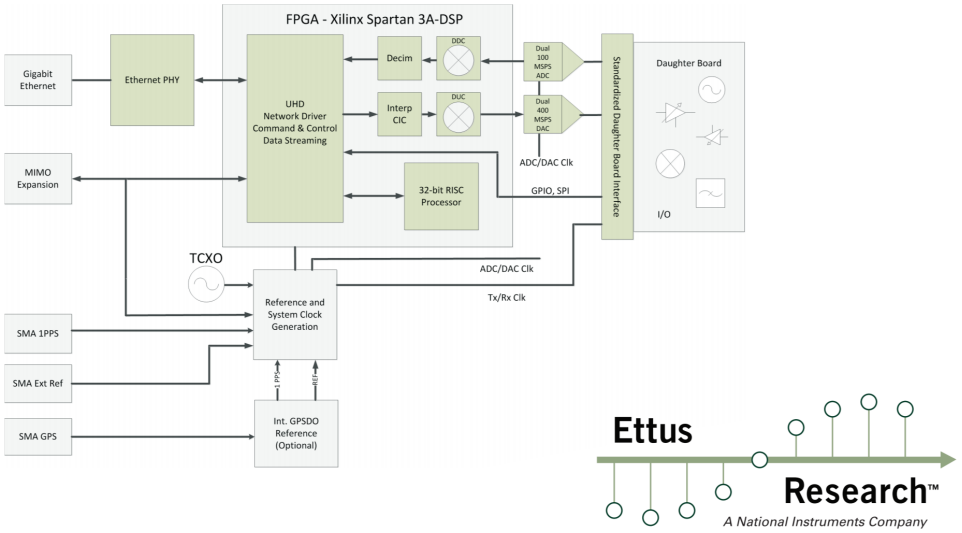


Figure 3.5: Schematic showing the architecture of the USRP N200. Image downloaded from `https://www.ettus.com/wp-content/uploads/2019/01/07495_Ettus_N200-210_DS_Flyer_HR_1.pdf` in April 2020.

The main motherboard is a Field Programmable Gate Array (FPGA) of the type Xilinx Spartan 3A-DSP. From the ethernet connection is data either fed from or provided to the computer by the UHD command & control unit within the FPGA. The FPGA also up-converts or down-converts signals for either transmission or receiving respectively. Between the FPGA and the daughterboard are Digital to Analog (DAC) and Analog to Digital Converters (ADC) which connects to wideband transformers within the daughterboard. The wideband transformers ensures that the impedance of the signal coming from the DAC/ADC-converters matches that of the 50-ohm impedance of the coaxial-connections on the RF-signal ports.

The synchronization of the various processes in the USRP is controlled by a reference and system clock generation unit. The reference clock system may take as input some periodic external timing coupled together with a 1 Pulse Per Second reference signal in order to synchronize all the steps performed by the different blocks. It may also use a GPS Disciplined Oscillator (GPSDO)-kit which couples a 10 MHz reference signal with the global GPS standard time

5. The actual allowed usable frequency range is within the boundaries set by the frequency license as described in section 3.2.

provided by a GPS antenna. The GPSDO-kit coupled together with the GPS antenna provides the best clock accuracy and is the option in use on the USRPs utilized in the Longyearbyen Ionosonde. The GPSDO provides a timing accuracy of 50 ns, which is well within the required 10 μs.

3.3.2 Architecture

The main architecture of the ionosonde consists of a transmitter and a receiver. The transmitter uses Software Defined Radio (SDR) to transmit electromagnetic waves between 1 MHz and 30 MHz. The receiver also utilises SDR to analyze the received signal power as a function of return delay and frequency to produce an ionogram and measure the range-Doppler distribution of the received echoes. Figure 3.6 shows a block-diagram of the system architecture with the transmitter on the left and the receiver on the right. The transmitting subsystem contains more components related to amplification and monitoring of the signal to ensure that the signal power levels are kept within a required range.

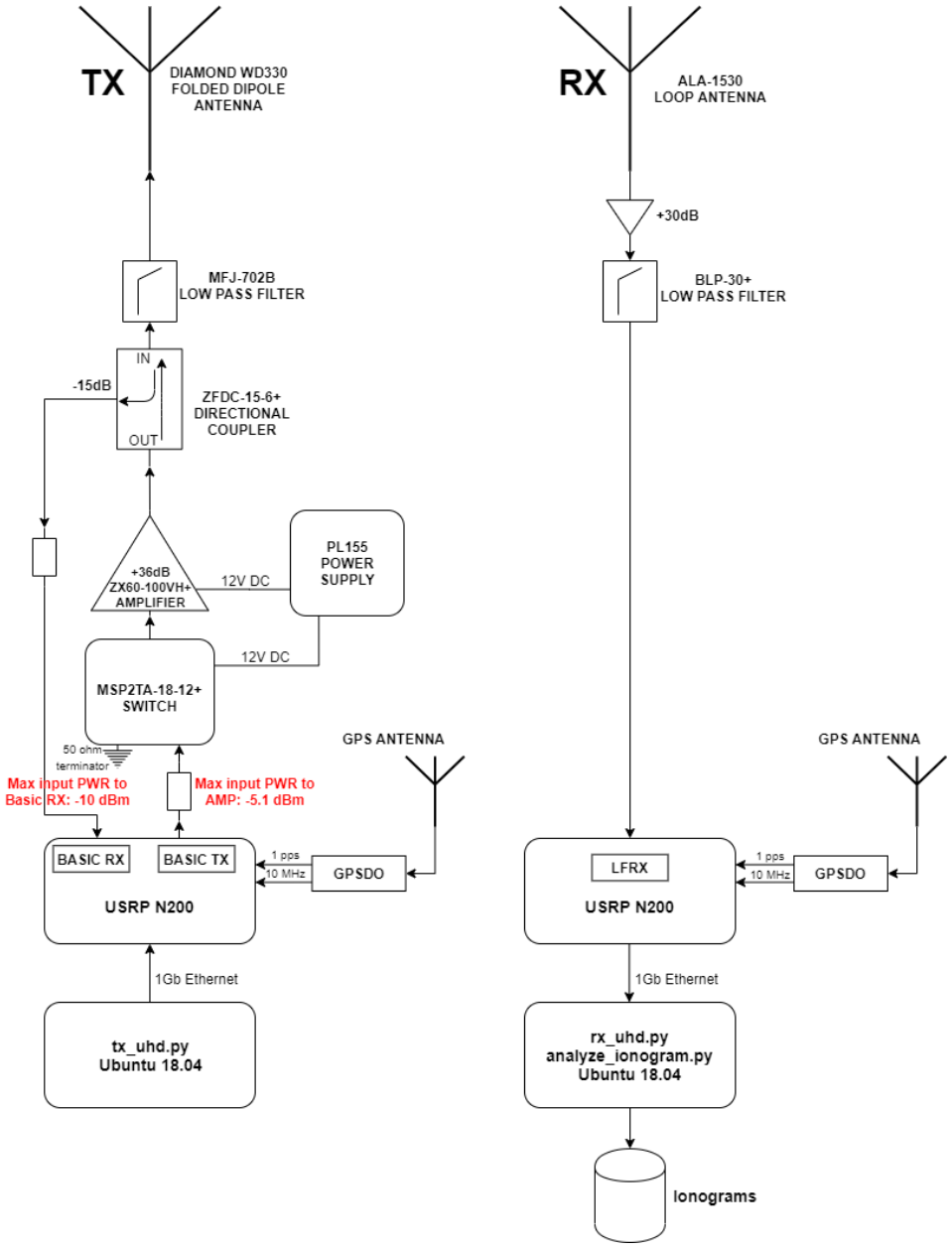


Figure 3.6: Block-diagram of the Longyearbyen Ionosonde.

With reference to the block-diagram, we will now be going through each step along the signal path provide an overview of the system. Following the overview, each hardware component is considered separately, before going over the software involved in running the ionosonde.

Transmit (TX)

The signal starts out in the lower left of figure 3.6. The computer is a Lenovo ThinkStation P320 Tiny which runs on the Ubuntu 18.04.4 LTS operating system. The computer runs the python script *tx_uhd.py* which provides the USRP N200 with the necessary commands to generate and transmit the signal via the Basic TX daughterboard. A second daughterboard within the transmit USRP, the Basic RX, is connected to a directional coupler further up the transmission line.

The purpose of the Basic RX is to monitor the reflected power to detect faults in the antenna or the feedline, which are the components that are most susceptible to damage due to external weathering effects. In the case of a bad feedline or faulty antenna, the reflected power would increase as a result of an emergent impedance mismatching. In order to protect the hardware components in case a bad feedline causes signal power to reflect back down the transmission path, attenuation on the signal is required. See section 3.3.3 for further description.

Connected to the USRP is a GPS Disciplined Oscillator (GPSDO) which is connected to an active antenna. The GPSDO uses the active antenna to connect to the GPS satellite constellation, it can then use the precise clock provided by the GPS satellites to synchronize time with other GPS connected USRPs to within a 50 ns precision of the GPS standard time. Having the clock reference provided by the GPSDO also improves the frequency accuracy of the USRP to 0.01 ppm.

From the Basic TX the signal goes to a ZX60-100VH+ amplifier via an MSP2TA-18-12+ switch. The switch ensures that the signal from TX is terminated to a 50 Ohm dummy-load when the PL155 power supply is OFF. When the power supply is ON the signal is directed to the amplifier. Between the Basic TX and the switch there is an attenuator to keep the input power to the amplifier within a safe range. The amplifier increases the signal power by 36 dB, and is then connected to a ZFDC-15-6+ directional coupler.

The signal is fed into the OUT-port of the ZFDC-15-6+ directional coupler, sending the signal out to the antenna through the IN-port. This is because the directional coupler measures the reflected power coming back into the IN-port,

this way the system can detect changes in the reflected power if the antenna breaks. Before the signal is transmitted to the antenna it goes through an MFJ-702B low pass filter to remove possible harmonics. The filter has a passband between 1.5 MHz and 30 MHz with 50 dB of attenuation at 54 MHz.

The transmitting antenna is a Diamond WD330 folded dipole. It is 25 m long, and designed for use in the HF-band.

Receive (RX)

In the upper right of the block-diagram is the receiving antenna. The receiving antenna is an ALA-1530 1 m diameter active magnetic loop antenna which amplifies signal power while lowering noise floor. The received signal is then filtered through a BLP-30+ low pass filter with a passband between 1 MHz and 30 MHz in order to further lower the noise floor.

After filtering, the signal is fed to the LFRX daughterboard within the receiving USRP N200. The receiving USRP utilizes the same reference clock provided by a GPS Disciplined Oscillator. Within the USRP the analog signal is converted to digital format before being sent to the computer. The computer is another Lenovo ThinkStation P320 Tiny which runs on the same Ubuntu 18.04.4 LTS operating system. The computer runs the python script *rx_uhd.py* to record the induced voltage in the receiver antenna. The python script *analyze_ionogram.py* uses the recorded data to generate ionograms using the matrix deconvolution technique described in section 2.5.4. The ionograms are then stored in a remote server.

3.3.3 Hardware

This section presents the hardware components provided in the system design and some of their operating parameters as provided by their manufacturers.

MSP2TA-18-12+ Switch

This component is a switch that controls the signal direction. It has 1 input port and 2 output ports which the input signal can alternate between. While the switch is powered, the signal will go through output port 1, and while it is not powered the signal will go through output port 2. The manufacturer datasheet [Mini-circuits, 2020a] states a typical insertion loss of 0.1 dB in the frequency range between DC and 1 GHz. The supply voltage is 12 V.

ZX60-100VH+ Amplifier

This component amplifies the signal power. It has a typical gain of 36 dB between 1 and 100 MHz and runs on 12 V and draws a maximum current of 370 mA. It can handle a maximum input power of +15 dBm. In the case of an open feedline or a large impedance mismatch at the output the amplifier, the power will be reflected back into the output port of the amplifier. To avoid damaging the amplifier in this scenario, the manufacturer datasheet [Mini-circuits, 2020b] states that a maximum of -5.1 dBm of power can be at the input to keep the signal from damaging the amplifier in case the antenna malfunctions.

PL155 Power supply

This component provides the necessary voltage and current to run the switch and amplifier.

ZFDC-15-6+ Directional coupler

The directional coupler redirects a fraction of signal going in the IN-OUT direction, to the CPL port. This means that any reflected signal from the antenna upstream, coming into the IN port will be redirected to the CPL port. The signal going in the IN-CPL direction has a 15 dB insertion loss, while the signal flowing in the OUT-IN direction has a 0.2 dB insertion loss. The directional coupler has a operational frequency range between 0.03 MHz to 35 MHz.

MFJ-702B & BLP-30+ Lowpass filter

These filters have a bandpass between 0 (Direct Current) and 30 MHz for the MFJ-702B and between 0 and 32 MHz for the BLP-30+. The BLP-30+ has a typical insertion loss of 0.2 dB at 10 MHz and can manage an input power of up to 0.5 W. The MFJ-702B can handle 200 W of input power.

Diamond WD330 Folded dipole

The Diamond WD330 is a 25 m long folded dipole designed for a frequency range between 2 and 30 MHz. [radioworld.co.uk, 2020].

ALA-1530 Magnetic loop

The ALA-1530 is a 1 m diameter magnetic loop antenna with a frequency range between 30 kHz and 30 MHz [Wellbrook, 2020b]. It responds primarily to the magnetic field and rejects locally radiated electric-field noise which it will reduce by up to 30 dB.

Attenuators

Signal attenuators are used to reduce the power of the signal. Attenuators are used between the Basic TX and the switch, and between the directional coupler and the Basic RX. The attenuation values will be determined after power measurements conducted in chapter 4.

3.3.4 Software

The software is written purely in Python and it relies on the Ettus Research USRP Hardware Driver library (UHD) to generate and receive radio signals. The software implements pseudorandom phase coded continuous wave transmit waveforms. The software relies on GPS to keep the transmitter and receiver clocks synchronous, which allows the transmitter and receiver to be located in different places. It is therefore easy to also have multiple receivers listening to the same transmitter. The software does not require internet to operate, which makes it possible to operate receivers and transmitters with minimal infrastructure. Other libraries that the software is reliant on includes Numpy, Scipy, Psutil, Matplotlib and H5py.

The software uses the UHD Python API (Application Programming Interface) to generate python modules from the source code of the UHD. In this way we can import modules to python that allows the programmer to issue commands and stream data to the USRP using python only.

NumPy is a powerful library for data processing in python. Among many other features, it contains efficient linear algebra routines and Fourier transforms which can be performed on N-dimensional arrays. The python modules in the UHD API is stated by Ettus Research to be optimized best for pre-allocated NumPy arrays[6] which the ionosonde software is based upon.

Scipy shares many attributes with NumPy, but has some additional utilities related to digital signal processing such as filters.

6. Statement from Ettus Research found at `https://kb.ettus.com/UHD_Python_API`

Psutil is a python library used for monitoring system parameters such as CPU usage, memory, disk-space and networks.

Matplotlib is a visualization tool in python, used for the plotting of ionograms. Its publication-quality image generation makes it a popular tool among data-scientists.

H5py is a package that enables large numpy-arrays to be stored in the HDF5 data format which is beneficial for handling large volumes of complex data. Since the radar is expected to operate continuously, large quantities of data need to be handled simultaneously.

The development of the software was done using git. Git is a version control system designed to keep track of changes in software as it is developed. It enables the programmer to test various additions to the source code before implementing the additions in the software itself.

In addition to python-code, the operating system and the BIOS firmware of the computers was modified to enable automatic startup of both the computers and python-scripts in the case of mains power loss. This way the system starts itself when mains power is returned.

The main code used on transmit is the *tx_uhd.py* script. A flowchart describing the script is show in figure 3.7.

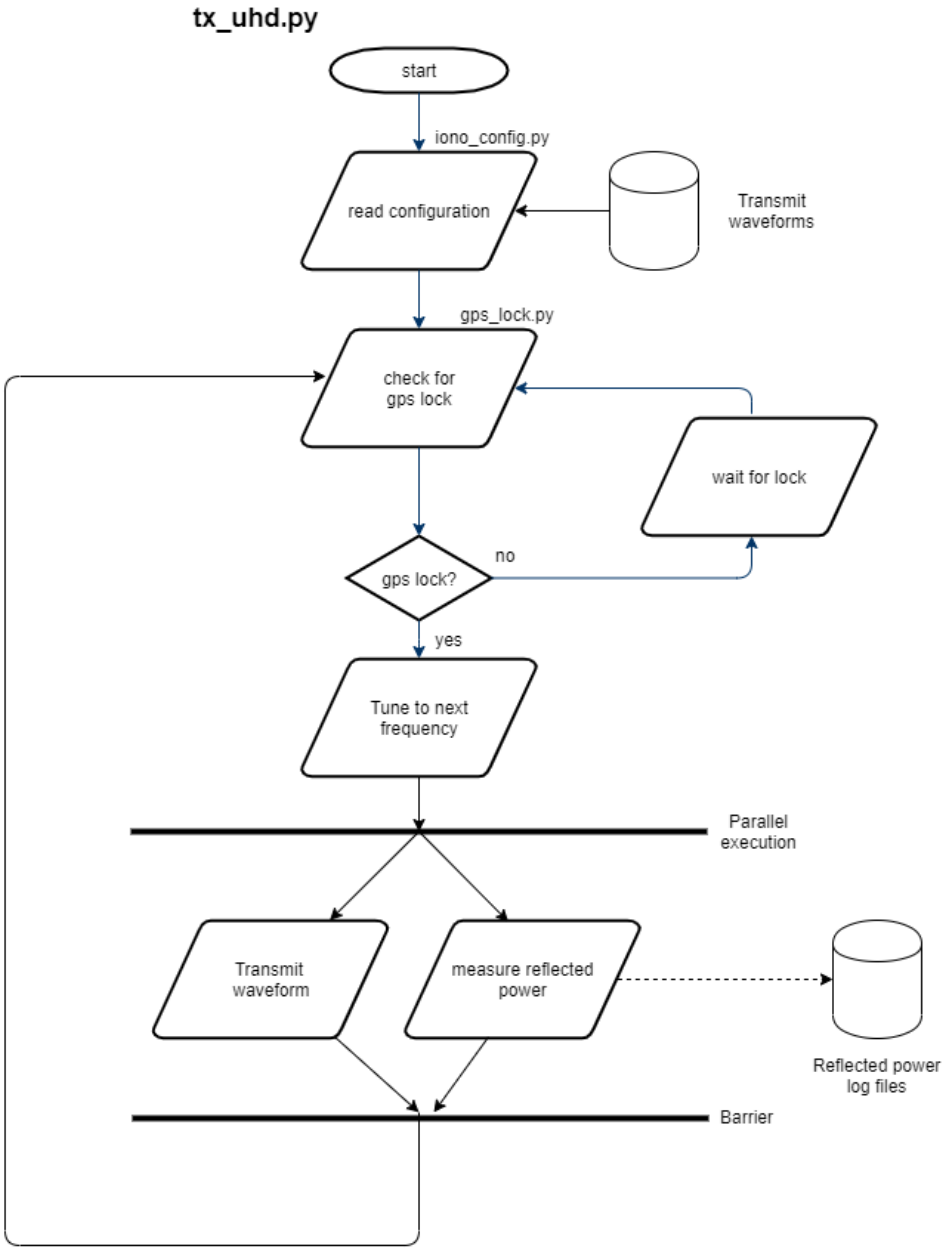


Figure 3.7: Flowchart for the python script that transmits a predefined waveform.

The transmit script starts by loading an ionosonde configuration which is also shared by the receiving script. The configuration defines the various USRP parameters such as the sample rate, TX and RX IP addresses and which ports to use. It also defines which waveforms should be used. The waveforms are predefined by a separate script. Commands such as where to store ionograms, signal amplitude and duration on each frequency is also defined by the configuration script which is labeled *iono_config.py* in the figure.

Before transmitting, the script makes sure that the USRP has connection locked to the GPS constellation with the GPSDO so that the transmitter and receiver are synchronized. The GPS lock is monitored by the *gps_lock.py* script which yields a True if locked and a False if there is no lock. If there is a lock, the script may proceed to transmit on a set of certain frequencies with a certain bandwidth which is defined by the *sweep.py* script contained in the configuration. The 30 frequencies currently in use in the sweep are shown in figure 3.8.

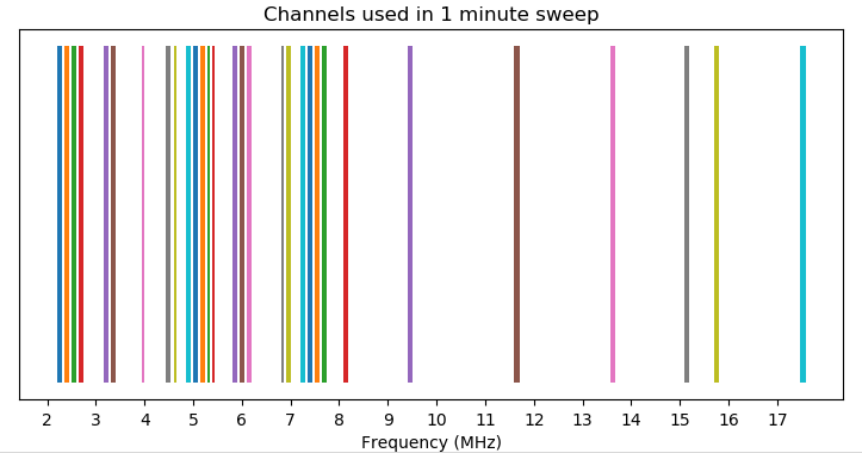


Figure 3.8: Frequency channels used for a 1 min sweep across the licensed frequency bands. 2 s are used on each of the 30 channels. The sweep can be extended to a 2 min duration by spending 4 s on each channel.

The reason for using 30 channels as opposed to the 33 channels allowed by the license is because of the simplicity it brings by being able to produce even 1 and 2 minute sweeps.

While transmitting, the script also uses a separate Basic RX within the transmitting USRP to measure the reflected power coming off of the system. The reflected power is then logged in the transmit computer.

The receiver runs on a similar python script as the transmitter. The flowchart of *rx_uhd.py* can be seen in figure 3.9.

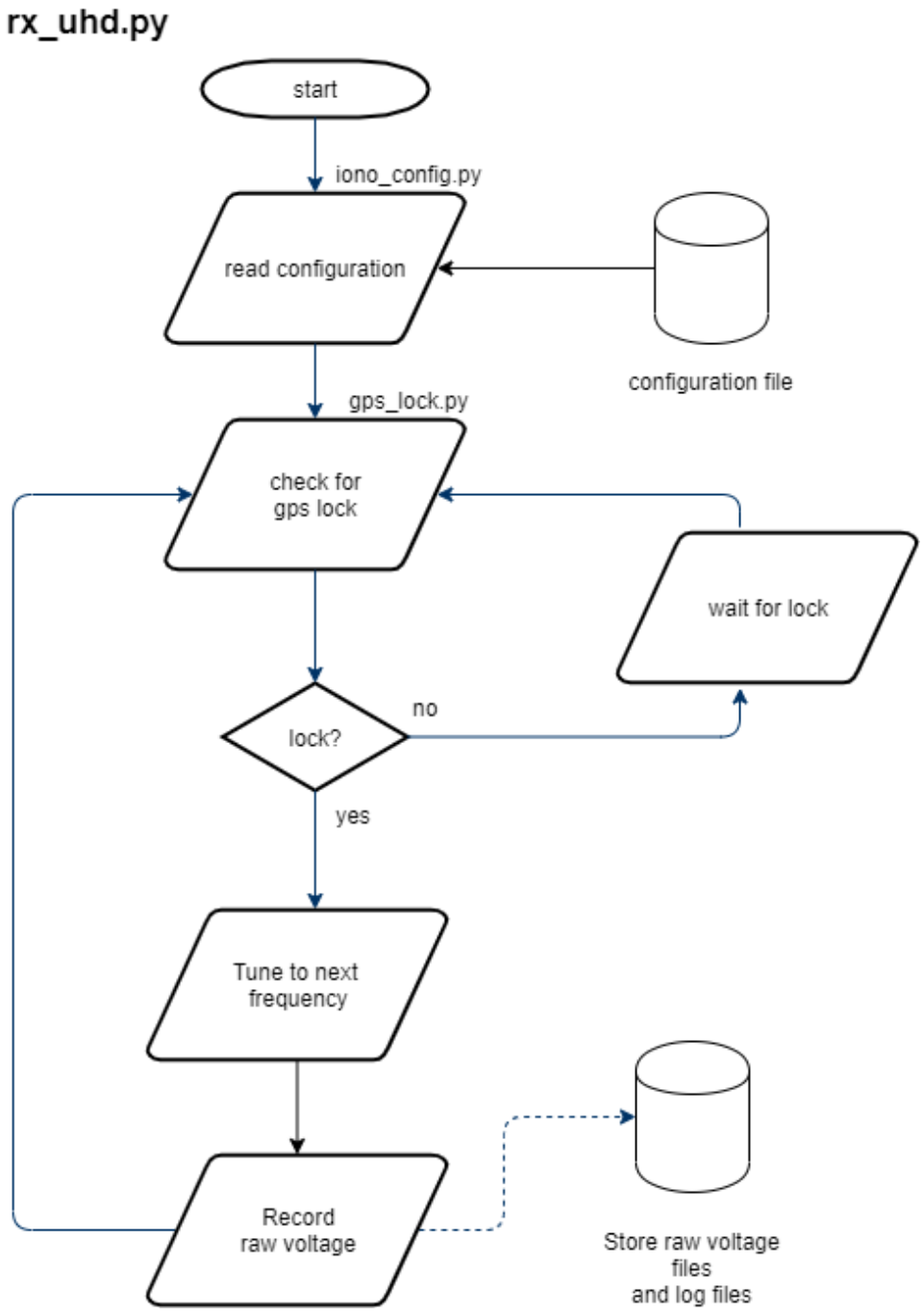


Figure 3.9: Flowchart for the python script that receives a predefined waveform.

The receiver script loads the same configuration and waits for the same GPS lock as the transmitter. As it sweeps across the same frequencies synchronously to the transmitter it records the induced voltage in the receiver antenna and stores it while also logging different housekeeping parameters such as CPU, RAM and disk-space usage.

The ionograms are produced with the *analyze_ionogram.py* script, which implements the coherent radar signal processing technique of matrix deconvolution described in section 2.5.4. The flowchart can be seen in figure 3.10.

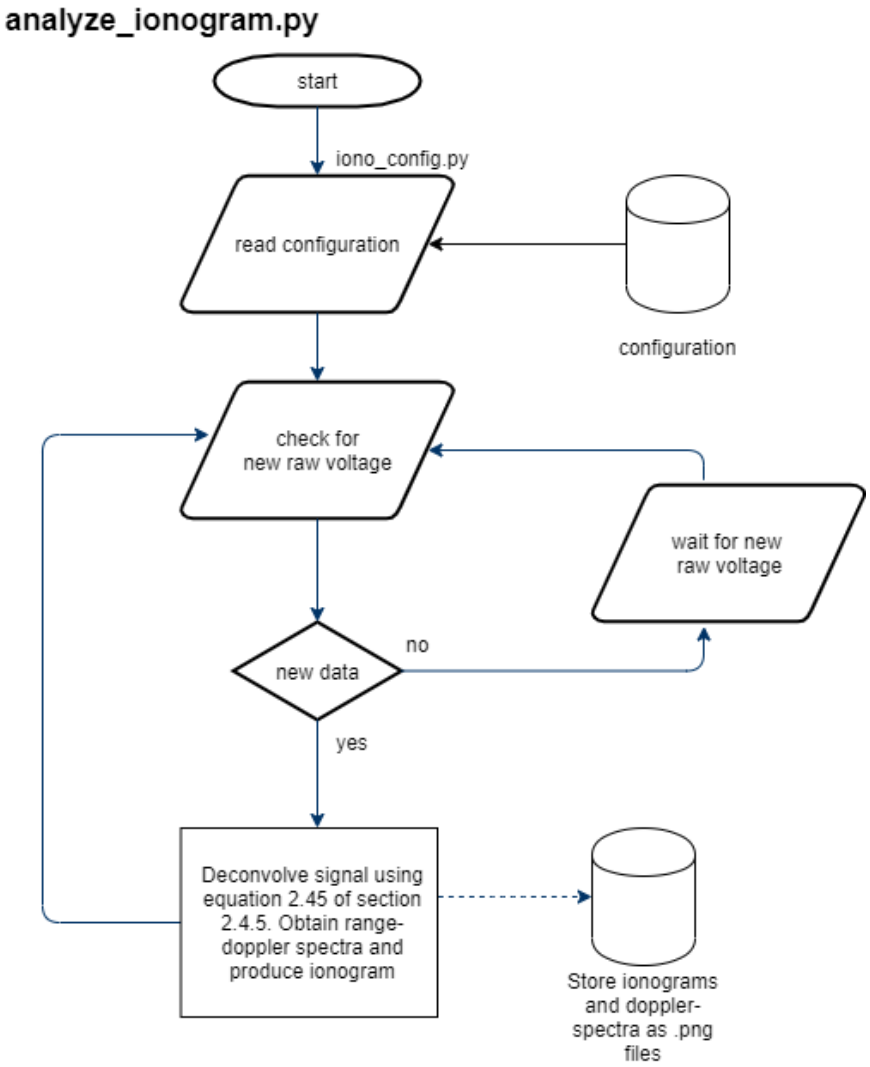


Figure 3.10: Flowchart for the python script that analyses the received signals to produce ionograms.

3.3.5 Transmit waveform

The transmit waveform is a continuous wave coded long pulse using a poly-phase modulation scheme as described in section 2.5.3. The pseudo-random code is defined by a length and a "seed" which determines the random sequence of phase-shifts, hence the term "pseudo-random". The code is then over-sampled to increase its length while retaining the same structure. The sample rate determines the length in time and the bandwidth of the code. The code is modulated onto carrier signals with a center frequency defined by the *sweep.py* script found in the source code in appendix D.

There are a total of three waveforms or coded long pulses defined in the frequency sweep. The three waveforms are used to accommodate for the three different bandwidth restrictions on the channels defined by the frequency licence.

With a 100 kHz bandwidth on the waveform, equation 2.41 yields a range resolution of 1.5 km. This means that when we measure a reflection, each sample point is separated by 1.5 km, any changes in the medium within those 1.5 km is not observed by us.

Figure 3.11 shows the Fourier transform of a square pulse in time-domain. In the frequency domain, the square pulse has the shape of a sinc-function.

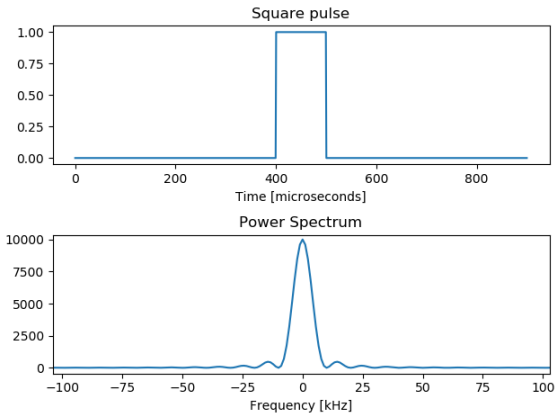


Figure 3.11: A square pulse in the time domain and its Fourier transform pair in the frequency domain.

The coded long pulse is essentially a train of square pulses in its complex baseband representation[7]. The Fourier transform of an entire pulse is shown in figure 3.12.

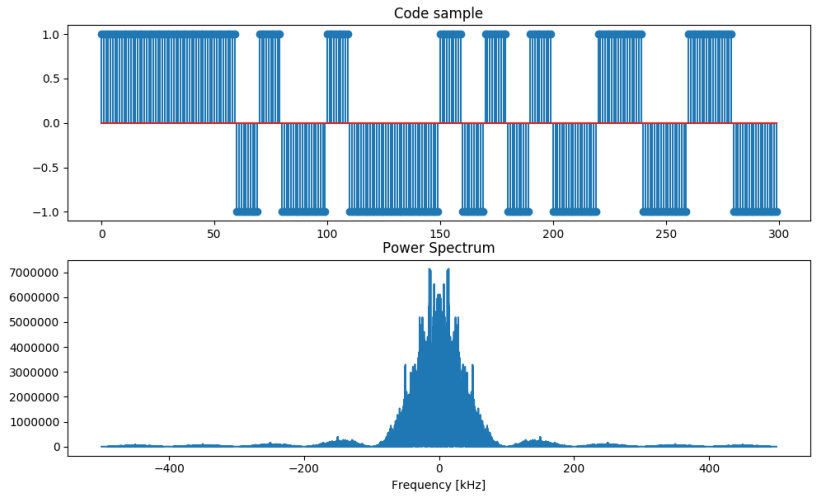


Figure 3.12: First 300 samples of the coded waveform in the first row. Frequency power spectrum of coded waveform in second row.

By filtering the signal, the ringing components outside of the 100 kHz band are removed. Filtering is done in the digital domain using software, which is partly why the code is over-sampled to allow for more precision filtering in the digital domain. The additional analog low-pass filter removes any resonances that may occur in the feedline.

7. See section 2.5.2 for a description of complex baseband signals

4

System Testing

During the development of the ionosonde, tests need to be conducted to choose suitable components to complete the design and then to ensure proper functionality. In this chapter we go through the tests as they were conducted both in a laboratory environment and in the field. The chapter chapter is separated into two sections, the first section covers the laboratory testing and the second section covers testing conducted in the field.

4.1 Laboratory Testing

Setting up the system on a laboratory bench is beneficial because it allows us to monitor the system behaviour in a controlled environment before taking it out in the field. A loopback test is done by sending the data signal from the transmitting source into the receiving end without any external modification of the signal (i.e. ionospheric effects) with the purpose of identifying potential errors along the transmission path and reveal bugs in the software. The test-setup only differs from the full-scale system in that it lacks transmission and receiving antennas for the signal to propagate through, instead we use a simple coaxial cable to feed the signal from the transmitter (TX) source to receive (RX). In a real-life scenario the transmitted signal decreases in power by several orders of magnitude as it propagates through the ionosphere as described by equation 2.22 in chapter 2, in a laboratory-setup this will not be the case. In order to not damage the sensitive receiver we use the appropriate signal

attenuators in order to not overload the RX hardware.

4.1.1 Measuring maximum output power

Measuring the maximum possible output power of the transmitter is necessary to make sure we stay within the licensed power requirements and to prevent exceeding the maximum input power to hardware components within the system to avoid damage. The measurements were conducted by connecting a 50 ohm termination at the transmitter output and a T-piece BNC adapter to connect an oscilloscope probe. The measurement setup is shown in figure 4.1. The computer is used to control the USRP using a GNURadio Companion

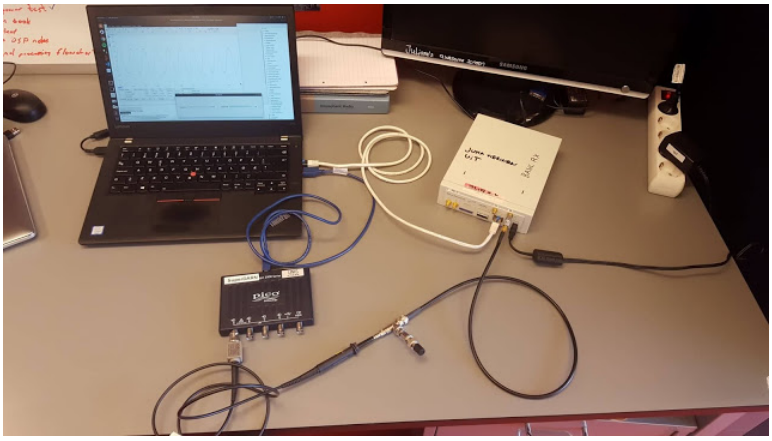


Figure 4.1: The output of the USRP is fed into a T-adapter with a 50 Ohm dummy load at the one end and a high-impedance oscilloscope probe at the other.

software script while measuring the output using an oscilloscope (Picoscope 2408B) with a high-impedance probe. As the script is running we can observe and do measurements on the transmitted waveform using the Picoscope. A signal waveform was generated using the GNURadio Companion software script shown in figure 4.2. In the script the signal source is set to zero frequency with an amplitude of 1 - this serves as the baseband signal. The signal is then modulated by the UHD: USRP sink block with an adjustable center frequency. The center frequency represents the carrier frequency to which the baseband signal is up-converted. The gain is set to "normalized" with the maximum value of 1. This ensures that the USRP is always transmitting at its highest power. Ensuring that the probe is correctly compensated, the peak-voltage is measured over a range of frequencies in the HF-spectrum. The peak-voltage V_p is related to the signal power in dBm, P_{dBm}, through the following relation:

$$P_{dBm} = 10 + 20\log_{10}(V_p) \tag{4.1}$$

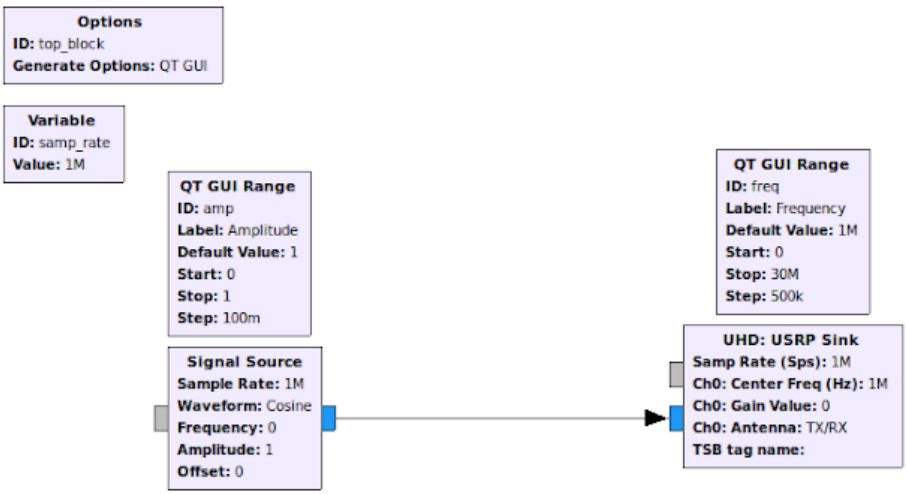


Figure 4.2: The GNURadio flowchart consists of a Signal Source block and a UHD: USRP Sink block. Sliders are added to adjust the amplitude and frequency.

For a USRP with a BasicTX daughterboard mounted inside, the maximum signal power as a function of frequency is given in table 4.1:

f [MHz]	P [dBm]	f [MHz]	P [dBm]
1	-1.7	9	-1.6
2	-1.6	10	-1.6
3	-1.6	11	-1.6
4	-1.6	12	-1.6
5	-1.6	13	-1.6
6	-1.6	14	-1.7
7	-1.6	15	-1.7
8	-1.6	16	-1.7

Table 4.1: Output power in dBm as a function of output transmit frequency. The power stays about the same over the lower half of the HF-spectrum.

From the measurements we see that we obtain a maximum output power of $-1.7\,\text{dBm}$, corresponding to a peak voltage of $V_p \sim 260\,\text{mV}$.

To measure how the output power responds to amplitude adjustments in the signal waveform, the frequency was set to 3 MHz while the amplitude was varied between a factor of 1 and 0.1 of maximum amplitude. The measured values are shown in table 4.2:

A	P [dBm]	A	P [dBm]
1	-1.7	0.5	-7.4
0.9	-2.5	0.4	-9.3
0.8	-3.5	0.3	-11.5
0.7	-4.6	0.2	-14.7
0.6	-5.9	0.1	-19.5

Table 4.2: Output power in dBm as a function of signal amplitude A relative to maximum at a transmit frequency of 3 MHz.

From the measurements we see that when keeping the amplitude at maximum we have an output power of -1.7 dBm. When halving the amplitude the power output is -7.4 dBm.

From section 3.3.3 the maximum input to the amplifier is restrained to -5.1 dBm. With a maximum output of -1.7 dBm we can conclude that a $-1.7+5.1 = 3.4$ dB attenuator is the minimum attenuation needed to obtain a maximum possible output to the amplifier of -5.1 dBm.

4.1.2 Software and preliminary loopback testing

This subsection describes the preliminary testing performed on the system as shown in figure 4.3. The purpose of testing the system in this setup is to make sure that the software runs as intended over a longer period of time without encountering programming errors or overloading the CPU. The left USRP labeled "TX" is the transmitter. On its front ports is connected an ethernet cable, a coaxial cable and the power supply. On the right is the receiving USRP labeled "RX". The commands and data are fed through the ethernet cables, which are connected to two separate computers, one running the software for transmitting and the other running the software for receiving and analyzing the data. The antennas for the GPS disciplined oscillator (GPSDO) which provide the reference clock signals used to synchronize the two USRPs are connected to the back panels of the USRPs.

The signal in the USRP is converted from digital to analog in TX and transmitted to RX via the coaxial cable. Attenuators at the receiving end reduce the signal power at the RX input port, and the signal is converted from analog to digital in the RX USRP. The attenuators equate to a reduction of power by 35 dB to ensure that the input power at the receiving USRP does not exceed the limit of -10 dBm.

In this test the receiving software produced an ionogram once every minute.

Figure 4.3: Setup of the transmitting USRP labeled TX and the receiving USRP labelled RX. The RX has 3 additional attenuators and a DC block connected to the input port.

The transmitting USRP spent 2 seconds of transmit time on each of the 30 frequency channels. During testing, the receiving computer would freeze rather sporadically after running for $3 - 4$ days. This was hypothesized to be due to a driver issue in the operating system. Whether or not this was caused by the ionosonde software was not investigated, as the issue was resolved by performing a clean install of Ubuntu 18.04, its dependencies and the ionosonde software.

4.1.3 Complete system test and measurements

This section describes the complete system loopback testing conducted in the laboratory, including all components as shown in the system diagram in figure 3.6 with the exception of the transmitter and receiver antennas. The complete setup is shown in figure 4.4.

In addition to testing that the system works, measurements are done on the signal output power and the reflected signal power. Measurements on the output power are to confirm that the theoretical output power matches that of the actual output power. Measurements on the reflected signal power are used to calibrate the Basic RX within the TX USRP so that it can measure actual reflected power. Knowing the level of reflected power during normal operation can be used for detecting faults in the feedline and antenna.

When measuring the output power from the complete system, the receiving USRP and adjustable attenuator have been replaced with a dummy load (50 ohm terminator). The measurement configuration is shown in figure 4.5.

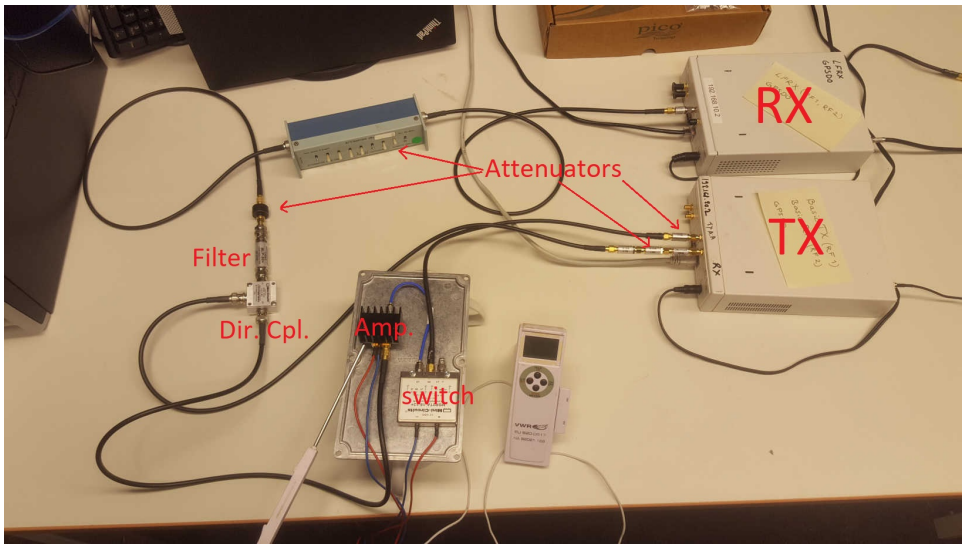


Figure 4.4: Loopback testing setup including the switch, amplifier, directional coupler, filter and attenuators. The RX USRP is connected to an adjustable attenuator seen as a blue box in the upper part of the figure. A temperature measuring probe measures the temperature of the heat sink on the amplifier.

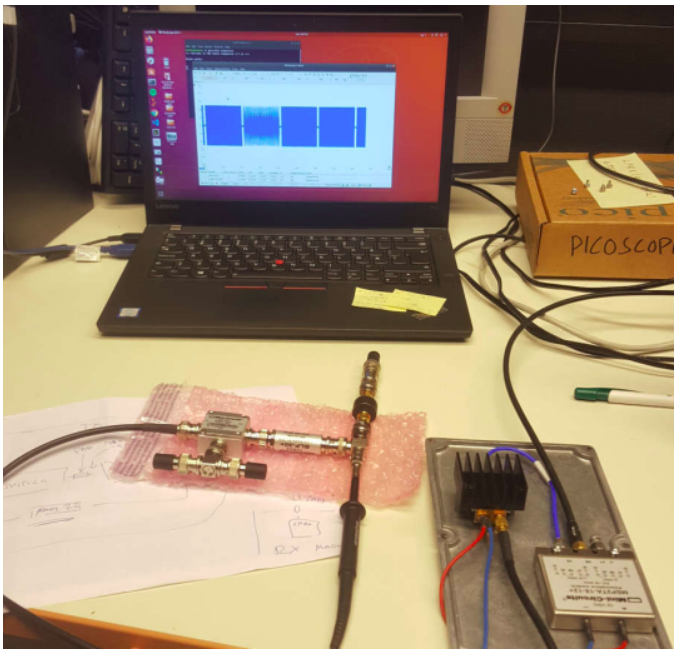


Figure 4.5: Configuration of the Picoscope probe going into the signal feedline using a BNC T-adapter. The figure shows the measurement setup of output transmit power. Reflected power is measured by connecting the Picoscope probe to the T-piece that is connected to the CPL port of the directional coupler. The signal measured by the Picoscope is seen on the computer-screen.

Setup

Within the TX USRP there is a Basic TX daughterboard for transmitting the signal, and a Basic RX daughterboard for measuring reflected signal power. From the output of the Basic TX the signal is fed through a −10 dB attenuator which then connects to the switch which is fastened to an aluminium plate alongside the amplifier. From the switch the signal goes to the amplifier which can be seen in figure 4.5 on the aluminium plate with a black heat sink. The switch and amplifier are both connected to a power supply which provides 12 V and 880 mA (∼ 10.6 W) to run the two components.

From the amplifier the signal is fed to the OUT port of a directional coupler and goes via its IN port to a low-pass filter. The directional coupler has a port labeled CPL that picks up any reflected signal coming back through the IN port, this is fed back through 25 dB of attenuation to the Basic RX within the TX USRP. From the low-pass filter, the output signal goes through a 6 dB attenuator which then feeds the signal to a dummy load.

Measuring

The Picoscope is used to measure signal power at the output of the low-pass filter and the reflected signal power from the output of the CPL port of the directional coupler. The configuration for measuring the output power and reflected power can be seen in figure 4.5.

The measurements are done by connecting a BNC T-piece to the low-pass filter output with the signal going through the 6 dB attenuator on to the terminator in one end and the Picoscope probe with a BNC-adapter in the other end. The reflected power is measured by connecting the Picoscope probe to one of the ports in the T-piece connected to the CPL port of the directional coupler. While measurements on the output power is done, the CPL port of the directional coupler connects to the Basic RX within the TX USRP as shown in figure 4.4.

Signal output power and reflected power

We will now present the measurements of the signal output power going to the simulated transmit antenna and the reflected signal power going back into the system. A comparison with theoretical calculations of the output power will then follow.

Using the measurement setup shown in figure 4.5, the peak-to-peak voltage of

the output signal is measured. The peak voltage is obtained by dividing the peak-to-peak voltage by 2. It can then be related to the power in dBm by using equation 4.1.

Measurements of the peak-to-peak voltage were conducted over the whole frequency sweep of the transmit signal. The highest peak-to-peak voltage was measured to be 4.2 V, while the lowest value measured was 4.0 V. These voltages correspond to an output power of $\sim 16 - 17$ dBm.

The reflected power were measured to be ~ 900 mV peak-to-peak. This corresponds to a reflected power of ~ 3 dBm.

Comparison to theoretical values

The transmit signal was used with 0.5 of maximum amplitude, from measurements in section 4.1.1 this corresponds to an output power of -7.4 dBm. This signal is then attenuated by 10 dB, and then further amplified by +36 dB. This gives a theoretical output power of 18.6 dBm. The measured output power was $16 - 17$ dBm. This gives a discrepancy of ~ 2 dBm between theoretical and measured output power. The discrepancy is within reason when taking into account insertion loss of the different components and the fact that the amplifier gain varies in frequency [Mini-circuits, 2020b].

Calibrating the Basic RX within the TX USRP

By measuring the actual reflected power we can calibrate the measurements done by the Basic RX within the TX USRP.

The measured reflected power of 3 dBm that goes into the Basic RX is further attenuated by the -25 dB attenuators that is connected between the output of the CPL port of the directional coupler and the input to the Basic RX. This means that we have -22 dBm of power going into the Basic RX that measures the reflected power.

During measurements of the output power the transmitting software showed a reflected power of ca. -39 dB. From the measurements of the reflected power we know that this should correspond to -22 dBm. Therefore, a calibration offset of 17 is added to the software that measures the reflected power and we have thus calibrated the software to display close to true reflected power measured by the Basic RX.

Loopback tests

We will now discuss the testing conducted on the system when configured to a loopback system as shown in figure 4.4. A block-diagram of the same testing setup can be seen in figure 4.6.

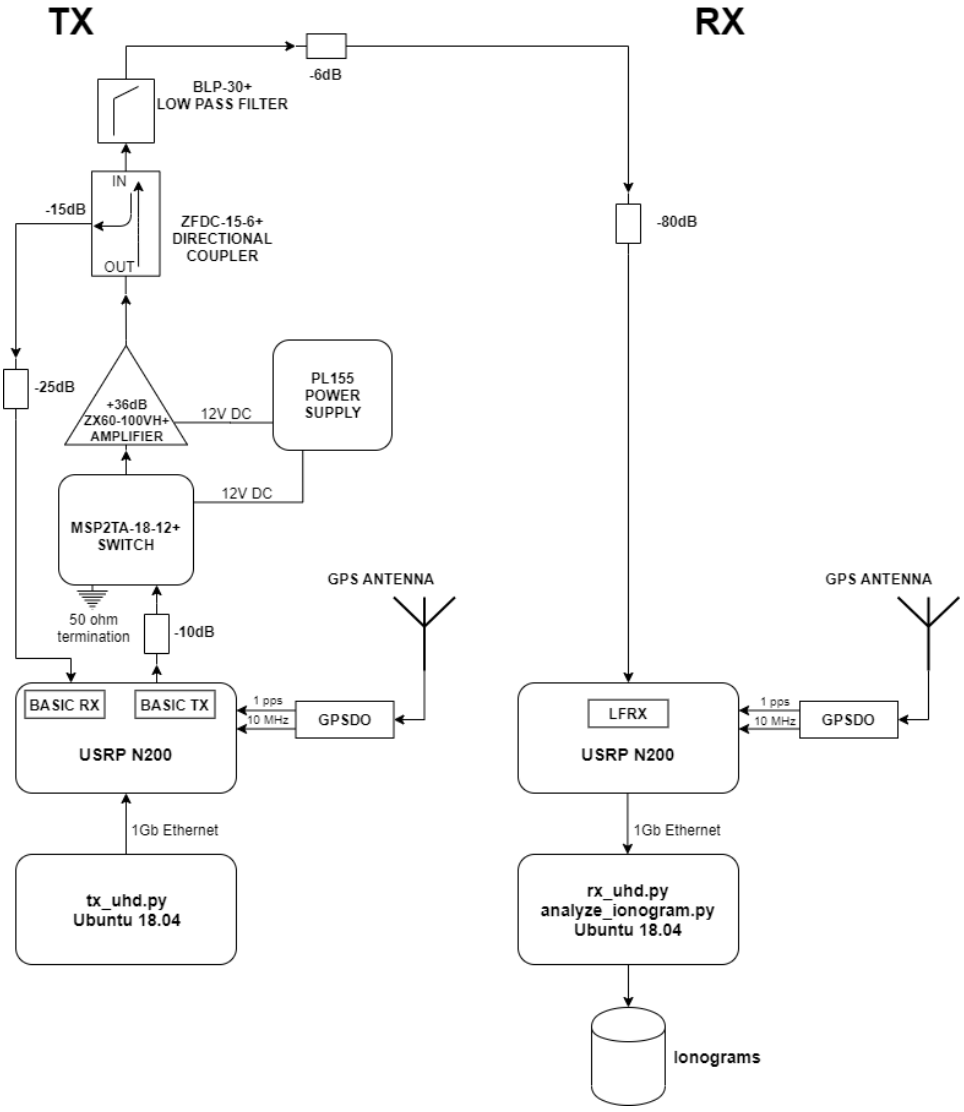


Figure 4.6: Block-diagram of the complete system loopback test setup.

The testing was conducted in the same way as the preliminary loopback testing; running the software for longer periods of time to ensure that the additional hardware components perform as expected. Additional tests were also performed at this stage. During operation, power supply to the amplifier and switch was turned off, the system responded as expected by redirecting the signal to the 50 ohm terminator. The *gps_lock.py* script was also tested by disconnecting the GPS antenna from the USRPs, at which point the USRPs would stop transmitting and receiving until GPS connection had been achieved.

The transmit waveform bandwidths were also measured to ensure the signal stays within the licensed bands. The measurements were done using the recorded voltage in the receiving USRP itself and with a separate dedicated spectrum analyzer. The code used to calculate the signal bandwidths can be found in the repository provided in appendix D.

The waveform bandwidths measured by the RX USRP is seen in figure 4.7

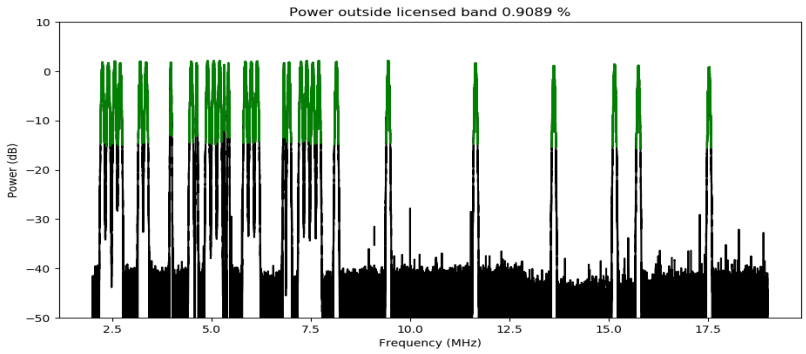


Figure 4.7: Spectrum of the transmit signals measured by the RX USRP. Frequency on x-axis and power on y-axis. The green-coloured part of the spectrum is the part of the transmit signal that is within the licensed bands. The frequency license allows a maximum of 1% of power outside the licensed bands.

As stated by the frequency license, we had some exceptions to the bandwidths of 100 kHz on the licensed bands. This was solved by having 3 separate waveforms for each required bandwidth as discussed in section 3.3.5. These channels were also measured to see if they had the required bandwidth. An extract of the measurements in figure 4.7 is seen in figure 4.8.

In addition, the head engineer of Kjell Henriksen Observatory conducted an independent test using a spectrum analyzer (Siglent SSA3032X) to confirm the transmission bandwidths.

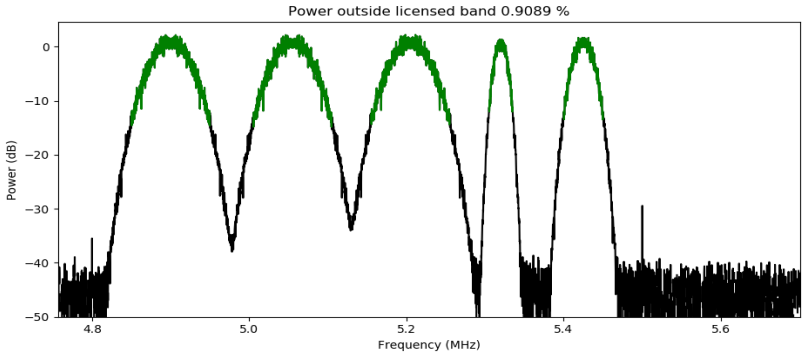


Figure 4.8: Bandwidth measurement of 3 different waveforms used to accommodate for the 3 different bandwidths required by the license. Required bandwidth on the channels are: 100 kHz (seen on the left), 50 kHz (seen on the right) and 30 kHz (seen in the middle).

4.2 Field Testing

After all the laboratory testing is conducted, the system is brought out to the field for final implementation. At this point the system is set up as shown in figure 5.8 of chapter 5, with the transmit and receive antennas included. With known power going into the transmit antenna, we can use the measured reflected power from the Basic RX within the transmit USRP to estimate the SWR discussed in section 3.2.3. Using equation 3.1, the SWR is obtained for the frequencies we use. Figure 4.9 displays the reflected power and SWR on each frequency being transmitted.

The SWR is found to be highest on the low frequencies, and lowest around 6 MHz which is approximately the resonant frequency for a 25 m dipole antenna.

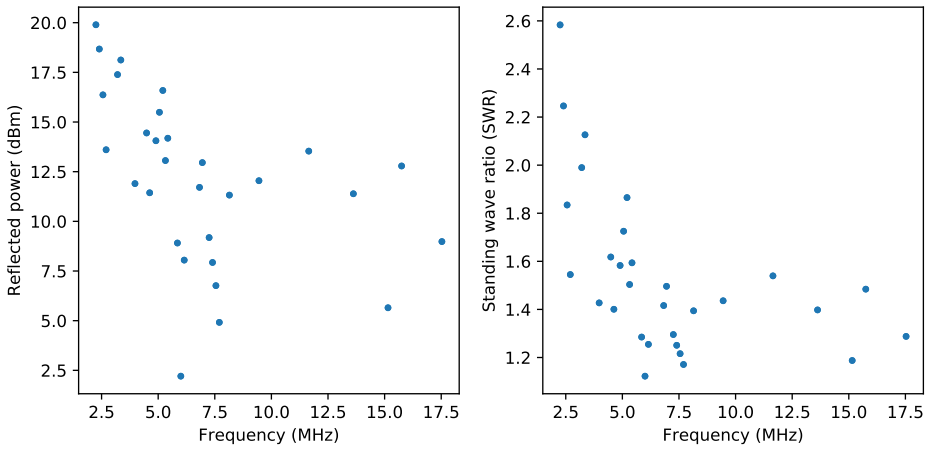


Figure 4.9: Reflected power measured with the transmitter USRP and transmitter software. Left: reflected power in units of dBm. Right: standing wave ratio calculated from the measured transmit power and measured reflected power using equation 3.1. On frequencies below 5 MHz, the antenna impedance matching is degraded, meaning that the antenna is less efficient. This is to be expected, as the antenna is somewhat small compared to the wavelength on these frequencies.

5

Results

Following implementation and testing in the laboratory, the ionosonde system was brought up to the radar site located on the Breinosa plateau outside Longyearbyen. Containers used in previous radar systems are used to house the receiver and transmitter hardware. The radar site is shown on figure 5.1, with the transmitter and receiver shown in figures 5.2 and 5.3, respectively.

This chapter presents the final results of the design and development conducted in chapter 3 and 4, and some of the data currently obtained by the system. Since this project entails both an engineering and a scientific aspect, the results are divided into two sections. The first section presents the scientific results in the form of ionograms and overview plots. The second section presents the ionosonde system itself. Lastly, a discussion of the results is provided.

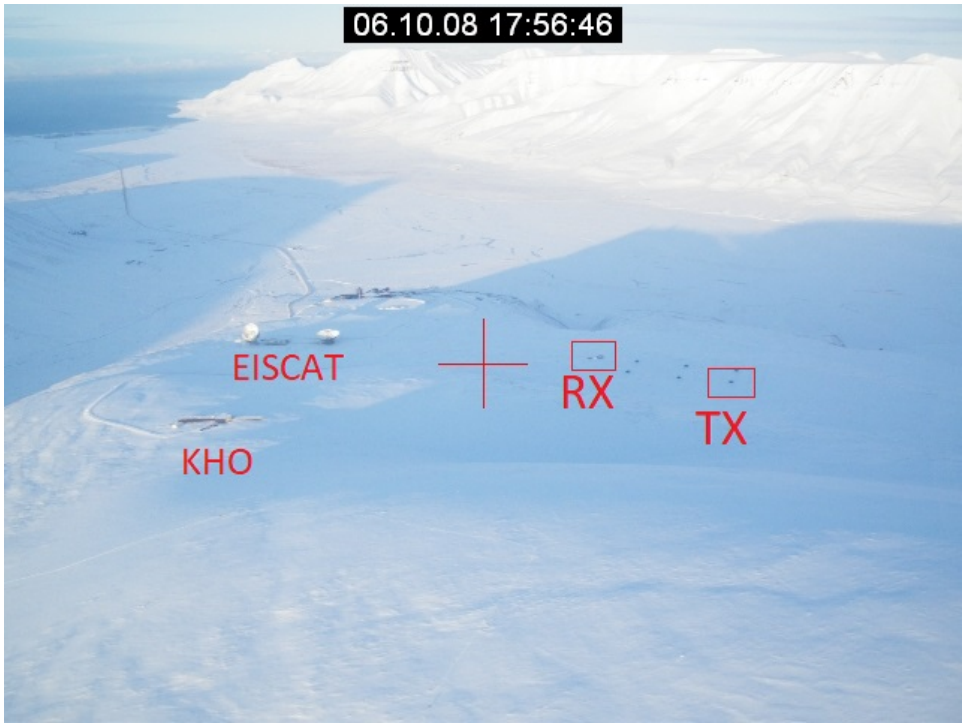


Figure 5.1: Aerial view of the radar site located on Breinosa, looking north-northwest. Transmit container labeled "TX" and receive container labeled "RX". The Kjell Henriksen Observatory (KHO) and the EISCAT facility is seen on the left. Image taken by Jorgen Lenks, 6th October 2008. Image adapted from `kho.unis.no`

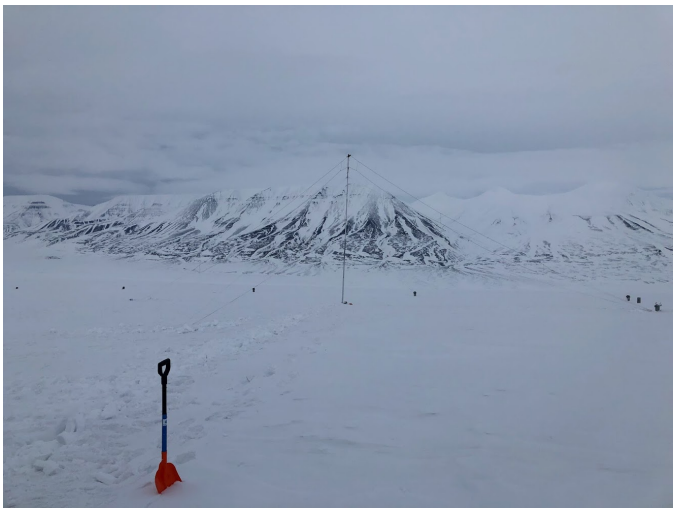


Figure 5.2: The transmit antenna at TX location. Diamond WD330 25 m folded dipole, setup as an inverted-V.

Figure 5.3: The receive antenna at RX location. ALA1530 1 m magnetic loop.

5.1 Scientific Results

In this section, we present some of the data obtained by the Longyearbyen ionosonde after being fully implemented. As of June 2020, the ionosonde sweeps over the 30 frequency channels displayed in figure 3.8 using 4 s on each channel. An ionogram is thus produced every 2 minutes.

Figure 5.4 shows an ionogram obtained on the 26th of May, 18:38 UT.

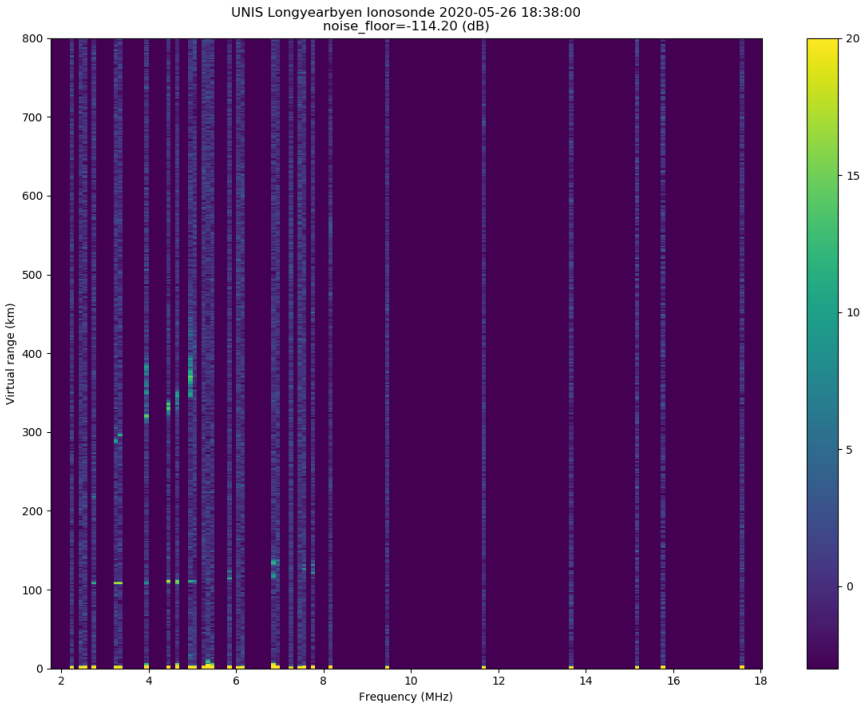


Figure 5.4: Ionogram obtained on the 26th May 2020 18:38 UT. Range displayed on the y-axis and frequency on the x-axis. Reflections can be seen from about 110 km and 290-400 km at frequencies between 3 and 8 MHz. The colorbar displays signal power in dB relative to the noise floor which in this ionogram is -114 dB.

The signal has a range extent of 1500 km, but the displayed range in the ionogram is adjusted to a maximum of 800 km since we are currently just focusing on sounding the ionosphere with vertical incidence.

In the ionogram we see reflections coming from the E-region around 110 km, and the F-region between 290-400 km. The reason we see reflections from both regions within the same frequency bands is because the signals are partially reflected in the E-region, while the rest of the signal is refracted and

propagates further up to the F-region. The strongest reflections appear to occur at the 3.3 MHz and 4.4 MHz channels. The approximate electron densities from these reflections is found to be 1.4×10^{11} m^{-3} and 2.4×10^{11} m^{-3}, using equation 2.1. These densities are slightly higher than the densities suggested in the E-region by the model parameters in table 2.1.

Using the processed data we can also make time-series plots to showcase the temporal evolution of the scattering medium. Figure 5.5 displays which ranges the reflections are coming from.

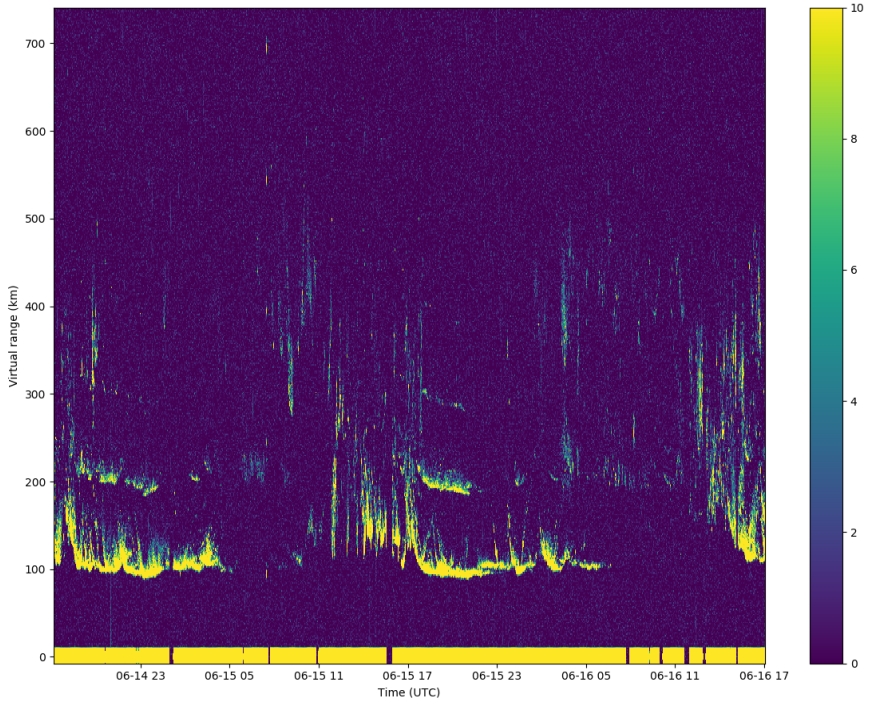


Figure 5.5: 48 hour overview plot of reflected signal power as a function of range and time. Plot shows reflected signals between 17:00 UT, 14th June and 17:00 UT, 16th June.

The signal around zero range is the direct overground transmission between TX and RX where the signal is not propagating via the ionosphere. The reason for the overground transmission having an apparent negative range is because of an artificial range shift introduced in the plotting script. The gaps that can be seen in the overground transmission is due to the airport-interlock system discussed in section 5.2.

A similar time-series of the reflected frequencies is also made, showcasing the frequencies at which reflected signals are obtained. Figure 5.6 is such a plot, displaying the frequencies at which reflections are coming from.

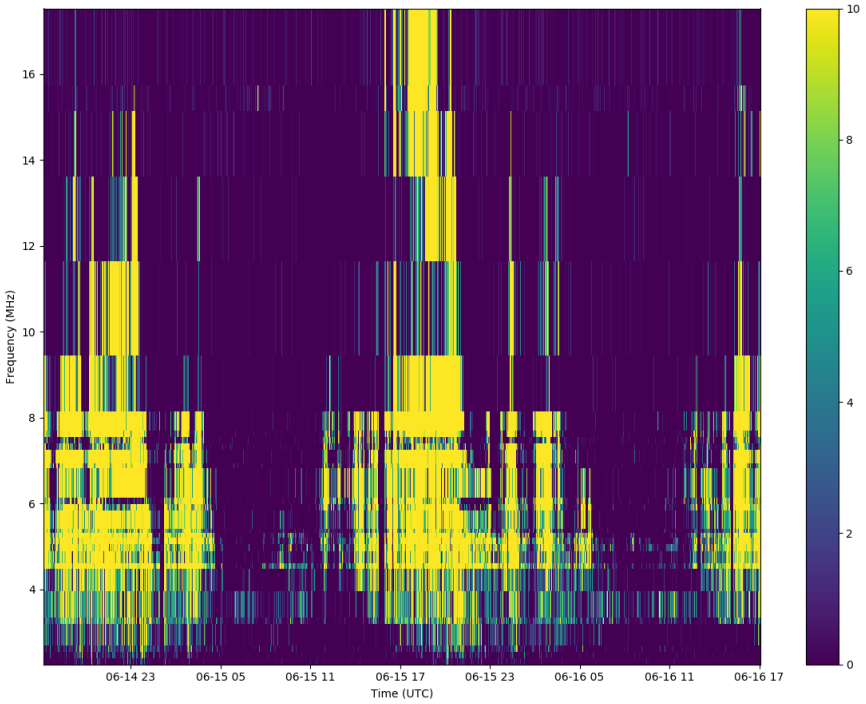


Figure 5.6: 48 hour overview plot of reflected signal power as a function of frequency and time. Plot shows reflected signals between 17:00 UT, 14th June and 17:00 UT, 16th June.

From figures 5.5 and 5.6 there appears to be some periodicity as to when most of the reflections are occurring. The quietest periods seem to be in the morning sector between 06:00 and 11:00 UT, while the active periods seem to be in the daytime and evening sector.

The software also produces range-Doppler spectra for each frequency channel throughout the sweep. Figure 5.7 displays an example obtained on June 27th, 21:00 UT.

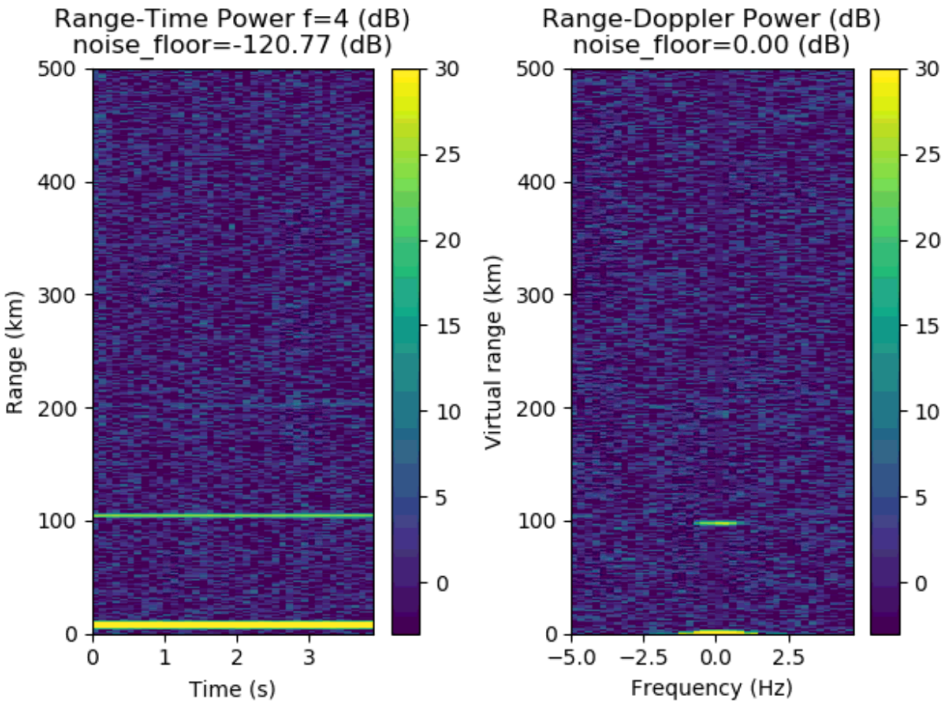


Figure 5.7: Range-Doppler spectra for a single frequency channel. Left: range-time power spectra. Reflection observed at ~105 km. Right: range-Doppler spectra. Reflection has a Doppler-bandwidth of ~1.5 Hz.

The frequency channel used in the Doppler spectrum is the 4th channel in the sweep. This channel has a center frequency of 3.205 MHz and a 100 kHz bandwidth. Using equation 3.8 we obtain a Doppler-width of 70±1 m/s, which is the line-of-sight velocity of the reflecting medium observed by the ionosonde.

For the latest ionograms and 48 h overview plots, I invite the reader to visit `http://kho.unis.no/Ionosonde/` The link is however subject to change, in which case the latest plots should be available under the data section found on `http://kho.unis.no/`

5.2 Engineering Results

The implementation has resulted in a functioning software defined ionosonde. The ionosonde consists of 2 computers, 2 USRPs and a small form factor circuitry, with a 25 m folded dipole antenna for transmit, and a 1 m magnetic loop antenna for receive. The complete system diagram is shown in figure 5.8.

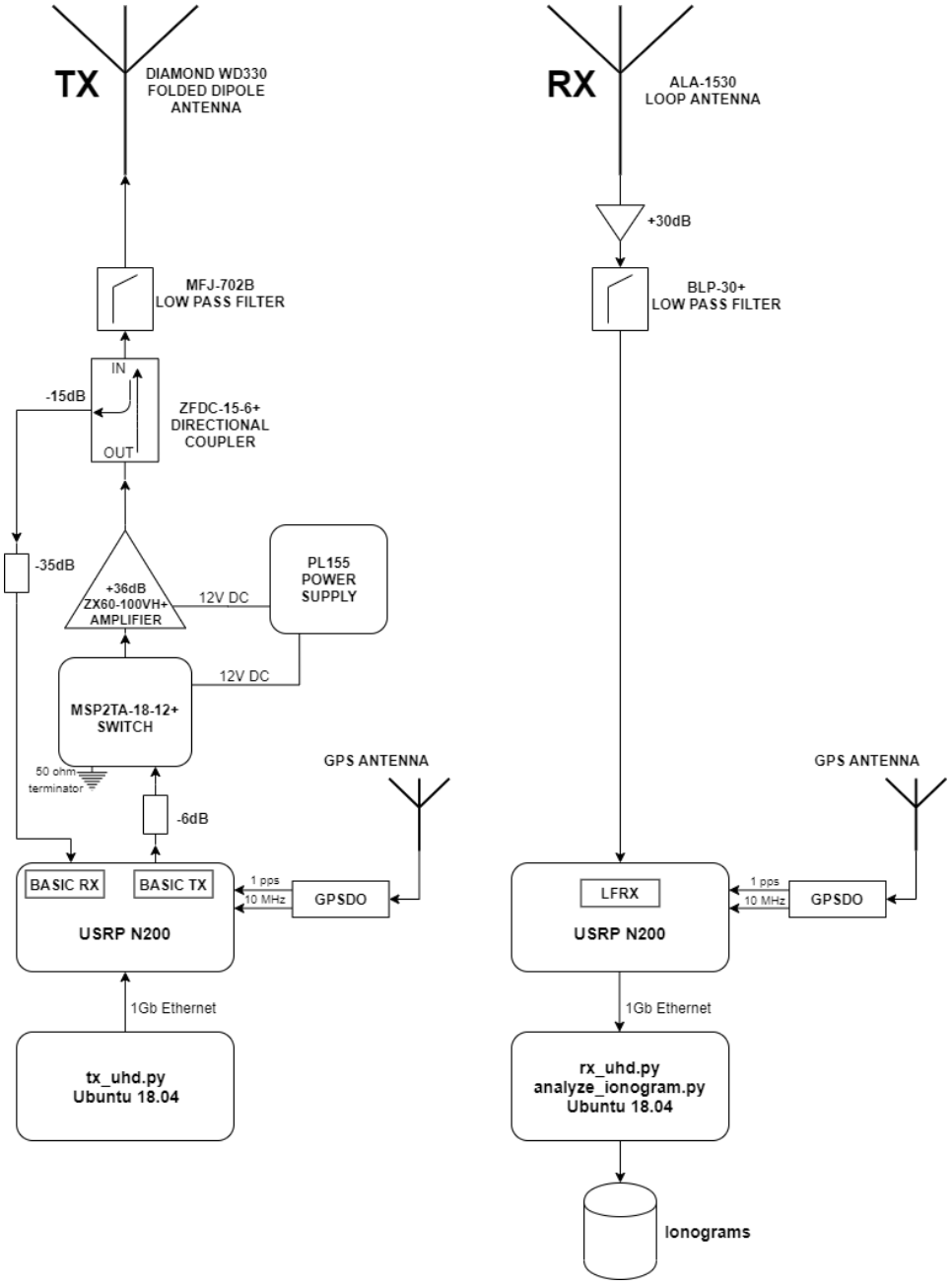


Figure 5.8: Complete system block-diagram of the Longyearbyen ionosonde. Transmitter is seen on the left and receiver is seen on the right. The two subsystems are synchronized by two GPS Disciplined Oscillators that are connected to the GPS constellation using a GPS antenna.

Within the transmitter circuitry there is a switch, an amplifier a directional coupler and attenuators. The amplifier is necessary to have sufficient transmit power P_T and the directional coupler is used to monitor reflected power. Attenuators are used to keep signal levels in a safe range for the components.

The main purpose of the switch is to re-direct the transmit signal from the USRP to a 50 ohm terminator whenever the power amplifier is switched off. This is to protect the amplifier electronics. The switch is controlled by an airport-interlock system that switches the transmit signal off whenever there is an overhead flight taking off or landing at the Longyearbyen airport to prevent potential HF-disturbances to airborne vehicles.

5.2.1 Testing results

Because of the constraint of -5.1 dBm maximum input power to the amplifier, we measured the maximum output power of the Basic TX USRP daughterboard to determine how much attenuation is needed between the Basic TX and the amplifier. The maximum TX output power from the Basic TX was measured to be -1.7 dBm. To keep the signal power below -5.1 dBm, it would need a 3.5 dB attenuator.

In the implementation, a 6 dB attenuator is added between the Basic TX and switch to fulfill the constraint while also adding a safety margin. The maximum input power the Basic TX can supply to the amplifier is then $-1.7 - 6 = -7.7$ dBm.

Between the directional coupler and the Basic RX within the USRP are 35 dB of attenuation. This is because the maximum input power the Basic RX can handle is -10 dBm. The maximum output power from the amplifier was measured to be about 17 dBm. We know from measurements that by increasing the signal amplitude and lowering the attenuation between the Basic TX and the amplifier, the output power from the amplifier can be as high as 27 dBm. To keep the reflected power from damaging the Basic RX in case the feedline breaks and the entire signal is reflected back into the system, 35 dB of attenuation is added between the directional coupler and the Basic RX.

5.2.2 Effective radiated power

According to the frequency licence, the Effective Radiated Power (ERP) may not exceed 10 W. With the testing results providing the input power going into the transmitter antenna, and the discussion about the antenna gain G in section 3.2.3 we can estimate the ERP of the transmit antenna.

In section 3.2.3 we concluded that the maximum gain of the transmit antenna was $G_T = D\epsilon = 1$, because while the directivity D was that of a half-wave dipole in the best case scenario, the efficiency ϵ could not be assumed to be 1 due to losses in the antenna.

To calculate the ERP, which can defined as

$$\text{ERP} = G_d P_{tx} \tag{5.1}$$

we need to know the power going into the antenna P_{tx} and the gain of the antenna relative to a perfect half-wave dipole G_d. A perfect half-wave dipole has no losses, and the gain is thus 1.64. The relative gain of our antenna is then:

$$G_d = \frac{G_T}{1.64} = 0.6 \tag{5.2}$$

This means that signal power actually decreases as a consequence of being directed through the transmit antenna, which means that to obtain an ERP above 10 W, we need to provide the antenna with at least more than 10 W of power, which we don't do.

While we have ensured that we do not exceed 10 W of ERP, we are still interested in what the actual ERP is. During testing, a 10 dB attenuator between the Basic TX and the amplifier was used, and the signal power set to -7.4 dBm coming out of the Basic TX. The highest peak-to-peak voltage measured on the output to the antenna using this setup was 4.2 V, corresponding to a signal power of about 17 dBm, corresponding to ~ 0.05 W.

With a gain of $G_d = 0.6$ and an input power to the antenna $P_{tx} = 0.05$ W, the ERP is calculated using equation 5.1 to be ERP $= 0.03$ W.

During operation at Breinosa, the signal attenuator in use is a 6 dB attenuator instead of a 10 dB attenuator, which increases the ERP. The signal amplitude is also set to 0.8 which is an increase from the 0.5 amplitude used in the laboratory. Table 4.2 provides the output power for the corresponding amplitudes used in the signal. We use this information to estimate the maximum possible ERP the transmitter can obtain.

Assuming that the amplitude is set to maximum, we simply add the additional dBms introduced to the signal strength by replacing the 10 dB attenuator with a 6 dB attenuator and increasing the signal amplitude: $P_{tx} = 17\,dBm + 4\,dB + 5.7\,dB \approx 27\,dBm$.

A 27 dBm signal has a power of 0.5 W, the ERP in this case is equates to ERP $= 0.3$ W.

In order for the ERP to exceed 10 W the antenna would need to have a gain of $G_d > 10/0.5 = 20$, or equivalently one could increase the input power to the antenna to be at least $P_{tx} > 10/0.6 \approx 16.7$ W.

The calculations to obtain the ERP assumed the gain G_T of the transmit antenna to be 1. However, from estimations of the SWR in figure 4.9 we saw that when transmitting close to the resonant frequency of 6 MHz, the SWR was close to 1. If we assume that at this frequency, the antenna is a perfect half-wave dipole antenna with zero loss, the gain $G_T = D\epsilon = 1.64$, which means that its gain relative to a perfect half-wave dipole $G_d = 1$. The ERP for a 0.5 W signal is then 0.5 W.

It must be noted that the values obtained for the ERP here are based on calculations since we have no real measurement of the antenna gain, and the input power P_{tx} used in equation 5.1 should actually be the root-mean-square (RMS) value of the power going into the antenna[1]. Having said that, the RMS power is less than the power used in our estimations and we can be certain that the ERP does not exceed the license restriction of 10 W.

5.2.3 Software

The software implements a basic ionosonde using a two computers and two USRPs. The repository containing the entire software can be found at `https://github.com/markusfloer/ionosonde.git`

5.3 Discussion

This section provides further insight into some of the results that are more open to interpretation. For instance, the distinction between O- and X-mode reflections discussed in chapter 2 is not made in the ionogram in figure 5.4. We will now go into the reason why this is the case, as well as explore some other aspects of the results.

The O- and X-mode propagation as discussed in section 2.2.1 has different cut-off frequencies, group velocities and polarizations. In order to distinguish the two modes from each other, another receiver antenna could be utilized with a phase offset from the other receive antenna so that the polarization of the two modes could be distinguished. Other types of ionosondes such as

1. ERP definition obtained from Wikipedia: `https://en.wikipedia.org/wiki/Effective_radiated_power`

the Digisonde 4D use receivers perpendicular to each other to achieve this [Reinisch, 2009].

An example were the Longyearbyen ionosonde observes both the O- and X-mode can be seen in figure 5.9.

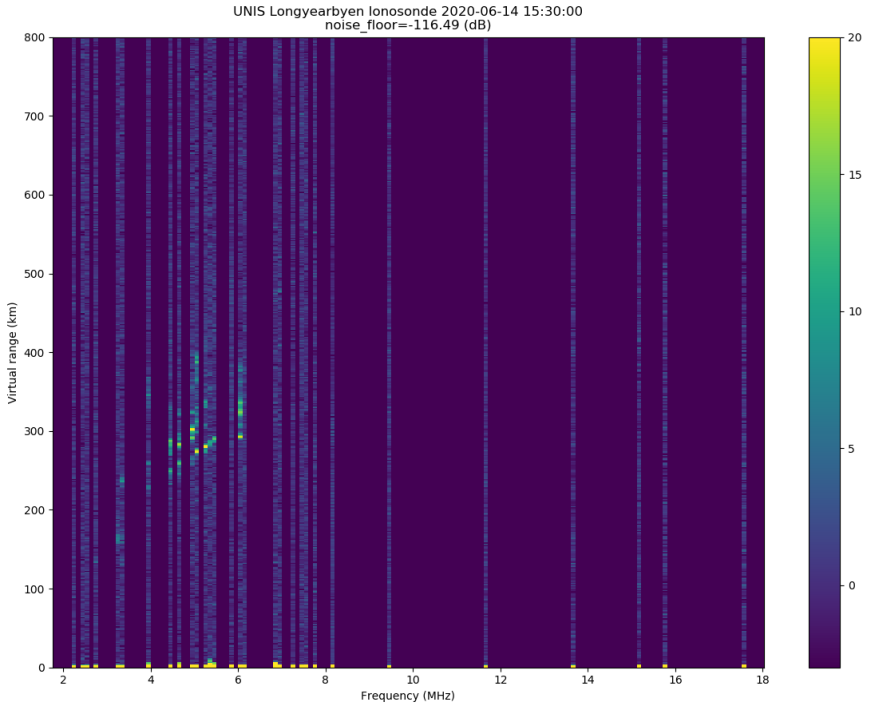


Figure 5.9: Ionogram obtained on the 14th June 2020 15:30 UT. Two similar reflections with some range and frequency offset from one another is observed.

The frequency offset between the O- and X-mode is described by equation 2.20 where the cut-offs are separated by half of the electron gyrofrequency defined in equation 2.7. Using the magnetic field strength around 300 km obtained in table 2.1 yields an electron gyrofrequency of 1.4 MHz. The two modes should then be separated by 0.7 MHz, which is approximately what is seen in figure 5.9, despite the low frequency resolution.

The ionograms displayed in figure 5.4 and 5.9 as well as the overview plot in figure 5.5 display reflected signal power as a function of range. However, the y-axis is labelled "virtual range". The reason for this is because the ionosonde assumes the signal undergoes regular reflection between the transmitter, ionosphere and receiver. It does not consider the effects of refraction of the signal through the ionosphere. The actual range can be obtained by true height analysis. There are several ways of performing true height analysis and some of

them are described in Liu et al. [1992].

Blank ionograms are also observed, this can sometimes be attributed to Polar Cap Absorption (PCA) whereby the plasma is at such low altitudes that collisions with the ambient neutral particles absorb the radar signal completely. It can also be that the level of ionization is so low that the electron-density is too low to provide enough reflected power back to the receiver, or that the plasma-frequency is below the incident wave frequency.

5.3.1 Power consumption and cost of system

An estimation for the power consumed by the Longyearbyen ionosonde is done by adding up the power consumed by all the components within the system. For hardware components such as antennas, switches and amplifiers, the power consumption stays roughly constant during operation. The power consumption of a computer however depends on the way it is used. When doing a first-order estimation of the power consumption by the system, we use the maximum power consumption each component can have, that way we obtain an estimate of the maximum possible power consumption. An estimation for the Longyearbyen ionosonde maximum power consumption is given in table 5.1.

Table 5.1: Maximum power consumption of components within the Longyearbyen ionosonde.

Component	Maximum power consumption
Computers	270 W
USRPs	36 W
Switch & Amp.	11 W
GPSDOs	4 W
ALA1530	2 W
TOTAL	323 W

The main power-consumers within the system are the computers, where their power adapters have a maximum capacity of 135 W each. This means that the computers alone can consume up to 270 W, which makes up about 84% of the total power. The estimation is done from the assumption that the computers use 130 W each, the regular load Wattage is however believed to be smaller. For a more accurate estimation, the power consumption of the computers would have to be measured using a power meter.

Disregarding the computers, the maximum power consumption of the TX/RX

system is 53 W. The transmitter has a maximum power consumption is 31 W. The receiver has a maximum consumption of 22 W.

Table 5.2 presents an estimation of the cost of the Longyearbyen ionosonde and its prototype components.

Table 5.2: Approximate cost of the Longyearbyen ionosonde components.

Components	Cost (NOK)
2x USRP	21000
2x GPSDO	9300
2x PC	8000
Power supply	5700
ALA1530	3000
Switch	2500
Diamond WD330	2400
Amplifier	2200
3x Basic TX/RX	900
2x GPS antenna	600
Directional coupler	500
MFJ LPF	400
BLP LPF	300
4x Attenuators	200
TOTAL	98 300,-

6

Conclusions

In this thesis we have successfully implemented a software defined ionosonde, capable of producing ionograms up to once every minute with up to 1.5 km range resolution. Through the power of software, the range and temporal resolution can be easily adjusted using predefined transmit waveforms. The ionosonde currently operates with a 2 minute temporal resolution and 30 frequency channels between 2 and 18 MHz. The frequency channels have a 1.5 km range resolution, with 4 exceptions of 3 km range resolution (50 kHz) and 1 exception with a 5 km range resolution (30 kHz) due to frequency license restrictions.

The ionosonde demonstrates the viability of using low powered, coded long pulses in the probing of the ionosphere, relying on the assumption that the scattering medium is a coherent radar target.

The ionosonde also demonstrates the viability of Python as the main language for implementation of USRP based radar systems, enabled by the Ettus Research UHD Python API (USRP Hardware Driver Python Application Programming Interface).

The ionosonde is implemented using commercially available, relatively inexpensive hardware. Along with the open source software, this thesis enables the implementation of similar systems by other users, whereby the only necessary prerequisite is a frequency licence.

The resulting system, the new Longyearbyen ionosonde, serves as an addition to the ongoing project of having distributed arrays of small instruments in the observation of the mesoscale dynamics of the upper atmosphere. The current revision of the Longyearbyen ionosonde can easily implement an additional receiver, which will enable higher resolution measurements. The system itself is a continuation of research and development within the field of software defined ionosondes.

6.1 Suggestions for Further Work

6.1.1 Separate O- and X-mode

By adding another receiver antenna perpendicular to the RX magnetic loop antenna, an additional receiver could be implemented by having an extra Basic RX installed in the receiver USRP.

6.1.2 Additional frequencies

With the feasibility of software defined ionosondes being confirmed by this system, it should provide valid reason for the implementation of several frequencies in such a system, providing higher resolution to the system. Higher frequency resolution relies in our case on the licensing by the Norwegian Communications Authority (NKOM). With the Longyearbyen ionosonde demonstrating the capability of software defined ionosondes, reconfiguration of the software allows a simple way to extend the frequency coverage.

6.1.3 Reducing power consumption

The systems main power consumption comes from the computers currently operating the ionosonde. Smaller scale computers such as the Raspberry Pi could alleviate some of the power consumption while still retaining the computational power.

6.1.4 ULF waves

It has been shown that a simple HF Doppler system can measure the small vertical component of the ExB drift velocity in the ionosphere (as a result of non-orthogonality between the Earth's magnetic field and the ground) associated with small scale ULF waves [Baddeley et al., 2005]. The small

scale refers to longitudinal wavelength which can be of the order of tens of kilometers. The waves are "invisible" to traditional instrumentation, such as ground magnetometers which essentially spatially averages over an area in the ionosphere comparable to several wavelengths of the ULF wave structure.

The vertical component of the ExB drift correspond to a measured frequency shift of a few Hz, making it possible to be observed by a system such as the one built here.

A

Dispersion Relation Derivation

Dispersion relation of EM-waves propagating in ionospheric plasma

Including electron-neutral collisions, magnetic field and electron density. Considering only the case for propagation parallel to the magnetic field.

Maxwell's equations describe the electromagnetic waves. We begin with Ampere's law and Faraday's law:

$$\nabla \times B = \frac{1}{c^2 \epsilon_0} J + \frac{1}{c^2} \frac{\partial E}{\partial t} \tag{A.1}$$

$$\nabla \times E = -\frac{\partial B}{\partial t} \tag{A.2}$$

Where $c^2 = 1/\mu_0 \epsilon_0$, c being the speed of light in vacuum, μ_0 and ϵ_0 being the permeability and permittivity of free space respectively. The time-derivative of Ampere's and the curl of Faraday's law yields

$$\nabla \times \frac{\partial B}{\partial t} = \frac{1}{c^2 \epsilon_0} \frac{\partial J}{\partial t} + \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} \tag{A.3}$$

$$\nabla \times \nabla \times E = -\nabla \times \frac{\partial B}{\partial t} \tag{A.4}$$

Inserting A.3 into A.4 gives

$$\nabla \times \nabla \times E = -\frac{1}{c^2\epsilon_0}\frac{\partial J}{\partial t} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} \tag{A.5}$$

This can be further evaluated using a vector identity

$$\nabla(\nabla \cdot E) - \nabla^2 E = -\frac{1}{c^2\epsilon_0}\frac{\partial J}{\partial t} - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} \tag{A.6}$$

$$\nabla^2 E = \frac{1}{c^2\epsilon_0}\frac{\partial J}{\partial t} + \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}, \tag{A.7}$$

Equation A.7 is the wave equation for an electromagnetic wave propagating through a plasma. The additional current term on the right-hand side accounts for the motion of charges that is induced in the plasma by the wave.
Note: The $\nabla(\nabla \cdot E)$ term can only be ignored if there are no longitudinal components of **E** along **k**. In cases where $E_1 \perp B_0$ the waves tend to be elliptically polarized, developing a component of E_1 along k, making the wave partly transverse and partly longitudinal[Chen, 1984, (p. 123)], in that case, the $\nabla(\nabla \cdot E)$ term must then be included, making equation A.6 the one to be considered.

Momentum equation

To describe the induced motion of the charges we use the electron momentum equation (ion motion being neglected because $m_e \ll m_i$) in a collisional, partly ionized, cold plasma ($k_b T_e \approx 0$) with uniform distribution in density and velocity (ie. $v_e \cdot \nabla v_e = 0$)

$$m_e n_e \frac{\partial v_e}{\partial t} = q_e n_e (E + v_e \times B) - m_e n_e \nu_{en} v_e \tag{A.8}$$

Where ν_e is the electron-neutral collision frequency, q_e is the elementary charge. We now describe the velocity and electric field in terms of harmonic functions

$$v_e = \hat{v}_e e^{i(kr-\omega t)} \tag{A.9}$$

$$E = \hat{E} e^{i(kr-\omega t)} \tag{A.10}$$

Where $\hat{E}$ and $\hat{v}_e$ are the amplitudes of the E and v_e waves. These also imply that the current J also will be an oscillating quantity. After canceling the exponential terms we can express J as

$$\widehat{J} = q_e n_e \widehat{v_e} \tag{A.11}$$

The ions are assumed stationary due to their massive weight, so we ignore their effect on the current. Going back to the eln. momentum equation A.8, we see that after harmonic expansion ($\partial_t \rightarrow -i\omega$)

$$-i\omega\hat{v}_e = \frac{q_e}{m_e}(\hat{E} + \hat{v}_e \times B) - \nu_e\hat{v}_e \tag{A.12}$$

$$(-i\omega + \nu_e)\hat{v}_e = \frac{q_e}{m_e}(\hat{E} + \hat{v}_e \times B) \tag{A.13}$$

Setting $(i\omega + \nu_e) = -iW$, using $W = (\omega - i\nu_e)$ the momentum equation becomes

$$-iW\hat{v}_e = \frac{q_e}{m_e}(\hat{E} + \hat{v}_e \times B) \tag{A.14}$$

Since the elementary charge here carries a negative sign due to it representing the electron we write the velocity as

$$\hat{v}_e = \frac{q_e}{im_eW}(\hat{E} + \hat{v}_e \times B) \tag{A.15}$$

Wave equation

We now consider the wave equation with the wave propagating along the magnetic field. B = $B\widehat{z}$, k = $k\widehat{z}$ and $E_1 = E_x\widehat{x} + E_y\widehat{y}$.
With k being strictly perpendicular to E_1 there is no longitudinal component of E_1 along k, we then only need consider equation (A.7)

$$\nabla^2 E = \frac{1}{c^2\epsilon_0}\frac{\partial J}{\partial t} + \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}$$

after harmonic expansion ($\partial_t \rightarrow -i\omega$, $\nabla \rightarrow ik$) the wave equation becomes

$$k^2\hat{E} = \frac{i\omega}{c^2\epsilon_0}\hat{J} + \frac{\omega^2}{c^2}\hat{E} \tag{A.16}$$

$$(\omega^2 - c^2k^2)\hat{E} = -\frac{i\omega}{\epsilon_0}\hat{J} \tag{A.17}$$

From the definition of J we have $\hat{J} = q_e n_e \hat{v}_e$, with the elementary charge q_e being negative since we consider electrons. Inserting this into our wave equation

$$(\omega^2 - c^2k^2)\hat{E} = \frac{i\omega q_e n_e}{\epsilon_0}\hat{v}_e \tag{A.18}$$

To proceed we need to insert for $\hat{v}_e$ from the momentum equation (A.15), but first we need to split $\hat{v}_e$ into its components

$$\hat{v}_e = \frac{q_e}{im_eW}(\hat{E} + \hat{v}_e \times B)$$

By carrying out the cross-product we get the two components of the velocity:

$$v_x = \frac{q_e}{im_eW}(E_x + v_yB_0) \tag{A.19}$$

$$v_y = \frac{q_e}{im_eW}(E_y - v_xB_0) \tag{A.20}$$

By inserting the expression for v_y into v_x we get

$$\begin{aligned}
v_x &= \frac{q_e}{im_eW}\left(E_x + \frac{q_e}{im_eW}(E_y - v_xB_0)B_0\right) \\
v_x &= \frac{q_e}{im_eW}E_x - \frac{q_e^2}{m_e^2W^2}(E_y - v_xB_0)B_0 \\
v_x - v_x\frac{q_e^2B_0^2}{m_e^2W^2} &= \frac{q_e}{im_eW}E_x - \frac{q_e^2B_0}{m_e^2W^2}E_y \\
v_x\left(1 - \frac{\omega_c^2}{W^2}\right) &= \frac{q_eE_x(-im_eW)}{im_eW(-im_eW)} - \frac{q_e^2B_0}{m_e^2W^2}E_y \\
v_x\left(1 - \frac{\omega_c^2}{W^2}\right) &= \frac{-iq_eE_x}{m_eW} - \frac{q_e^2B_0}{m_e^2W^2}E_y \\
v_x\left(1 - \frac{\omega_c^2}{W^2}\right) &= \frac{q_e}{m_eW}\left(-iE_x - \frac{\omega_c}{W}E_y\right) \\
v_x &= \frac{q_e}{m_eW}\left(-iE_x - \frac{\omega_c}{W}E_y\right)\left(1 - \frac{\omega_c^2}{W^2}\right)^{-1}
\end{aligned}$$

Where ω_c is the electron cyclotron frequency $\omega_c = \frac{q_eB_0}{m_e}$. Similarly, by inserting expression for v_x in A.20 into v_y we get

$$v_y = \frac{q_e}{m_eW}\left(-iE_y + \frac{\omega_c}{W}E_x\right)\left(1 - \frac{\omega_c^2}{W^2}\right)^{-1}$$

To make use of the velocity components we also need to separate the wave equation into its respective components

$$(\omega^2 - c^2k^2)\hat{E}_x = \frac{i\omega q_en_e}{\epsilon_0}\hat{v}_x \tag{A.21}$$

$$(\omega^2 - c^2k^2)\hat{E}_y = \frac{i\omega q_en_e}{\epsilon_0}\hat{v}_y \tag{A.22}$$

Inserting v_x and v_y:

$$(\omega^2 - c^2k^2)\hat{E}_x = \frac{\omega_p^2}{1 - \omega_c^2/W^2}\left(E_x - \frac{i\omega_c}{W}E_y\right) \tag{A.23}$$

$$(\omega^2 - c^2k^2)\hat{E}_y = \frac{\omega_p^2}{1 - \omega_c^2/W^2}\left(E_y - \frac{i\omega_c}{W}E_x\right) \tag{A.24}$$

Where $\omega_p = \frac{q_e^2 n_e}{m_e \epsilon_0}$ is the electron plasma frequency. We define α as

$$\alpha = \frac{\omega_p^2}{1 - \omega_c^2/W^2}$$

and rewrite the coupled equations as:

$$\begin{aligned}(\omega^2 - c^2k^2 - \alpha)E_x + i\alpha\frac{\omega_c}{W}E_y = 0 \\ (\omega^2 - c^2k^2 - \alpha)E_y - i\alpha\frac{\omega_c}{W}E_x = 0\end{aligned} \tag{A.25}$$

Equation (A.25) is still our wave-equation, separated into components of x an y, whilst propagating along z. The determinant of this system of equations must equate to zero

$$det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = 0 \tag{A.26}$$

Where

$$\begin{aligned}A &= \omega^2 - c^2k^2 - \alpha \\ B &= -i\alpha\omega_c/W \\ C &= \omega^2 - c^2k^2 - \alpha \\ D &= i\alpha\omega_c/W\end{aligned}$$

That is;

$$\begin{aligned}AD &= BC \\ (\omega^2 - c^2k^2 - \alpha)^2 &= \left(\frac{\alpha\omega_c}{W}\right)^2 \\ &\Rightarrow \omega^2 - c^2k^2 - \alpha = \pm\frac{\alpha\omega_c}{W}\end{aligned}$$

Thus, we have

$$\omega^2 - c^2k^2 = \alpha(1 \pm \frac{\omega_c}{W}) \tag{A.27}$$

$$\omega^2 - c^2k^2 = \frac{\omega_p^2}{1 - \omega_c^2/W^2}(1 \pm \frac{\omega_c}{W}) \tag{A.28}$$

$$\omega^2 - c^2k^2 = \omega_p^2\frac{1 \pm \frac{\omega_c}{W}}{(1 + \frac{\omega_c}{W})(1 - \frac{\omega_c}{W})} \tag{A.29}$$

$$\omega^2 - c^2k^2 = \frac{\omega_p^2}{1 \mp \frac{\omega_c}{W}} \tag{A.30}$$

The $\mp$ sign indicates that there are two possible solutions to the system of coupled equations (A.25), corresponding to two different waves that can propagate along B. The dispersion relations are

$$\frac{c^2k^2}{\omega^2} = 1 - \frac{\omega_p^2}{1 - \frac{\omega_c}{W}} \tag{A.31}$$

$$\frac{c^2k^2}{\omega^2} = 1 - \frac{\omega_p^2}{1 + \frac{\omega_c}{W}} \tag{A.32}$$

Since $W = (\omega - i\nu_n)$

$$\frac{c^2k^2}{\omega^2} = 1 - \frac{\omega_p^2}{1 - \frac{\omega_c}{(\omega - i\nu_e)}} \qquad \text{(R-wave)} \tag{A.33}$$

$$\frac{c^2k^2}{\omega^2} = 1 - \frac{\omega_p^2}{1 + \frac{\omega_c}{(\omega - i\nu_e)}} \qquad \text{(L-wave)} \tag{A.34}$$

The R-wave referring to a right-hand circularly polarization and L-wave referring to a left-hand circularly polarization. The collision frequency ν_e acts as an attenuating factor.

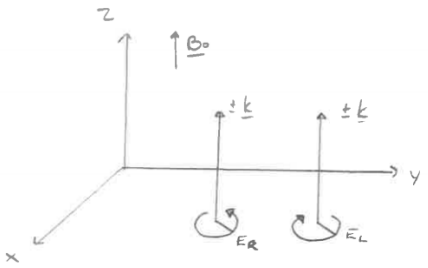


Figure A.1: The two wavemodes that propagate along **B**. One with the E-field right-hand circularly polarized and the other with its E-field left-hand circularly polarized.

B

Appleton-Hartree Derivation

Deriving Appleton Hartree equation

Since the ionospheric plasma is an anisotropic medium we have

$$\mathrm{D} = \epsilon_0 \mathrm{E} + \mathrm{P} \tag{B.1}$$

The magnetic fields is that of free space

$$\mathrm{B} = \mu_0 \mathrm{H} \tag{B.2}$$

Considering a plane wave traveling in the x-direction

$$E = E_0 \cos(\omega t - kx) \tag{B.3}$$
$$= E_0 e^{i(\omega t - kx)} \tag{B.4}$$

The properties of the differential of this wave is

$$\frac{\partial}{\partial t} = i\omega, \qquad \frac{\partial^2}{\partial t^2} = -\omega^2, \qquad \nabla = ik$$

This yields for the electric and magnetic field

$$D_x = \epsilon_0 E_x + P_x \tag{B.5}$$
$$B_x = \mu_0 H_x \tag{B.6}$$

Applying Faradays law and Amperes law yields:

$$\begin{aligned} -ikH_z &= i\omega(\epsilon_0 E_y + P_y) \\ ikE_z &= -\mu_0 i\omega H_y \\ ikH_y &= i\omega(\epsilon_0 E_z + P_z) \\ ikE_y &= -\mu_0 i\omega H_z \end{aligned} \tag{B.7}$$

This is equivalent to

$$\epsilon_0 E_y + P_y = -\frac{k}{\omega} H_z = D_y \tag{B.8}$$

$$\epsilon_0 E_z + P_z = \frac{k}{\omega} H_y = D_z \tag{B.9}$$

$$kE_z = \mu_0 \omega H_y \tag{B.10}$$

$$-kE_y = \mu_0 \omega H_z \tag{B.11}$$

Next we express the refractive index in terms of E's and P's by substituting H_z from B.8 and B.11:

$$\epsilon_0 E_y + P_y = -\frac{k}{\omega}\left(-\frac{kE_y}{\mu_0 \omega}\right) = \frac{k^2}{\omega^2 \mu_0} E_y \tag{B.12}$$

$$1 + \frac{1}{\epsilon_0}\frac{P_y}{E_y} = \left(\frac{k}{\omega}\right)^2 \frac{1}{\mu_0 \epsilon_0} = \left(\frac{ck}{\omega}\right)^2 \tag{B.13}$$

From one of the definitions of the refractive index we have

$$n = \frac{ck}{\omega} \tag{B.14}$$

Hence

$$n^2 = 1 + \frac{1}{\epsilon_0}\frac{P_y}{E_y},$$

similarly for the other component: (B.15)

$$n^2 = 1 + \frac{1}{\epsilon_0}\frac{P_z}{E_z}$$

It follows from B.15 that

$$\frac{E_y}{E_z} = \frac{P_y}{P_z} \tag{B.16}$$

and from B.10 and B.11 we obtain

$$-\frac{H_y}{H_z} = \frac{E_z}{E_y} = \frac{P_z}{P_y} = R \tag{B.17}$$

where R is defined as the wave-polarization. We now turn to the electron momentum equation to obtain some constitutive relations, consider a wave with propagation along the x-axis and the magnetic field in the zy-plane. The components of the electron momentum equation are

$$\begin{aligned} m\ddot{x}_x &= qE_x - q\dot{x}_z B_T - m\nu\dot{x}_x \\ m\ddot{x}_y &= qE_y - q\dot{x}_z B_L - m\nu\dot{x}_y \\ m\ddot{x}_z &= qE_z + q\dot{x}_x B_T - q\dot{x}_y B_T - m\nu\dot{x}_z \end{aligned} \tag{B.18}$$

Multiplying through by the electron density n_e, replacing the differentials with ω and replacing $n_e q x_x$ with P_x etc. yields the constitutive relations

$$\begin{aligned} \epsilon_0 X E_x &= -P_x + iP_x Z + iP_Z Y_T \\ \epsilon_0 X E_y &= -P_y + iP_y Z + iP_Z Y_L \\ \epsilon_0 X E_z &= -P_z + iP_z Z - iP_x Y_T + iP_y Y_L \end{aligned} \tag{B.19}$$

where X, Y and Z are the same as in 2.13, with subscripts T and L referring to the transverse and longitudinal component. From B.5 we see that

$$\epsilon_0 X E_x = P_x \tag{B.20}$$

Substituting this into the first relation in B.19 gives

$$\frac{P_x}{P_y} = \frac{iY_T}{1 - X - iZ} \tag{B.21}$$

From (57) we have that $P_y E_z = P_z E_y$. Multiplying the second equation in B.5 by P_z and the third equation in B.5 by P_y yields upon subtracting the two

$$P_z^2 Y_L = P_x P_y Y_T - P_y^2 Y_L \tag{B.22}$$

By rearranging and inserting for $R = P_z/P_y$ we get

$$Y_L R^2 - \frac{-Y_T^2}{1 - X - iZ} R + Y_L = 0 \tag{B.23}$$

This quadratic equation has two solutions, given by

$$R = \frac{i}{2Y_L}\left[\frac{Y_T^2}{1 - X - iZ} \mp \sqrt{\frac{Y_T^4}{(1 - X - iZ)^2} + 4Y_L^2}\right] \tag{B.24}$$

From B.19 we get

$$-\frac{P_y}{\epsilon_0 E_y} = \frac{X}{1 - iZ + iY_L R} \tag{B.25}$$

which, when substituting into B.15 yields

$$n^2 = 1 - \frac{X}{1 - iZ - \left(\frac{Y_T^2}{2(1-X-iZ)}\right) \pm \left(\frac{Y_T^4}{4(1-X-iZ)^2} + Y_L^2\right)^{\frac{1}{2}}} \tag{B.26}$$

which is the Appleton Hartreee equation.

C

Set-up and Installation Procedure

This chapter describes the process of installing and setting up the necessary software in order to transmit and receive with the USRP platforms. The installation procedure is described here using the Linux Ubuntu operating system, for a description of the installation procedure on Windows or Mac OS, I invite the reader to visit https://files.ettus.com/manual/page_install.html for further instructions.

C.1 Hardware dependencies

The software relies on USRP N2xo software defined radio hardware. The minimum requirement is:

- Ettus Research USRP N2xo + internal GPSDO with BasicRX or LFRX daughterboard for receiving
- Ettus Research USRP N2xo + internal GPSDO with BasicRX or LFRX daughterboard for receiving and BasicTX or LFTX daugherboard of transmitting. The receiver card on the transmitter can be used to measure relfected or transmitted power using a directional coupler.

- We rely on the internal Ettus Research GPSDO and use the UHD commands to interface with the GPSDO. It is possible to use another made, but you'll need to come up with an alternative interface to check for GPS lock.

- You choice of transmit and receiver antennas, and associated RF plumbing. An example is shown in figure 5.8.

- 2 PCs - one to control the transmitter and one to control the receiver. These will often be in different places. For a monostatic radar, it is possible to use the same PC to control the transmitter and receiver. The receiver signal processing is not extremely demanding, and should work with a 10+ year old entry level CPU.

C.2 Software dependencies

- Requires Linux. We've tested the program using Ubuntu 18.04 LTS. It should be possible to adapt the code relatively easily to any operating system and platform that supports Python.

- Requires UHD Library 3.15. The UHD library needs to be compiled with the Python API enabled.

- Numpy, Matplotlib, Scipy, Psutil

C.3 Installation Instructions

Installing UHD 3.15 with Python API enabled:

```
sudo apt-get install libopenblas-dev python3-matplotlib \
  python3-psutil python3-h5py python3-setuptools

sudo apt-get -y install git swig cmake doxygen \
  build-essential libboost-all-dev libtool libusb-1.0-0 \
  libusb-1.0-0-dev libudev-dev libncurses5-dev libfftw3-bin \
  libfftw3-dev libfftw3-doc libcppunit-1.14-0 libcppunit-dev \
  libcppunit-doc ncurses-bin cpufrequtils python-numpy \
  python-numpy-doc python-numpy-dbg python-scipy python-docutils \
```

```
qt4-bin-dbg qt4-default qt4-doc libqt4-dev libqt4-dev-bin \
python-qt4 python-qt4-dbg python-qt4-dev python-qt4-doc \
python-qt4-doc libqwt6abi1 libfftw3-bin libfftw3-dev \
libfftw3-doc ncurses-bin libncurses5 libncurses5-dev \
libncurses5-dbg libfontconfig1-dev libxrender-dev  \
libpulse-dev swig g++ automake autoconf libtool \
python-dev libfftw3-dev libcppunit-dev libboost-all-dev \
libusb-dev libusb-1.0-0-dev fort77 libsdl1.2-dev \
python-wxgtk3.0 git libqt4-dev python-numpy ccache \
python-opengl libgsl-dev python-cheetah python-mako \
python-lxml doxygen qt4-default qt4-dev-tools \
libusb-1.0-0-dev libqwtplot3d-qt5-dev pyqt4-dev-tools \
python-qwt5-qt4 cmake git wget libxi-dev gtk2-engines-pixbuf \
r-base-dev python-tk liborc-0.4-0 liborc-0.4-dev libasound2-dev \
python-gtk2 libzmq3-dev libzmq5 python-requests python-sphinx \
libcomedi-dev python-zmq libqwt-dev libqwt6abi1 python-six \
libgps-dev libgps23 gpsd gpsd-clients python-gps python-setuptools
```

```
wget https://github.com/EttusResearch/uhd/archive/v3.15.0.0.tar.gz
tar xvfz v3.15.0.0.tar.gz
cd uhd
cd host
mkdir build
cd build
cmake -DENABLE_PYTHON_API=ON ../
make
sudo make install
```

D

Source Code

The complete repository with all the software used throughout this project is found on: `https://github.com/markusfloer/ionosonde`

The reader is free to download, clone, modify or otherwise use the code as seen fit according to the GNU General Public License, which can be found in the repository.

The three main main components of the source code for the ionosonde is listed here; *tx_uhd.py*, *rx_uhd.py* and *prc_lib.py*(which *analyze_ionograms.py* relies on). See the link to repository for further description of other internal modules such as *sweep.py*, *iono_config.py* and *gps_lock.py*.

Listing D.1: *tx_uhd.py*; Frequency hopping pulse modulated signal transmitter

```
#!/usr/bin/env python3
#
# Copyright 2020, Juha Vierinen, Markus Floer
#
# SPDX-License-Identifier: GPL-3.0-or-later
#
"""
Generate and TX samples using a set of waveforms, and waveform characteristics
"""

import argparse
import uhd
import time
import threading
import numpy as n
import matplotlib.pyplot as plt
import os

# internal modules related with the ionosonde
import sweep
import gps_lock as gl
```

```
import iono_logger as l
import iono_config


def tune_at(u, t0, f0=4e6, gpio_state=0):
    """
    tune radio to frequency f0 at t0_full
    use a timed command.
    """
    u.clear_command_time()
    t0_ts = uhd.libpyuhd.types.time_spec(t0)
    print("tuning_to_%1.2f_MHz_at_%1.2f" % (f0 / 1e6, t0_ts.get_real_secs()))
    u.set_command_time(t0_ts)
    tune_req = uhd.libpyuhd.types.tune_request(f0)
    u.set_tx_freq(tune_req)
    u.set_rx_freq(tune_req)

    if gpio_state == 0:
        out = 0x00
    else:
        out = 0x01
    gpio_line = 0x01  # pin 0
    print("TX_A_GPIO_pin_0=%d" % (out))
    u.set_gpio_attr("TXA", "OUT", out, gpio_line, 0)
    u.clear_command_time()


def tx_send(tx_stream, waveform, md, timeout=11.0):
    # this command will block until everything is in the transmit
    # buffer.
    tx_stream.send(waveform, md, timeout=(len(waveform) / float(iono_config.sample_rate)) + 1.0)


def rx_swr(u, t0, recv_buffer, f0, log):
    """
    Receive samples for a reflected power measurement
    USRP output connected to input with 35 dB attenuation gives
    9.96 dB reflected power.
    """
    N = len(recv_buffer)
    stream_args = uhd.usrp.StreamArgs("fc32", "sc16")
    rx_stream = u.get_rx_stream(stream_args)
    stream_cmd = uhd.types.StreamCMD(uhd.types.StreamMode.num_done)
    stream_cmd.num_samps = N
    stream_cmd.stream_now = False
    stream_cmd.time_spec = uhd.types.TimeSpec(t0)
    rx_stream.issue_stream_cmd(stream_cmd)
    md = uhd.types.RXMetadata()
    num_rx_samps = rx_stream.recv(recv_buffer, md, timeout=float(N / iono_config.sample_rate) + 1.0)
    pwr = n.mean(n.abs(recv_buffer) ** 2.0)
    rx_stream = None
    refl_pwr_dBm = 10.0 * n.log10(pwr) + iono_config.reflected_power_cal_dB
    log.log("reflected_pwr_%1.4f_(MHz)_%1.4f_(dBm)" % (f0, refl_pwr_dBm))


def transmit_waveform(u, t0_full, waveform, swr_buffer, f0, log):
    """
    Transmit a timed burst
    """
    t0_ts = uhd.libpyuhd.types.time_spec(n.uint64(t0_full), 0.0)
    stream_args = uhd.usrp.StreamArgs("fc32", "sc16")
    md = uhd.types.TXMetadata()
    md.has_time_spec = True
    md.time_spec = t0_ts

    print("transmit_start_at_%1.2f" % (t0_full))

    # wait for moment right before transmit
    t_now = u.get_time_now().get_real_secs()
    print("setup_time_%1.2f" % (t_now))
    if t_now > t0_full:
        log.log("Delayed_start_for_transmit_%1.2f_%1.2f" % (t_now, t0_full))
    if t0_full - t_now > 0.1:
        time.sleep(t0_full - t_now - 0.1)

    try:
        # transmit the code
        tx_stream = u.get_tx_stream(stream_args)
        tx_thread = threading.Thread(target=tx_send, args=(tx_stream, waveform, md))
        tx_thread.daemon = True  # exit if parent thread exits
        tx_thread.start()

        # do an swr measurement
        rx_thread = threading.Thread(target=rx_swr, args=(u, t0_full, swr_buffer, f0, log))
```

```
            rx_thread.daemon = True  # exit if parent thread exits
            rx_thread.start()
            tx_thread.join()
            rx_thread.join()
            tx_stream = None
        except:
            exit(0)


def main():
    """
    The main loop for the ionosonde transmitter
    """
    t_start = time.time()
    logfname = "tx-%d.log" % (t_start)
    log = l.logger(logfname)
    os.system("ln -sf %s tx-current.log" % (logfname))
    log.log("Starting TX sweep", print_msg=True)

    # this is the sweep configuration
    s = iono_config.s

    sample_rate = iono_config.sample_rate

    # use the address configured for the transmitter
    usrp = uhd.usrp.MultiUSRP("addr=%s" % (iono_config.tx_addr))
    usrp.set_tx_rate(sample_rate)
    usrp.set_rx_rate(sample_rate)

    rx_subdev_spec = uhd.usrp.SubdevSpec(iono_config.rx_subdev)
    tx_subdev_spec = uhd.usrp.SubdevSpec(iono_config.tx_subdev)

    usrp.set_tx_subdev_spec(tx_subdev_spec)
    usrp.set_rx_subdev_spec(rx_subdev_spec)

    # wait until GPS is locked, then align USRP time with global ref
    gl.sync_clock(usrp, log, min_sync_time=iono_config.min_gps_lock_time)

    # start with first frequency on tx and rx
    tune_req = uhd.libpyuhd.types.tune_request(s.freq(0))
    usrp.set_tx_freq(tune_req)
    usrp.set_rx_freq(tune_req)

    # hold SWR measurement about half of the transmit waveform length, so
    # we have no timing issues
    swr_buffer = n.empty(int(0.5 * sample_rate * s.freq_dur), dtype=n.complex64)

    # figure out when to start the cycle
    t_now = usrp.get_time_now().get_real_secs()
    t0 = n.uint64(n.floor(t_now / (s.sweep_len_s)) * s.sweep_len_s + s.sweep_len_s)
    print("starting next sweep at %1.2f" % (s.sweep_len_s))

    gpio_state = 0
    while True:
        log.log("Starting sweep at %1.2f" % (t0))
        for i in range(s.n_freqs):
            f0, dt = s.pars(i)

            print("f=%f bandwidth %d kHz" % (f0, s.bw(i) * 1e3))
            transmit_waveform(usrp, n.uint64(t0 + dt), s.waveform(i), swr_buffer, f0, log)

            # tune to next frequency 0.0 s before end
            next_freq_idx = (i + 1) % s.n_freqs
            tune_at(usrp, t0 + dt + s.freq_dur - 0.05, f0=s.freq(next_freq_idx), gpio_state=gpio_state
                )
            gpio_state = (gpio_state + 1) % 2

            # check that GPS is still locked.
            gl.check_lock(usrp, log, exit_if_not_locked=True)

        t0 += n.uint64(s.sweep_len_s)


if __name__ == "__main__":
    main()
```

Listing D.2: *rx_uhd.py*; Frequency hopping raw voltage recorder

```
#!/usr/bin/env python
#
# Copyright 2020, Juha Vierinen, Markus Floer
#
# SPDX-License-Identifier: GPL-3.0-or-later
```

```
#
"""
 Ionosonde receiver. Receive signals based on a fixed program using
 GPS timing.
"""
import argparse
import numpy as np
import uhd
import time
import threading
import numpy as n
import matplotlib.pyplot as plt
import sweep
import iono_logger as l
import gps_lock as gl
import stuffr
import glob
import re
import os
import iono_config
import scipy.signal
import os
import psutil

def tune_at(u,t0,f0=4e6):
    """
    tune radio to frequency f0 at t0_full
    use a timed command.
    """
    u.clear_command_time()
    t0_ts=uhd.libpyuhd.types.time_spec(t0)
    print("tuning_at_%1.2f"%(t0_ts.get_real_secs()))
    u.set_command_time(t0_ts)
    tune_req=uhd.libpyuhd.types.tune_request(f0)
    u.set_tx_freq(tune_req)
    u.set_rx_freq(tune_req)
    u.clear_command_time()

def delete_old_files(t0,data_path="/dev/shm"):
    """
    Deleting files older than two cycles ago
    """
    # delete older files
    fl=glob.glob("%s/raw*.bin"%(data_path))
    fl.sort()
    for f in fl:
        try:
            tfile=int(re.search(".*/raw-(.*)-....bin",f).group(1))
            if tfile < t0:
                os.system("rm_%s"%(f))
        except:
            print("Error_deleting_file_%s"%(f))


def write_to_file(recv_buffer,fname,log,dec=10,fl=20):
    print("writing_to_file_%s"%(fname))

    # filter and decimate with Blackmann-Harris window
    w = n.zeros(fl, dtype=n.complex64)
    w[0:fl] = scipy.signal.blackmanharris(fl)
    # filter, time shift, decimate, and cast to complex64 data type
    obuf=n.array(n.roll(n.fft.ifft(n.fft.fft(w,len(recv_buffer))*n.fft.fft(recv_buffer)),-int(fl/2))
          [0:len(recv_buffer):dec],dtype=n.complex64)

    # rectangular impulse response. better for range resolution,
    # but not very good for frequency selectivity.
#    obuf=stuffr.decimate(recv_buffer,dec=dec)
    obuf.tofile(fname)

def receive_continuous(u,t0,t_now,s,log,sample_rate=1000000.0):
    """
    New receive script, which processes data incoming from the usrp
    one packet at a time.
    """

    # sweep timing and frequencies
    fvec=[]
    t0s=[]
    for i in range(s.n_freqs):
        f,t=s.pars(i)
        fvec.append(f)
        t0s.append(t)

    # setup usrp to stream continuously, starting at t0
```

```
stream_args=uhd.usrp.StreamArgs("fc32","sc16")
rx_stream=u.get_rx_stream(stream_args)
stream_cmd = uhd.types.StreamCMD(uhd.types.StreamMode.start_cont)
stream_cmd.stream_now=False
stream_cmd.time_spec=uhd.types.TimeSpec(t0)
rx_stream.issue_stream_cmd(stream_cmd)
md=uhd.types.RXMetadata()

# this is how many samples we expect to get from each packet
max_samps_per_packet = rx_stream.get_max_num_samps()

# receive buffer size large enough to fit one packet
recv_buffer=n.zeros(max_samps_per_packet,dtype=n.complex64)

# initial timeout is long enough for us to receive the first packet, which
# happens at t0
timeout=(t0-t_now)+5.0

# store data in this ringbuffer, and offload it to ram disk once
# one full cycle is finished
output_buffer = n.zeros(2*int(s.freq_dur*sample_rate),dtype=n.complex64)

# use this buffer to write to a file
wr_buff = n.zeros(int(s.freq_dur*sample_rate),dtype=n.complex64)
bl=len(output_buffer)
bi=0

# frequency index
fi = 0
# samples since 1970 for the previous packet (no previous packet at first)
prev_samples = -1

# samples since 1970 for the first packet.
samples0=int(stream_cmd.time_spec.get_full_secs())*int(sample_rate) + int(stream_cmd.time_spec.
      get_frac_secs()*sample_rate)

# number of samples per frequency in the sweep
n_per_freq=int(s.freq_dur*sample_rate)
n_per_sweep=int(s.sweep_len_s*sample_rate)

sweep_num=0
freq_num=0
next_sample = samples0 + n_per_freq
cycle_t0 = t0

tune_at(u,t0+s.freq_dur,f0=s.freq(1))

locked=True
try:
    while locked:
        num_rx_samps=rx_stream.recv(recv_buffer,md,timeout=timeout)
        if num_rx_samps == 0:
            # shit happened. we probably lost a packet. gotta try again
            log.log("dropped_packet._number_of_received_samples_is_0")
            continue

        # the start of the buffer is at this sample index
        samples=int(md.time_spec.get_full_secs())*int(sample_rate) + int(md.time_spec.
              get_frac_secs()*sample_rate)

        # this is how many samples we have jumped forward.
        step = samples-prev_samples

        if prev_samples == -1:
            step = num_rx_samps

        if step != 363 or num_rx_samps != 363:
            log.log("anomalous_step_%d_num_rx_samps_%d_"%(step,num_rx_samps))

        prev_samples=samples

        # write the result into the output buffer
        output_buffer[ n.mod(bi+n.arange(num_rx_samps,dtype=n.uint64),bl) ]=recv_buffer

        bi=bi+step

        if samples > next_sample:
            # this should be correct now.
            idx0=sweep_num*n_per_sweep+freq_num*n_per_freq

            wr_buff[:]=output_buffer[n.mod(idx0+n.arange(n_per_freq,dtype=n.uint64),bl)]

            # spin of a thread to write all samples obtained while sounding this frequency
            wr_thread=threading.Thread(target=write_to_file,args=(wr_buff,"%s/raw-%d-%03d.bin"%(
```

```
                    iono_config.data_path,cycle_t0,freq_num),log))
                wr_thread.start()
                freq_num += 1

                # setup tuning for next frequency
                tune_at(u,cycle_t0 + (freq_num+1)*s.freq_dur,f0=s.freq(freq_num+1))
                print("Tuning to %1.2f at %1.2f"%(s.freq(freq_num+1)/1e6,cycle_t0 + (freq_num+1)*s.
                    freq_dur))

                # the cycle is over
                if freq_num == s.n_freqs:
                    cycle_t0 += s.sweep_len_s
                    freq_num=0
                    sweep_num+=1
                    locked=gl.check_lock(u,log,exit_if_not_locked=False)
                    log.log("Starting new cycle at %1.2f"%(cycle_t0))

                # we've got a full freq step
                next_sample += n_per_freq

            timeout=0.1
    except:
        print("interrupt")
        pass
    print("Issuing stop command...")

    stream_cmd = uhd.types.StreamCMD(uhd.types.StreamMode.stop_cont)
    rx_stream.issue_stream_cmd(stream_cmd)
    time.sleep(1)
    print("Stream stopped")
    exit(0)
    return

def housekeeping(usrp,log,s):
    while True:
        t0=usrp.get_time_now().get_real_secs()
        delete_old_files(int(t0)-int(s.sweep_len_s)*3,iono_config.data_path)
        t0+=np.uint64(s.sweep_len_s)

        process = psutil.Process(os.getpid())
        log.log("Memory use %1.5f (MB)"%(process.memory_info().rss/1e6))

        time.sleep(s.sweep_len_s)

def main():
    """
    Start up everything and run main loop from here.
    """
    # setup a logger
    logfile="rx-%d.log"%(time.time())
    log=l.logger(logfile)
    log.log("Starting receiver")
    os.system("rm rx-current.log;ln -s %s rx-current.log"%(logfile))

    s=iono_config.s
    log.log("Sweep freqs:")
    log.log(str(s.freqs))
    log.log("Sweep length %1.2f s Freq step %1.2f"%(s.sweep_len_s,s.freq_dur))

    # Configuring USRP
    sample_rate=iono_config.sample_rate

    # number of samples per freq
    N=int(sample_rate*s.freq_dur)

    # configure usrp
    usrp = uhd.usrp.MultiUSRP("addr=%s,recv_buff_size=500000000"%(iono_config.rx_addr))
    usrp.set_rx_rate(sample_rate)
    subdev_spec=uhd.usrp.SubdevSpec(iono_config.rx_subdev)
    usrp.set_rx_subdev_spec(subdev_spec)

    # Synchronizing clock
    gl.sync_clock(usrp,log)

    # figure out when to start the cycle.
    t_now=usrp.get_time_now().get_real_secs()
    t0=np.uint64(np.floor(t_now/(s.sweep_len_s))*s.sweep_len_s+s.sweep_len_s)
    print("starting next sweep at %1.2f in %1.2f s, time now %1.2f"%(t0,t0-t_now,t_now))

    # start with initial frequency
    tune_req=uhd.libpyuhd.types.tune_request(s.freq(0))
    usrp.set_rx_freq(tune_req)
```

```
    # start reading data
    housekeeping_thread=threading.Thread(target=housekeeping,args=(usrp,log,s))
    housekeeping_thread.daemon=True
    housekeeping_thread.start()

    # infinitely loop on receive
    receive_continuous(usrp,t0,t_now,s,log)

if __name__ == "__main__":
    main()
```

Listing D.3: *prc_lib.py*; Matrix deconvolution of known pseudo-random waveform echo

```
#!python
# ----------------------------------------------------------------------------
# Copyright (c) 2017 Massachusetts Institute of Technology (MIT)
# All rights reserved.
#
# Distributed under the terms of the BSD 3-clause license.
#
# The full license is in the LICENSE file, distributed with this software.
# ----------------------------------------------------------------------------
"""Script for analyzing pseudorandom-coded waveforms.
See the following paper for a description and application of the technique:
Vierinen, J., Chau, J. L., Pfeffer, N., Clahsen, M., and Stober, G.,
Coded continuous wave meteor radar, Atmos. Meas. Tech., 9, 829-839,
doi:10.5194/amt-9-829-2016, 2016.
"""
import datetime
import glob
import itertools
import math
import os
import time
import h5py
from argparse import ArgumentParser
import stuffr

import numpy as np
import scipy.signal
import scipy.fftpack as sf


import create_waveform


def periodic_convolution_matrix(envelope, rmin=0, rmax=100):    # Function that implements a Toeplitz
    matrix
    """
    we imply that the number of measurements is equal to the number of elements
    in code
    """
    L = len(envelope)
    ridx = np.arange(rmin, rmax)
    A = np.zeros([L, rmax - rmin], dtype=np.complex64)
    for i in np.arange(L):
        A[i, :] = envelope[(i - ridx) % L]
    result = {}
    result['A'] = A
    result['ridx'] = ridx
    return (result)


def create_estimation_matrix(code, rmin=0, rmax=1000, cache=True):

    r_cache = periodic_convolution_matrix(envelope=code, rmin=rmin, rmax=rmax)
    A = r_cache['A']
    Ah = np.transpose(np.conjugate(A))
    # least-squares estimate
    # B=(A^H A)^{-1}A^H
    B_cache = np.dot(np.linalg.inv(np.dot(Ah, A)), Ah)

    r_cache['B'] = B_cache
    B_cached = True
    return (r_cache)


def analyze_prc(zin,
                clen=10000,
                station=0,
                Nranges=1000,
```

```
            rfi_rem=True,
            spec_rfi_rem=False,
            cache=True,
            gc_rem=False,
            #                wfun=scipy.signal.blackmanharris(N),
            wfun=scipy.signal.tukey,
            #                wfun=1.0,
            gc=20,
            fft_filter=False,
            code_type="prn",
            dec=10):
"""
Analyze pseudorandom code transmission for a block of data.
idx0 = start idx
an_len = analysis length
clen = code length
station = random seed for pseudorandom code
cache = Do we cache (\conj(A^T)\*A)^{-1}\conj{A}^T for linear least squares
    solution (significant speedup)
rfi_rem = Remove RFI (whiten noise).
"""
if code_type == "perfect":
    code = create_waveform.create_prn_dft_code(clen=clen, seed=station)
else:
    code = create_waveform.create_pseudo_random_code(clen=clen, seed=station)
an_len = len(zin) / dec
N = int(an_len / clen)
res = np.zeros([N, Nranges], dtype=np.complex64)
#    reso = np.zeros([N, Nranges], dtype=np.complex64)

# use cached version of (A^HA)^{-1}A^H if it exists.
if os.path.exists("waveforms/b-%d-%d.h5" % (station, Nranges)):
    hb = h5py.File("waveforms/b-%d-%d.h5" % (station, Nranges), "r")
    B = hb["B"].value
    hb.close()
else:
    r = create_estimation_matrix(code=code, cache=cache, rmax=Nranges)
    B = r['B']
    hb = h5py.File("waveforms/b-%d-%d.h5" % (station, Nranges), "w")
    hb["B"] = B
    hb.close()

spec = np.zeros([N, Nranges], dtype=np.complex64)

if dec > 1:
    z = stuffr.decimate(zin, dec=dec)
else:
    z = zin
z.shape = (N, clen)
if rfi_rem:
    bg = np.median(z, axis=0)
    z = z - bg
#    print(len(z))
if fft_filter:
    S = np.zeros(clen, dtype=np.float32)
    for i in np.arange(N):
        S += np.abs(sf.fft(z[i, :])) ** 2.0
    S = np.sqrt(S / float(N))
for i in np.arange(N):
    #        z = stuffr.decimate(zin[(i*clen*dec):((i+1)*clen*dec)],dec=dec)
    # B=(A^H A)^{-1}A^H
    # B*z = (A^H A)^{-1}A^H*z = x_ml
    # z = measurement
    # res[i,:] = backscattered echo complex amplitude
    if fft_filter:
        zw = np.array(sf.ifft(sf.fft(z[i, :]) / S), dtype=np.complex64)
    else:
        zw = z[i, :]
    res[i, :] = np.dot(B, zw)
#        res[i,:] = reso[i,:]#/np.median(np.abs(reso[i,:]))
if gc_rem:
    for i in range(gc, Nranges):
        res[:, i] = res[:, i] - np.median(res[:, i])

window = wfun(N)
for i in np.arange(Nranges):
    spec[:, i] = np.fft.fftshift(np.fft.fft(
        window * res[:, i]
    ))

if spec_rfi_rem:
    median_spec = np.zeros(N, dtype=np.float32)
    for i in np.arange(N):
        median_spec[i] = np.median(np.abs(spec[i, :]))
```

```
    for i in np.arange(Nranges):
        spec[:, i] = spec[:, i] / median_spec[:]
ret = {}
ret['res'] = res
ret['spec'] = spec
return (ret)
```

Bibliography

Akasofu, S.-I.
1964. The development of the auroral substorm. *Planetary and Space Science*, 12(4):273–282.

Akasofu, S.-I.
1981. Energy coupling between the solar wind and the magnetosphere. *Space Science Reviews*, 28(2):121–190.

Akasofu, S.-I.
2013. The relationship between the magnetosphere and magnetospheric/auroral substorms. In *Annales Geophysicae*, volume 31, P. 387. Copernicus GmbH.

Appleton, E. V.
1932. Wireless studies of the ionosphere. *Institution of Electrical Engineers-Proceedings of the Wireless Section of the Institution*, 7(21):257–265.

Appleton, E. V. and M. Barnett
1925. On some direct evidence for downward atmospheric reflection of electric rays. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 109(752):621–641.

Baddeley, L.
2019. Unis agf223 course compendium.

Baddeley, L., T. Yeoman, and D. Wright
2005. Hf doppler sounder measurements of the ionospheric signatures of small scale ulf waves.

Bailey, D.
1964. Polar-cap absorption. *Planetary and Space science*, 12(5):495–541.

Baker, D. et al.
2013. *Solar and space physics: A science for a technological society*. National

Academies Press.

Baumjohann, W. and R. A. Treumann
1997. *Basic space plasma physics*. World Scientific.

Birkeland, K.
1908. *The Norwegian aurora polaris expedition 1902-1903*, volume 1. H. Aschelhoug & Company.

Breit, G. and M. A. Tuve
1926. A test of the existence of the conducting layer. *Physical Review*, 28(3):554.

Brekke, A.
2012. *Physics of the upper polar atmosphere*. Springer Science & Business Media.

Bullett, T., N. Zabotin, R. Livingston, R. Grubb, J. Mabie, and C. Negrea
2014. Hf radar measurements of the ionosphere using dynasonde methods. In *2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*, Pp. 1–1. IEEE.

Chapman, S.
1931. The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth. *Proceedings of the Physical Society*, 43(1):26.

Chartier, A. T., J. Vierinen, and G. Jee
2020. First observations of the mcmurdo-south pole ionospheric hf channel. *Atmospheric Measurement Techniques Discussions*, 2020:1–11.

Chen, F. F.
1984. *Introduction to plasma physics and controlled fusion*, volume 1. Springer.

Cooley, J. W., P. A. Lewis, and P. D. Welch
1967. Historical notes on the fast fourier transform. *Proceedings of the IEEE*, 55(10):1675–1677.

COSPAR and URSI
2019. International reference ionosphere. Available at `http://irimodel.org/`.

Cowley, S. and M. Lockwood
1992. Excitation and decay of solar wind-driven flows in the magnetosphere-

ionosphere system. In *Annales geophysicae*, volume 10, Pp. 103–115.

Crowley, G. and F. Rodrigues
2012. Characteristics of traveling ionospheric disturbances observed by the tiddbit sounder. *Radio Science*, 47(04):1–12.

Davies, K.
1990. *Ionospheric radio*. IET.

Foster, J. C.
2005. Dasi: Distributed arrays of small instruments for space weather research. In *2nd Symposium on Space Weather*.

Hartinger, M., M. B. Moldwin, S. Zou, J. W. Bonnell, and V. Angelopoulos
2015. Ulf wave electromagnetic energy flux into the ionosphere: Joule heating implications. *Journal of Geophysical Research: Space Physics*, 120(1):494–510.

Hulburt, E.
1974. Early theory of the ionosphere. *Journal of Atmospheric and Terrestrial Physics*, 36(12):2137–2140.

Hunsucker, R.
1990. Atmospheric gravity waves and traveling ionospheric disturbances: Thirty years of research.

Hysell, D., M. Milla, and J. Vierinen
2016. A multistatic hf beacon network for ionospheric specification in the peruvian sector. *Radio Science*, 51(5):392–401.

IAGA
2019. International geomagnetic reference field. Available at `https://www.ngdc.noaa.gov/IAGA/vmod/igrf.html`.

Kaipio, J. and E. Somersalo
2006. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media.

Kirkwood, S. and H. Nilsson
2000. High-latitude sporadic-e and other thin layers–the role of magnetospheric electric fields. *Space Science Reviews*, 91(3-4):579–613.

Lanchester, B., T. Nygren, A. Huuskonen, T. Turunen, and M. J. Jarvis
1991. Sporadic-e as a tracer for atmospheric waves. *Planetary and space science*, 39(10):1421–1434.

Liu, J., F. Berkey, and S. Wu
1992. A study of the true height analysis methods. *Terrestrial, Atmospheric and Oceanic Sciences*, 3(2):129–146.

Lott, O., W. Stark, and M. Bail
2006. Multi-location long-term hf noise measurements and comparison to itu-r p. 372-8.

Milan, S., M. Lester, and N. Sato
2003. Multi-frequency observations of e-region hf radar aurora. *Annales Geophysicae*, 21.

Mini-circuits
2020a. Msp2ta-18-12+ datasheet. Available at `https://www.minicircuits.com/pdfs/MSP2TA-18-12+.pdf`.

Mini-circuits
2020b. Zx60-100vh+ datasheet. Available at `https://www.minicircuits.com/pdfs/ZX60-100VH+.pdf`.

Nygrén, T., L. Jalonen, A. Huuskonen, and T. Turunen
1984. Density profiles of sporadic e-layers containing two metal ion species. *Journal of atmospheric and terrestrial physics*, 46(10):885–893.

Nygren, T., L. Jalonen, J. Oksman, and T. Turunen
1984. The role of electric field and neutral wind direction in the formation of sporadic e-layers. *Journal of atmospheric and terrestrial physics*, 46(4):373–381.

on Distributed Arrays of Small Instruments for Research, C., M. in Solar-Terrestrial Physics, N. R. Council, et al.
2006. *Distributed Arrays of Small Instruments for Solar-Terrestrial Research: Report of a Workshop*. National Academies Press.

Perkins, F.
1973. Spread f and ionospheric currents. *Journal of Geophysical Research*, 78(1):218–226.

Perreault, P. and S. Akasofu
1978. A study of geomagnetic storms. *Geophysical Journal International*, 54(3):547–573.

Pilipenko, V.
1990. Ulf waves on the ground and in space. *Journal of atmospheric and*

terrestrial physics, 52(12):1193–1209.

radioworld.co.uk
2020. Diamond wd330 datasheet. Available at https://www.radioworld.co.uk/generate_product_pdf?product_id=1700135.

Reinisch, B.
2009. Digisonde 4d technical manual (version 1.0). *Lowell Digisonde International*.

Reinisch, B. W.
1986. New techniques in ground-based ionospheric sounding and studies. *Radio Science*, 21(3):331–341.

Reinisch, B. W., I. A. Galkin, G. Khmyrov, A. Kozlov, K. Bibl, I. Lisysyan, G. Cheney, X. Huang, D. Kitrosser, V. Paznukhov, et al.
2009. New digisonde for research and monitoring applications. *Radio Science*, 44(1).

Turunen, T., T. Nygren, and A. Huuskonen
1993. Nocturnal high-latitude e-region in winter during extremely quiet conditions. *Journal of atmospheric and terrestrial physics*, 55(4-5):783–795.

Verhulst, T., D. Altadill, J. Mielich, B. Reinisch, I. Galkin, A. Mouzakis, A. Belehaki, D. Burešová, S. Stankov, E. Blanch, et al.
2017. Vertical and oblique hf sounding with a network of synchronised ionosondes. *Advances in Space Research*, 60(8):1644–1656.

Vierinen, J., J. L. Chau, N. Pfeffer, M. Clahsen, and G. Stober
2016. Coded continuous wave meteor radar. *Atmospheric Measurement Techniques*, 9(2):829–839.

Wang, C., J. Wang, R. E. Lopez, L. Zhang, B. Tang, T. Sun, and H. Li
2016. Effects of the interplanetary magnetic field on the location of the open-closed field line boundary. *Journal of Geophysical Research: Space Physics*, 121(7):6341–6352.

Wellbrook
2020a. Ala-1530 antenna factor. Available at https://www.wellbrook.uk.com/loopantennas/pdf/ALA1530AntennaFactorChart.pdf.

Wellbrook
2020b. Ala-1530 antenna overview. Available at https://www.wellbrook.uk.com/loopantennas/pdf/ALA1530.pdf.