

Mathias Fontell

Numerical Ray Tracing of Medium and High Frequency Radio Waves in the Terrestrial Ionosphere

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Otaniemi, Espoo

12th December, 2018

Supervisor:

Prof. Esa Kallio, Aalto University School of
Electrical Engineering

Thesis advisors:

PhD Kirsti Kauristie, Finnish Meteorological
Institute

Docent Heikki Vanhamäki, University of Oulu

Author: Mathias Fontell

Title: Numerical Ray Tracing of Medium and High Frequency Radio Waves in the Terrestrial Ionosphere

Date: 12/12/2018

Language: English

Number of pages: 10+123

Degree programme: Engineering physics

Supervisor: Prof. Esa Kallio

Advisors: PhD Kirsti Kauristie, Docent Heikki Vanhamäki

This thesis develops a numerical ray tracing program to study the propagation of medium and high frequency radio waves in the terrestrial ionosphere. Based on the cold plasma wave approximation and geometrical optics, a ray tracing program is developed in C++ to numerically solve the paths of radio waves transmitted as skywaves.

The thesis presents the key properties of the terrestrial ionosphere as a medium for radio waves. We show how the Appleton-Hartree dispersion relation can be used with the Hamiltonian ray equations to obtain a solution to the radio ray path by adaptive numerical integration. Depending on the polarization of the radiating antenna, we present methods for evaluating the signal attenuation by magnetoionic splitting, ground reflections, and D-layer absorption. Moreover, we present a Monte Carlo method for estimating the geometric attenuation of skywaves.

The performance of the developed ray tracing program is validated against theoretical and empirical results. We find that the computed predictions of the maximum usable frequency agree well with the predictions of commercial software, representing at most a difference of 8% between our ray tracer and the commercial program. Moreover, we ascertain that the developed ray tracer produces reversible ray paths, a property that is expected on the basis of first principles. Finally, the developed program is used to predict coverage maps of a high frequency receiver in orbit, demonstrating how ray tracing and ray reversibility can be utilized to deduce high frequency ground stations that are within reach of a satellite.

Keywords: Ray tracing, ionosphere, skywave radio

Författare: Mathias Fontell		
Titel: Numerisk strålföljning av medium- och högfrekvensradiovågor i jordens jonosfär		
Datum: 12/12/2018	Språk: Engelska	Sidantal: 10+123
Utbildningsprogram: Teknisk fysik		
Övervakare: Prof. Esa Kallio		
Handledare: FD Kirsti Kauristie, Docent Heikki Vanhamäki		
<p>I detta diplomarbete utvecklas ett program för numerisk strålföljning (ray tracing) som används för att studera utbredningen av medium- och högfrekventa radiovågor i jordens jonosfär. Approximationen av vågor i kalla plasmer samt geometrisk optik används som basis för att utveckla ett strålföljningsprogram i C++ som numeriskt beräknar rymdvågornas utbredning.</p> <p>Arbetet presenterar centrala egenskaper hos jordens jonosfär som ett medium för radiovågor. Det demonstreras hur Appleton-Hartree-dispersionsrelationen kan användas med de Hamiltonska strålekvationerna för att lösa radiostrålens bana med hjälp av adaptiv numerisk integrering. Beroende på sändande antennens polarisation presenteras metoder för att beräkna signaldämpningen vid magnetojonisk uppdelning och reflektion vid markytan, samt dämpningen på grund av absorption i jonosfärens D-skikt. Dessutom presenteras en Monte Carlo-metod för att uppskatta den dämpning som orsakas av rymdvågornas utbredning över långa sträckor från sändaren.</p> <p>Det utvecklade strålföljningsprogrammets prestanda valideras med både teoretiska och empiriska mått. Beräknade värden för den maximala användbara frekvensen stämmer väl överens med resultat som producerats av ett kommersiellt program, vilket bevisas av att de förutspådda värdena skiljer sig med högst 8% mellan det utvecklade och kommersiella programmet. Dessutom påvisas reversibilitet av de beräknade strålbana, vilket är en egenskap som förväntas på basis av fysikaliska grundprinciper. Slutligen används programmet för att förutspå räckvidden av en HF-mottagare som befinner sig i omloppsbana. Dessa resultat demonstrerar hur strålföljning och strålreversibilitet kan användas för att hitta HF-markstationer som är inom räckhåll av en satellit.</p>		
Nyckelord: Strålföljning, jonosfären, radio, rymdvågor		

Preface

On December 3 2018, the Suomi 100 satellite was successfully launched from Vandenberg Air Force Base in California. This small satellite, designed and built by students at Aalto University, is what originally got me into working with this thesis topic in the first place. In 2015, still a somewhat inexperienced but eager physics student, I approached Professor Esa Kallio to learn if there were any opportunities for me to get involved with the newly conceived Suomi 100 mission. Thanks to his initiative, I started working with the radio instrument payload and ray tracing in the summer of 2016. The topic really struck a chord with me, eventually leading me to pursue my master's thesis on numerical ray tracing in 2018. Of all the people involved with my thesis, I am without a doubt most indebted to Esa for his supervision during my time at the Aalto Space Physics research group. The interest and support you have expressed towards my work has been truly invaluable, and I am greatly thankful for your mentorship throughout all these years.

I would like to express my most sincere gratitude to my instructors, Kirsti Kauristie and Heikki Vanhamäki, for providing vital commentary on my thesis on several occasions. Together with Esa, your professional experience in ionosphere research and radio physics has helped me correct many mistakes and shortcomings I might have been completely oblivious to otherwise. Furthermore, I am most thankful to Markku Alho and Riku Järvinen for making my stay at the Space Physics group both enjoyable and productive. In addition to your continued assistance with computer-related wizardry, the cheerful attitude and sense of humour you displayed really made it a fun experience to work on this project. To the Suomi 100 satellite systems engineering team, Petri Koskimaa and Arno Alho in particular, you have my gratitude for enlightening me on the technical details and practical engineering problems related to the satellite and the payload. I would also like to thank Tuija Pulkkinen and the Academy of Finland (grant 267073) for the financial support with my thesis. Moreover, thanks are due to Johannes Norberg and Lasse Häkkinen (Finnish Meteorological Institute) respectively for assistance with EISCAT-related ionosphere data and Proplab Pro 3, as well as Antti Kero and Alexander Kozlovsky (Sodankylä Geophysical Observatory) for providing insight into the fundamentals of ionosonde operation. At the Aalto Department of Applied Physics, I would like to express my sincere gratitude to Mathias Groth and Taina Kurki-Suonio for their tutelage. Your enthusiasm and professionalism has truly inspired me, and I am fortunate to have received such amazing mentors. I would also like to thank Seppo Sipilä, whose outstanding master's thesis on ray tracing in tokamak reactors served as an inspiration for my own thesis. Moreover, thanks are due to Konsta Särkimäki for advise on numerical integration, stemming from Konsta's experience with orbit-following in tokamak plasmas.

Lastly, I want to extend my most heartfelt appreciation to my friends and family for all the support that kept me going throughout my years as a student. If you are reading this, I am sure you know who you are. Nina, Andreas, Sara and Mathias, my dearest colleagues and partners in crime, thank you for all the unforgettable memories. Maria, my dear friend, thank you for helping me pull me through when

the going got rough. To my sisters, Laura and Sara, and my parents, Maria and Pekka, words alone cannot express how much your support means to me. It has been a truly wonderful journey.

Otaniemi, December 12, 2018

Mathias Fontell

Contents

Abstract	ii
Abstract (in Swedish)	iii
Preface	iv
Contents	vi
Symbols and abbreviations	ix
1 Introduction	1
2 The terrestrial ionosphere	5
2.1 D, E and F layers of the ionosphere	5
2.2 Earth-centered Earth-fixed (ECEF) coordinate system	7
2.3 The geomagnetic field	9
2.4 Global ionosphere and atmosphere models	10
2.4.1 International Reference Ionosphere (IRI)	10
2.4.2 NRLMSISE-00 atmosphere model	11
2.5 Comparison of IRI output with empirical data	12
3 Cold plasma waves	14
3.1 Magnetoionic theory: approximations and restrictions	14
3.2 Macroscopic description of a cold magnetoplasma	15
3.3 Plane wave approximation	16
3.4 Conductivity tensor of a cold magnetoplasma	17
3.5 Dielectric tensor and the wave equation	18
3.6 Dispersion relations: the Booker quartic and Appleton-Hartree formula	20
3.7 Cutoffs, resonances and coupling points of the refractive index	22
3.8 Polarization of magnetoionic modes	24
3.8.1 Constitutive relations	24
3.8.2 Polarization formulae	24
3.8.3 Polarization coordinate system transformations	27
3.8.4 Polarization ellipses	28
3.9 Wave mode interchange	30
3.10 Energy transfer by EM waves	32
4 Ray tracing in the ionosphere	34
4.1 Group velocity of wave packets	34
4.2 Geometrical optics in slowly varying media	35
4.3 Ray equations	37
4.4 Choosing the Hamiltonian	38
4.4.1 Appleton-Hartree Hamiltonian	39
4.4.2 Booker quartic Hamiltonian	40
4.5 Adaptive numerical integration	40

4.6	Antenna radiation patterns	43
4.6.1	Isotropic antenna	45
4.6.2	Ideal dipole	46
4.6.3	Half-wave dipole	48
4.6.4	Ionosonde rhombic antenna	49
4.7	Magnetoionic splitting and ground reflections	50
4.7.1	The limiting region of the ionosphere	50
4.7.2	Magnetoionic splitting formulae	51
4.7.3	Signal loss and polarization shift at ground reflection	54
4.8	Signal attenuation	58
4.8.1	Attenuation by absorption	59
4.8.2	Photon mapping to estimate geometric attenuation	60
4.9	Reversibility and reciprocity of radio waves in the ionosphere	63
4.9.1	Physical reversibility of ray paths	64
4.9.2	Irreversibility caused by numerics and model approximations	65
4.9.3	Reciprocity of EM fields	66
5	Ray tracing program	70
5.1	Dependencies and software architecture	71
5.2	Inputs, outputs and options	73
5.3	Execution of ray tracing	77
5.4	Exception handling	81
5.5	Parallelization on the CPU	82
5.6	Benchmark performance	84
6	Ray tracing results	89
6.1	Code validation using Proplab Pro 3	89
6.1.1	Predicted ray paths	90
6.1.2	Predicted maximum usable frequencies	92
6.2	Reversibility analysis	94
6.3	Attenuation of characteristic modes by absorption	95
6.4	Coverage map prediction using photon mapping	100
7	Summary	103
	References	107
	Appendix	111
A	Derivatives of the Hamiltonian	111
A.1	Appleton-Hartree formula	112
A.1.1	With field and with collisions (AHWFWC)	113
A.1.2	With field and no collisions (AHWFNC)	114
A.1.3	No field and with collisions (AHNFWC)	115
A.1.4	No field and no collisions (AHNFNC)	115
A.2	Booker quartic	116

A.2.1	With field and with collisions (BQWFWC)	116
A.2.2	With field and no collisions (BQWFNC)	117
B	Derivatives of the geomagnetic dipole field	118
C	Ray tracer configuration file options	119

Symbols and abbreviations

Symbols

e	elementary charge $\approx 1.6022 \times 10^{-19}$ C
m_e	electron rest mass $\approx 9.1094 \times 10^{-31}$ kg
c	speed of light in vacuum = 299 792 458 m/s
ε_0	vacuum permittivity $\approx 8.854 \times 10^{-12}$ F/m
μ_0	vacuum permeability = $4\pi \times 10^{-7}$ Vs/(Am)
Z_0	wave impedance of free space = $\sqrt{\mu_0/\varepsilon_0} \approx 376.73$ Ω
k_B	Boltzmann constant $\approx 1.381 \times 10^{-23}$ J/K
\mathbf{E}	electric field [V/m]
\mathbf{B}	magnetic flux density [T]
\mathbf{D}	electric displacement field [Cm ⁻²]
\mathbf{H}	magnetic field [A/m]
ρ_f	free electric charge density [Cm ⁻³]
\mathbf{j}_f	free electric current density [Am ⁻²]
N_α	particle density of species α [1/m ³]
ω_p	Electron plasma frequency = $\sqrt{N_{e0}e^2/m_e\varepsilon_0}$ [radians/s]
Ω_c	Electron gyrofrequency = $e \mathbf{B}_0 /m_e$ [radians/s]
ν_{en}	Electron collision frequency with neutrals [Hz]
$\hat{\mathbf{u}}$	unit vector of vector \mathbf{u}
$ \mathbf{u} $	Euclidean norm (i.e. length) of vector \mathbf{u}
i	unit imaginary number $\sqrt{-1}$
\mathcal{Z}^*	complex conjugate of complex number \mathcal{Z}
$\text{Re}[\mathcal{Z}]$	real part of complex number \mathcal{Z}
$\text{Im}[\mathcal{Z}]$	imaginary part of complex number \mathcal{Z}
\mathcal{I}	identity matrix
ω	wave angular frequency [radians/s]
f	= $\omega/2\pi$, wave frequency [Hz]
\mathbf{k}	wave vector [1/m]
k	= $ \mathbf{k} $, modulus of wave vector [1/m]
n	= kc/ω , wave refractive index
μ	= $\text{Re}[n]$
χ	= $-\text{Im}[n]$

Operators

$\mathbf{A} \cdot \mathbf{B}$	scalar product of vectors \mathbf{A} and \mathbf{B}
$\mathbf{A} \times \mathbf{B}$	cross product of vectors \mathbf{A} and \mathbf{B}
∇	the nabla operator = $\hat{\mathbf{e}}_x \frac{\partial}{\partial x} + \hat{\mathbf{e}}_y \frac{\partial}{\partial y} + \hat{\mathbf{e}}_z \frac{\partial}{\partial z}$
$\nabla \cdot \mathbf{A}$	divergence of vector field \mathbf{A}
$\nabla \times \mathbf{A}$	curl of vector field \mathbf{A}
$\frac{df}{dx}$	total derivative of f with respect to variable x
\dot{f}	total time derivative of f
$\frac{\partial f}{\partial x}$	partial derivative of f with respect to variable x
$\frac{\partial f}{\partial \mathbf{r}}$	= $\hat{\mathbf{e}}_x \frac{\partial f}{\partial x} + \hat{\mathbf{e}}_y \frac{\partial f}{\partial y} + \hat{\mathbf{e}}_z \frac{\partial f}{\partial z}$, spatial gradient of f
$\frac{\partial f}{\partial \mathbf{k}}$	= $\hat{\mathbf{e}}_x \frac{\partial f}{\partial k_x} + \hat{\mathbf{e}}_y \frac{\partial f}{\partial k_y} + \hat{\mathbf{e}}_z \frac{\partial f}{\partial k_z}$, gradient of f in k-space

Abbreviations

AM	amplitude modulation
CPU	central processing unit
ECEF	Earth-centered and Earth-fixed
EM	electromagnetic
GPU	graphics processing unit
HF	high frequency radio band, i.e. 3 to 30 MHz
IRI	International Reference Ionosphere
I/O	input/output
LEO	low Earth orbit
MF	medium frequency radio band, i.e. 0.3 to 3 MHz
MUF	maximum usable frequency
NRLMSISE-00	US Naval Research Laboratory mass spectrometer and incoherent scatter radar global atmosphere model year 2000
NVIS	near vertical incidence skywave
ODE	ordinary differential equation
OTH	over the horizon
O-mode	ordinary mode
X-mode	extraordinary mode
UTC	Coordinated Universal Time
WKB approximation	Wentzel-Kramers-Brillouin approximation

1 Introduction

Excerpt from the Times of London, December 16, 1901 [8]:

TRANS-ATLANTIC MESSAGE

From our Correspondent, St. Johns, NF, Dec. 14;

Signor Marconi authorizes me to announce that he received on Wednesday and Thursday electrical signals at his experimental station here from the station at Poldhu, Cornwall, thus solving the problem of telegraphing across the Atlantic without wire. He has informed the governor, Sir Cavendish Boyle, requesting him to apprise the British Cabinet of the discovery, the importance of which it is impossible to overvalue.

Prior to Guglielmo Marconi's historical experiments in 1901 [8], the ability of wirelessly relaying messages across the vast Atlantic ocean would have certainly been perceived as something out of science fiction. How is it possible, that electrical signals at the terminals of one transmitter can be reproduced so far beyond the horizon without any apparent medium to convey the transmission without line of sight? Little did Marconi know, that in a few decades his experiments would lead to the discovery of the responsible medium – namely, the ionosphere, an electrically ionized region in the upper atmosphere capable of refracting radio waves over the horizon. Fast forward by a century to the year 2018. In this day and age, many of the former mysteries of the ionosphere have been solved. Already for more than half a century now, the ionosphere has been utilized for both civilian and military purposes for over-the-horizon radio communication in a form of propagation known as skywave radio.

Miniaturization of satellites provides novel opportunities for studying the ionosphere and radio physics from orbit. In late 2018, Aalto University is set to launch the Suomi 100 centenary nanosatellite to celebrate a hundred years of Finnish independence. The Suomi 100 satellite conforms to the 1U CubeSat standard [37]: it is only $10 \times 10 \times 10$ cm in size and 1.1 kg in mass. This particular satellite is to be launched to a polar circular LEO orbit of 575 kilometers, and is fitted with a visible-spectrum camera and an AM radio receiver. The payload instrument of interest for this thesis is the AM radio receiver, which has been designed at Aalto University and is specifically tuned to MF/HF frequencies in the range of 1-10 megahertz. These frequencies are in the typical range of the ionospheric plasma oscillations. The instrument is thus designed to receive a myriad of MF/HF signals of artificial and natural origin alike, making the Suomi 100 satellite a miniaturized spaceborne laboratory for studying ionospheric radio physics. The MF/HF antennas are electrically small ferrite rods with radiation patterns akin to that of an ideal dipole [27].

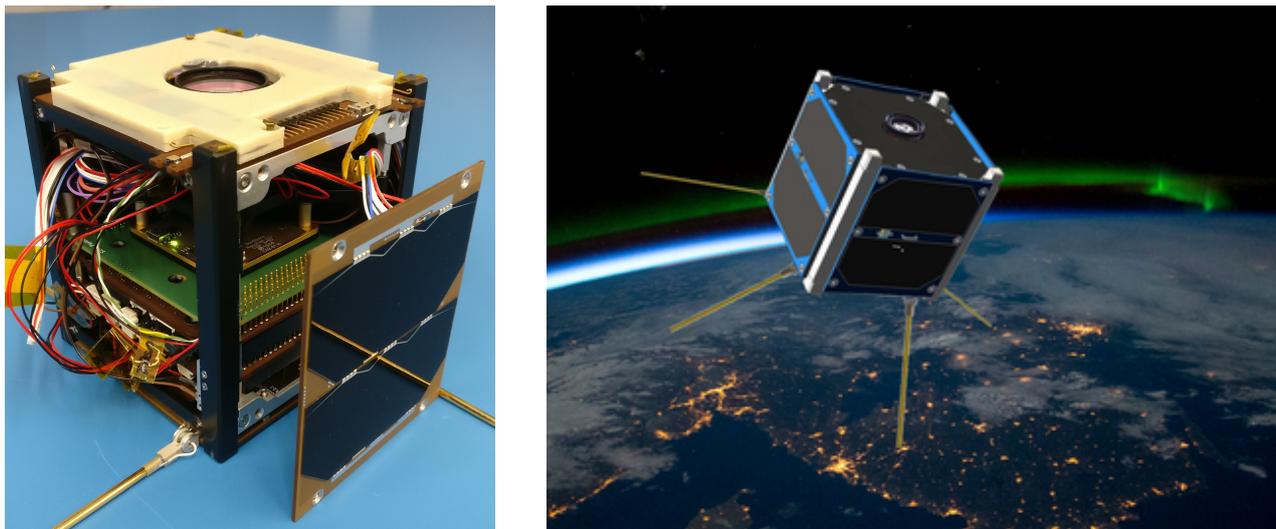


Figure 1: The Suomi 100 nanosatellite. The image on the left depicts the partially assembled flight model of the satellite. The two embedded MF/HF ferrite rods are visible on the beige top face as darkened depressions. The image on the right depicts an artistic rendering of the satellite in orbit. Courtesy of Aalto University.

This thesis is concerned with the development of software to be used in data analysis of the Suomi 100 MF/HF measurements. For such an endeavor, the software is to be fitted with a comprehensive ability to simulate the propagation of electromagnetic waves in the ionosphere. In over-the-horizon radio systems, signals of less than a kilometer in wavelength typically propagate hundreds to thousands of kilometers. A full three-dimensional electromagnetic solution of such a system would come at a staggering cost in terms of both memory and processing power; Typical finite element or finite difference time-domain methods (see e.g. [35]) require for accuracy that the grid cell size be only a fraction of the wavelength (which requires memory), while numerical stability requires that the time step is decreased proportionally when the cells get smaller (which requires processing power). Although it would be an overstatement to claim that such a solution is impossible, especially with modern improvements in computing power, certain approximations of EM wave propagation are still practically viable options for capturing the essential phenomena of interest. Arguably the most common and widely used approximation in ionospheric radio physics is that of geometrical optics, in which rays of short wavelengths are traced by sequential refractions in a predefined medium to visualize the travel paths of waves transmitted in certain directions. This method of computational wave propagation is known as ray tracing.

Ultimately, this thesis seeks to develop a complete ionospheric ray tracing program. The findings of the thesis are divided into six core sections:

- Section 2 is concerned with identifying the primary properties of the ionosphere that are indispensable for the development of a comprehensive ray tracer. We formulate analytic expressions for evaluating the plasma frequency, gyrofrequency and electron-neutral collision frequency in the ionosphere from meas-

urable quantities. We also present a global Cartesian coordinate system and vector basis, in which all vector quantities of the ray tracer can be expressed unambiguously. A large emphasis is placed on procedurally generating realistic ionosphere profiles from open-source software packages, which allows testing of the ray tracer in a representative medium.

- Section 3 presents the necessary magnetoionic theory of EM wave propagation in a cold magnetoplasma (such as the ionosphere). In this section, we derive the formulae for the refractive index and wave polarization from the cold plasma fluid equations and Maxwell's equations. We will see, that the geomagnetic field introduces anisotropy that manifests itself as two possible magnetoionic modes known as ordinary (O) and extraordinary (X) waves. In addition, we present problematic points of the refractive index, namely cutoffs, resonances and coupling points, which necessitate additional diagnostics in the ray tracer to detect these exceptions.
- Section 4 is the most elaborate section of this thesis, that describes the mathematical machinery needed to implement a ray tracer capable of simulating the key characteristics of skywave propagation. Here we derive the Hamiltonian ray equations, and present two Hamiltonian functions suitable for tracing MF/HF waves in the ionosphere. Moreover, we document some common antenna types and their radiation patterns, and describe how the resulting polarization affects skywave propagation by magnetoionic splitting and ground reflections. Furthermore, methods for estimating signal losses from D-layer absorption and free-space path loss are formulated. Finally, a suitable adaptive numerical integrator is described. Some attention is given to contemplating the effects of reversibility and nonreciprocity of skywaves, and how a ray tracer may reproduce these effects.
- Section 5 documents the software engineering aspects of the developed ray tracer. We describe the software architecture, dependencies, and required inputs, as well as the possible outputs of the traced ray paths. It is imperative to elaborate on how certain exceptions are handled, and for what kind of inputs such problems are most likely encountered. We describe the used methods of hardware acceleration on the CPU, as well as the achieved benchmark performance of some representative ray tracing sessions.
- Section 6 is then dedicated to presenting the results of ray tracing using the developed software. We delve into some methods of code validation using commercial software, and explore how the output differs when using a procedurally generated ionosphere instead of an empirically-measured one. We also attempt to justify some of the design choices by proving their efficiency in practice, such as the achieved path reversibility. Finally, we use the ray tracer to predict a coverage map of an MF/HF transmitter using photon mapping.
- Finally, section 7 summarizes the results obtained in this thesis. Based on the results of code validation in section 6, we discuss the possibilities for future

improvement of the developed ray tracer.

2 The terrestrial ionosphere

At an altitude of more than approximately 60 kilometers above the ground, the influx of energetic solar photons and particles of both near-Earth space and cosmic origin ionize the neutral terrestrial atmosphere to generate densities of free charges in excess of 10^8 m^{-3} . The resulting region of the upper atmosphere is known as the ionosphere, and is of vital practical importance for over-the-horizon radio propagation. It is fair to say that the study of the ionosphere went historically hand in hand with the study of radio physics; after Marconi's first successful transatlantic message in 1901, the ionosphere was probed for more than two decades by use of radio waves before Appleton's conclusive proof of its existence in 1927 [2].

As stated before, the primary goal of this thesis is to develop a computational tool to model MF/HF radio propagation in the ionosphere. As such, this section characterizes the quantities and properties of the ionosphere most vital to this effort. In order to properly model radio propagation in the ionosphere, the knowledge of (at least) three ionospheric quantities is required:

- The number density of free electrons $N_e(\mathbf{r}, t)$, which ultimately determines the electron plasma frequency of the medium.
- The geomagnetic field $\mathbf{B}_0(\mathbf{r}, t)$, which in turn determines the gyrofrequency of the medium. The intensity and orientation of the field turns out to determine the anisotropic properties of the ionosphere.
- The effective collision frequency of electrons $\nu_{eff}(\mathbf{r}, t)$, which determines the absorption per path length of a passing radio wave. In practice, ν_{eff} will represent the collision frequency of electrons with neutral constituents.

The quantities above are denoted as being functions of both position and time, reminding us that the ionosphere demonstrates an intricate substructure as well as temporal variation. However, the computational tool developed for this thesis will treat the ionosphere as being static and only varying with altitude within a sufficiently small geographical region.

2.1 D, E and F layers of the ionosphere

Ionosondes, suborbital rockets, incoherent scatter radars and in situ satellite measurements reveal that the ionosphere exhibits an intricate substructure related to different chemical processes and spectra of solar radiation. An illustration of a typical vertical profile is presented in figure 2. The consequences of myriad chemical and physical processes divide the daytime ionosphere into four distinct main regions: The D region (roughly 60 to 90 km altitude), the E region (105-160 km), the F1 region (160-180 km) and the F2 region (roughly 180 km and above). The ionosphere exhibits significant diurnal behaviour: At night, the absence of ionizing solar radiation causes the F1 and D regions to disappear, while the ionization degree of the remaining E and F layers is greatly reduced. The F2 region, consisting of mainly O^+ -ions and formed by vertical ion drift processes, is the region that harbors the

maximum electron density at an altitude of roughly 300 km [8]. The plasma frequency corresponding to the temporary maximum electron density $N_{e,max}$ is known as the F2 (or simply F) critical frequency, and is given by the expression

$$f_{p,max} = \frac{1}{2\pi} \sqrt{\frac{N_{e,max}e^2}{\epsilon_0 m_e}}. \quad (2.1)$$

The F critical frequency is typically in the order of a few MHz, lying in the MF/HF frequency range.

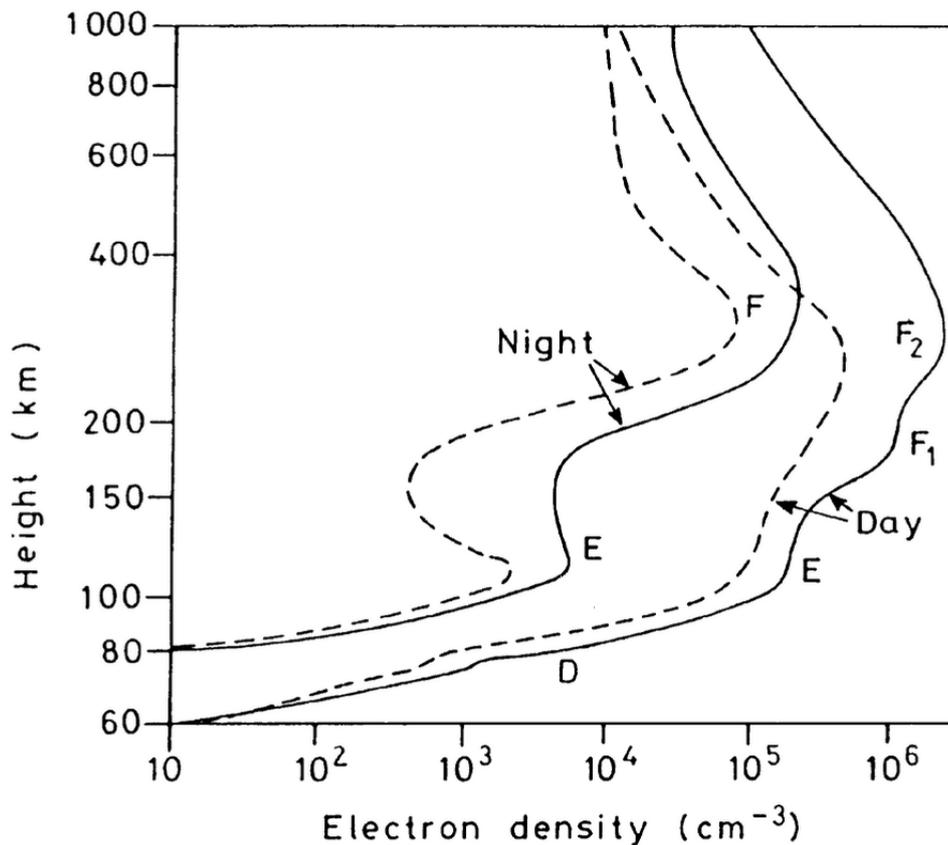


Figure 2: Typical vertical electron density profiles of the ionosphere, illustrating the main layers. The solid lines represent profiles at maximum solar activity, while the dashed lines are during minimum solar activity. [8]

Several centuries of telescopic observations indicate, that the number of spots on the Sun varies in an oscillatory fashion with a period of roughly 11 years [10]. Subsequently, the solar activity varies with the same period, which has been well observed to cause a similar variation in the structure of the ionosphere. During solar maximum, the F2 maximum density is typically greater by roughly a magnitude than during solar minimum. Apart from annual variation, the ionosphere exhibits seasonal variation as well. However, contrary to intuition, the noon F critical frequency is usually greater in winter than in summer – a phenomenon known as the seasonal anomaly. [8]

The D region accounts for the lowest part of the ionosphere, and is formed by the most penetrating radiation of the extreme ultraviolet spectrum and the Lyman- α line [8]. The density of neutral constituents is high in the D region, causing free electrons generated by ionization to collide frequently with mainly molecular oxygen and nitrogen. The resulting high value of the electron-neutral collision frequency ν_{en} accounts for the significant attenuation of low-frequency radio waves passing through the D region. In [16], semiempirical formulae for computing the electron-nitrogen and electron-oxygen collision frequencies are presented as

$$\nu_{N_2} = 2.33 \times 10^{-17} N_{N_2} (1 - 1.21 \times 10^{-4} T_e) T_e \text{ [Hz]}, \quad (2.2)$$

$$\nu_{O_2} = 1.82 \times 10^{-16} N_{O_2} (1 + 3.6 \times 10^{-2} \sqrt{T_e}) \sqrt{T_e} \text{ [Hz]}. \quad (2.3)$$

In equations (2.2)-(2.3), the electron temperatures T_e are given in Kelvin, and the neutral particle densities in m^{-3} . In this thesis, the effective collision frequency of electrons in all regions of the ionosphere is evaluated as the weighted average of ν_{N_2} and ν_{O_2} :

$$\nu_{eff} = \nu_{en} = \frac{N_{N_2} \nu_{N_2} + N_{O_2} \nu_{O_2}}{N_{N_2} + N_{O_2}}. \quad (2.4)$$

One is to keep in mind, that equations (2.2)-(2.4) for evaluating ν_{eff} are not to be extended outside the D region without caution. At higher altitudes, the collision processes between other neutral and ion species may become significant, and the semiempirical expressions (2.2)-(2.3) are no longer valid. In case that ν_{N_2} (2.2) becomes negative for high values of T_e , we set the collision frequency ν_{N_2} to zero.

From time to time, the E and F regions demonstrate peculiar anomalies that manifest as electron densities much higher than usual for the layer. Such localized anomalies in the E region are termed sporadic-E, and are characterized by layers of elevated electron density only a few kilometers across in altitude. The cause of frequent sporadic-E at low latitudes (within $20^\circ - 30^\circ$ from the equator) is usually attributed to instabilities in the equatorial electrojet [8], while high-latitude sporadic-E is typically attributed to large-scale convective electric fields [29]. In addition, auroral electron precipitation is known to cause rapid and spatially extensive variations in the E and F regions. The elevated electron densities of anomalous E and F regions show as plasma frequencies above 5 or even 10 MHz, temporarily allowing VHF waves to propagate over the horizon when they would usually just penetrate through the ionosphere. [8]

2.2 Earth-centered Earth-fixed (ECEF) coordinate system

For future reference, we set to define an orthonormal basis of vectors that is Earth-centered and Earth-fixed. By doing so, all the vector quantities used by the developed ray tracer can be expressed in some global system of coordinates, greatly facilitating transformations from one system to another.

For the remainder of this thesis, the global system is defined as follows:

- The origin is placed at the center of the Earth
- The positive x-axis is taken to point from the origin to the equator on the Greenwich meridian (0°N , 0°E)
- The positive y-axis is taken to point from the origin to the equator on the 90°E meridian (0°N , 90°E)
- The positive z-axis is taken to point from the origin to the geographic North pole (90°N , 0°E)

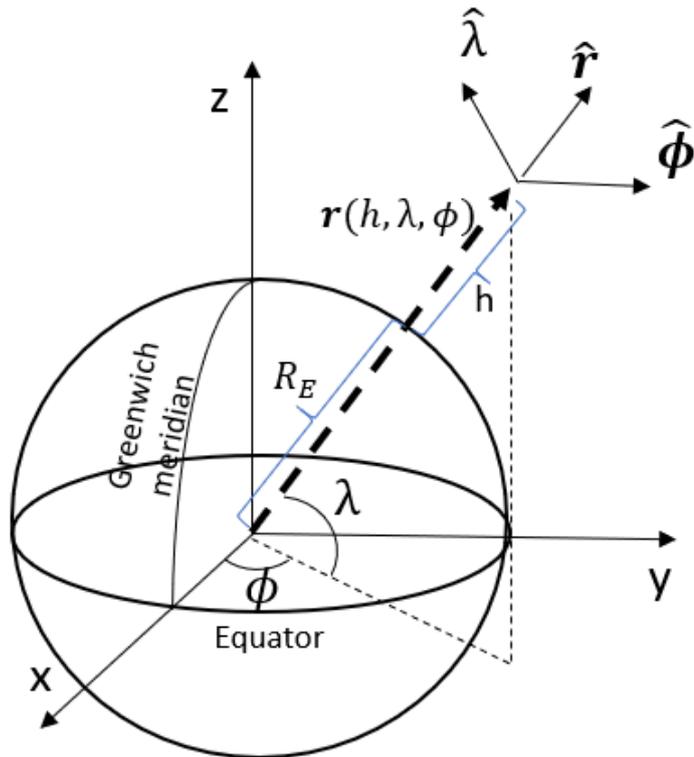


Figure 3: The global Earth-centered Earth-fixed (ECEF) coordinate system. The local orthonormal unit vector $\hat{\mathbf{r}}$ is (approximately) normal to the Earth's surface, while $\hat{\boldsymbol{\lambda}}$ is pointing northward and $\hat{\boldsymbol{\phi}}$ eastward.

This particular choice of coordinate axes is presented in figure 3. Given the geographic latitude λ , longitude ϕ and the altitude h above the ground, the position is given in the above global coordinate system by the expression

$$\mathbf{r}(h, \lambda, \phi) = x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z \equiv \begin{bmatrix} x \\ y \\ z \end{bmatrix} = (R_E + h) \begin{bmatrix} \cos \lambda \cos \phi \\ \cos \lambda \sin \phi \\ \sin \lambda \end{bmatrix}. \quad (2.5)$$

In equation (2.5), R_E denotes the mean radius of the Earth, taken to be 6371.0 km. Conversely, the spherical coordinates can be reproduced from known Cartesian

coordinates by

$$h = r - R_E = \sqrt{x^2 + y^2 + z^2} - R_E, \quad (2.6)$$

$$\lambda = \arcsin\left(\frac{z}{r}\right), \quad (2.7)$$

$$\phi = \text{atan2}(y, x). \quad (2.8)$$

Here $\text{atan2}(y, x)$ denotes the 2-argument arctangent function that is capable of identifying the correct quadrant of a point (x, y) .

The local orthonormal unit vectors $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\phi}}\}$ are particularly useful for expressing the elevation and azimuth directions of radio transmissions. The unit vector $\hat{\mathbf{r}}$ is (approximately) normal to the Earth's surface, while $\hat{\boldsymbol{\lambda}}$ is pointing northward and $\hat{\boldsymbol{\phi}}$ eastward. These vectors are given by the simple expressions

$$\hat{\mathbf{r}} = \frac{\mathbf{r}}{r} = \begin{bmatrix} \cos \lambda \cos \phi \\ \cos \lambda \sin \phi \\ \sin \lambda \end{bmatrix}, \quad (2.9)$$

$$\hat{\boldsymbol{\lambda}} = \begin{bmatrix} -\sin \lambda \cos \phi \\ -\sin \lambda \sin \phi \\ \cos \lambda \end{bmatrix}, \quad (2.10)$$

$$\hat{\boldsymbol{\phi}} = \begin{bmatrix} -\sin \phi \\ \cos \phi \\ 0 \end{bmatrix}. \quad (2.11)$$

2.3 The geomagnetic field

The magnetosphere of the Earth is of an exceedingly complex and dynamic nature, being the result of interaction between the Earth's intrinsic magnetic field and the solar wind. Fortunately, the sum magnetic field near the surface of the Earth is approximately that of the intrinsic geomagnetic dipole field [10]

$$\mathbf{B}_0(\mathbf{r}) = B_E \left(\frac{R_E}{r}\right)^3 [3(\hat{\mathbf{m}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \hat{\mathbf{m}}]. \quad (2.12)$$

Equation (2.12) provides a simple analytical model for evaluating the geomagnetic field anywhere near the surface of the Earth, and will be used in this form for ray tracing through the ionosphere in this thesis. The value B_E denotes the strength of the geomagnetic field at the surface of the Earth ($r = R_E$) on the geomagnetic equator where $\hat{\mathbf{m}} \cdot \hat{\mathbf{r}} = 0$. The precise value of B_E demonstrates slight secular variation; a representative value of $B_E = 30.4 \mu\text{T}$ is given in [10] for the year 1986. Furthermore, the vector $\hat{\mathbf{m}}$ is the magnetic dipole moment unit vector. Assuming that the dipole moment responsible for the intrinsic geomagnetic field is placed at the center of the Earth, the vector $\hat{\mathbf{m}}$ then points towards the South geomagnetic pole. The NOAA 2015 World Magnetic Model [41] gives the geographic coordinates of the South geomagnetic pole as $(80.31^\circ\text{S}, 107.38^\circ\text{E})$. This allows us to evaluate the

magnetic dipole orientation in the global coordinate system using eqn. (2.9):

$$\hat{\mathbf{m}} = \begin{bmatrix} \cos \lambda \cos \phi \\ \cos \lambda \sin \phi \\ \sin \lambda \end{bmatrix} = \begin{bmatrix} \cos(-80.31^\circ) \cos(107.38^\circ) \\ \cos(-80.31^\circ) \sin(107.38^\circ) \\ \sin(-80.31^\circ) \end{bmatrix} \approx \begin{bmatrix} -0.05028 \\ 0.16063 \\ -0.98573 \end{bmatrix}. \quad (2.13)$$

For the purpose of modeling radio propagation in the ionosphere, the spatial derivatives of eqn. (2.12) need to be evaluated as well. Analytic expressions for these derivatives with respect to Cartesian coordinates are given in appendix B. Moreover, the developed ray tracing program allows for the option of a constant geomagnetic field. Provided the coordinates (λ, ϕ) for determining $\hat{\mathbf{m}}$ and the magnitude B_E , the orientation and magnitude of the constant geomagnetic field is given by the expression

$$\mathbf{B}_0(\mathbf{r}) = \mathbf{B}_0 = B_E \hat{\mathbf{m}}. \quad (2.14)$$

2.4 Global ionosphere and atmosphere models

If we wish to computationally model the propagation of radio waves, we are going to need a background ionosphere in which to do so. Since Marconi's first successful transatlantic message in 1901, the ionosphere has been probed empirically from top to bottom to ever increasing accuracy, ushering the development of more and more precise empirical models of the ionosphere and atmosphere. Since some of these models are published as open-source code, they provide a feasible means of testing our ray tracing software in a realistic ionosphere without the excessive labor of manually generating ionospheric profiles. Moreover, procedurally generated realistic ionospheric profiles provide exciting opportunities for analyzing radio propagation from ground stations to satellites or other ground stations as a network design tool. In this thesis, we present two such common models: the International Reference Ionosphere (IRI), and the NRLMSISE-00 atmosphere model.

2.4.1 International Reference Ionosphere (IRI)

The International Reference Ionosphere, or IRI for short, is an ongoing international project by the Committee on Space Research (COSPAR) and the International Union of Radio Science (URSI) [40]. The project was established in the late sixties in an effort to conceive a useful empirical model of the ionosphere. The latest release of IRI was published in 2016, with source code available in Fortran 77. IRI allows the generation of vertical electron density, main ion density and electron temperature profiles over a range of geographic coordinates and time of day and year, making it particularly useful for ray tracing. Moreover, IRI is fitted with tabled measurements of solar activity and geomagnetic indices from the years 1957 to 2017, allowing IRI to demonstrate some annual variation in its output.

An example of vertical profiles generated by IRI are presented in figure 4. Only two quantities are collected from the output: the electron density $N_e(h)$, and the electron temperature $T_e(h)$ as a function of altitude h . The former is needed for

reproducing the ionospheric layers and plasma frequencies, while the latter is needed later for evaluating the collision frequency ν_{en} (eqn. (2.4)).

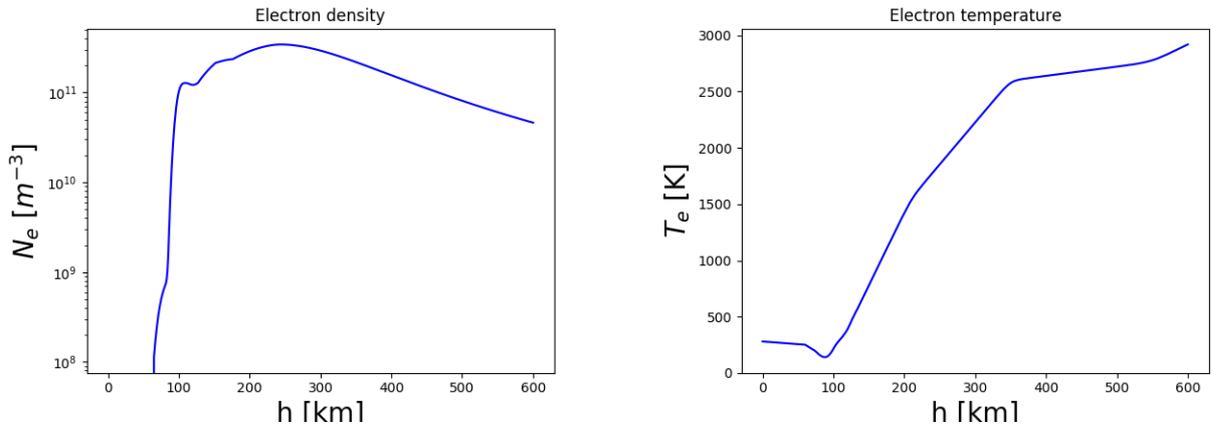


Figure 4: Standard output of IRI version 2016. The profiles are generated as vertical profiles above Espoo (60.1°N, 24.8°E) for the 15th of June 2016 at 12:00 local time. The profiles are generated from the minimal altitude 65 km using steps of 1 km.

2.4.2 NRLMSISE-00 atmosphere model

NRLMSISE-00, is an atmosphere model developed by the US Naval Research Laboratory, and is based on the earlier MSISE90 model [39]. The need of such a model for our ray tracer was recognized from the necessity of evaluating wave absorption by electron collisions: The semiempirical expressions (2.2)-(2.3) for the collision frequency require knowledge of not only the electron temperature, but of the neutral densities as well. The electron temperature values are conveniently computed by using IRI, while the nitrogen and oxygen density values are produced from NRLMSISE-00. The model source code is available in both Fortran 77 and the C language – however, since the developed ray tracer is programmed in C++, the C version of NRLMSISE-00 was favored due to compatibility between C and C++.

Examples of vertical profiles generated by NRLMSISE-00 are shown in figure 5. The model itself generates two quantities useful to us, namely the molecular nitrogen and oxygen densities. In post-processing, these values are used together with the electron temperature output of IRI to estimate ν_{en} using eqns. (2.2)-(2.4). In addition to the electron temperature produced by IRI, the NRLMSISE-00 code is called using solar activity and geomagnetic index values fetched by IRI. This way, the profiles generated by both models are expected to represent the same ionosphere.

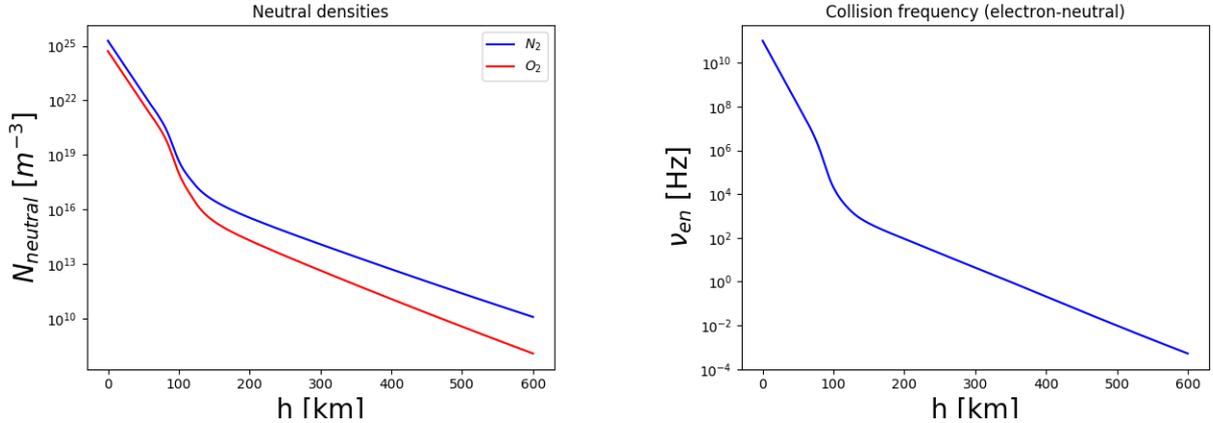


Figure 5: Standard output of the NRLMSISE-00 C-version. The profiles are generated as vertical profiles above Espoo (60.1°N, 24.8°E) for the 15th of June 2016 at 12:00 local time. The profiles are generated from the minimal altitude 65 km using steps of 1 km, using solar activity values and geomagnetic indices fetched by IRI. The electron-neutral collision frequency values are then computed using eqns. (2.2)-(2.4) to later generate an estimate of wave absorption.

2.5 Comparison of IRI output with empirical data

Of course, IRI is just a model, and as such, cannot be guaranteed to reproduce the actual ionosphere to perfect accuracy. In order to validate the performance of IRI for use with the developed ray tracer, it is prudent to compare its output with empirical observations of the ionosphere. For this purpose, we have elected to employ the EISCAT Dynasonde database [38]. EISCAT stands for European Incoherent Scatter Scientific Association, and is an international collaboration of research institutes to maintain a network of radar equipment for observing the ionosphere and its interactions with the thermosphere and magnetosphere. In particular, the association maintains a specialized HF ionosonde in Tromsø, Norway, titled the Tromsø dynasonde [31]. Measured ionograms as acquired by the dynasonde are freely available on the web, and thus appear to provide ample opportunity to benchmark the performance of IRI.

Figure 6 shows a comparison between IRI and the Tromsø dynasonde for some representative day in midsummer. The electron density output of IRI has been converted to a plasma frequency value using equation (2.1). For this particular sample, it appears that the profiles match remarkably well in the D and F1 regions, while the E and F2 region plasma frequencies demonstrate a discrepancy of up to 0.5 MHz.

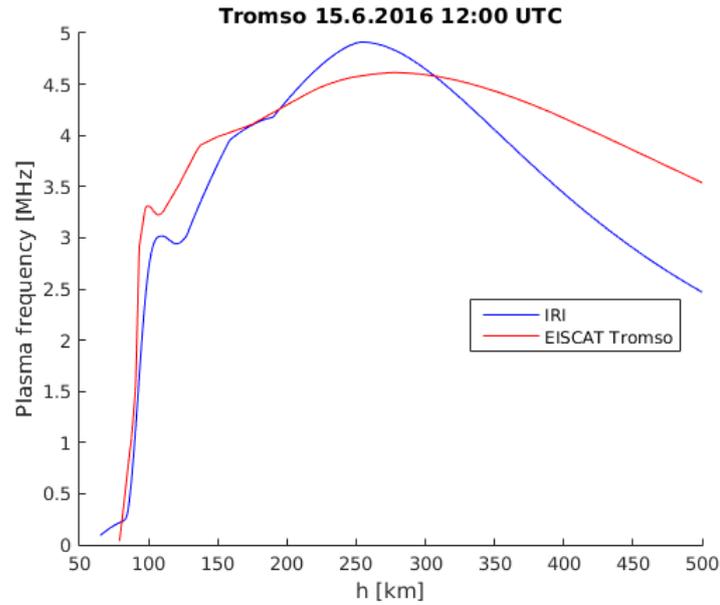


Figure 6: EISCAT dynasonde data [38] plotted against IRI output for the same time and location in Tromsø. The greatest differences are apparent in the E and F2 regions, where the predicted plasma frequencies differ by up to approximately 0.5 MHz.

Moreover, unpredictable anomalies in the ionosphere constitute another valid criticism against the use of IRI for ray tracing. An example of such an anomaly is the sporadic-E, which is known to affect the maximum usable frequency in practice by generating temporary clouds of high electron density. The possibilities of IRI to model the occurrence of sporadic-E are limited, but not unheard of [21]. In addition, IRI-2016 is equipped with F2 and auroral E storm models to produce some semblance of ionospheric anomalies [40], extending the utility of IRI for radio prediction.

3 Cold plasma waves

3.1 Magnetoionic theory: approximations and restrictions

The study of wave propagation in a cold plasma immersed in an external magnetic field is known as magnetoionic theory. The need for such a theoretical treatment was spearheaded by research in ionospheric physics, when physicists Kennelly and Heaviside proposed in 1902 that successful transatlantic radio communications by Guglielmo Marconi could be explained by radio waves being reflected from an ionized layer in the atmosphere [2]. In 1925, Englishmen Appleton and Hartree with Americans Nichols and Schelleng recognized the importance of the geomagnetic field on ionospheric wave propagation, which was followed by pioneering experimental work in probing the ionosphere to better understand its constitution and extent. In 1931, Hartree published his work on magnetoionic theory in collaboration with Appleton, in which the all-important Appleton-Hartree refractive index formula for an electron plasma was presented. The same result for a more general collisional plasma was independently derived by German radio physicist Lassen in 1927 [2][3].

The original treatment of Appleton, Hartree and Lassen included only the effects of plasma electrons on wave propagation. While it is perfectly possible to generalize the magnetoionic treatment with an inclusion of ion species, it has proved to be unnecessary for ionospheric radio unless very low frequencies (< 1 kHz) are used. The reason for this is that ions are several thousand times more massive than electrons, which equates to ion gyrofrequencies that are smaller than the corresponding electron gyrofrequency by at least three orders of magnitude. More importantly, the effects of ions can be neglected only if the radio wave frequency is much greater than the ion gyrofrequency [3], which holds demonstrably well in the terrestrial ionosphere: The ion gyrofrequency is at best around 1 kHz, while the MF/HF-bands span 0.3-30 MHz. In this thesis, we thus consider only the effects of electrons on the refractive index.

In the ionosphere, temperatures are typically of the order of a few hundred Kelvin in the D and E regions, and up to a few thousand Kelvin in the E and F regions. For us to use the cold plasma approximation and neglect (most) thermal effects, it is necessary that the wave phase velocity be much greater than the plasma thermal speed $\sqrt{2k_B T/m}$ [11]. This requirement is most notably violated at so-called resonances in magnetoplasmas, when the refractive index grows abruptly in magnitude and the phase speed approaches zero. However, one should keep in mind that not all thermal effects are neglected in the big picture; The electron collisions between other ionospheric constituents require knowledge of the nonzero electron temperature, and are of great practical importance when determining signal attenuation by absorption. In this thesis, we include the effects of electron collisions on wave propagation.

Moreover, common magnetoionic theory assumes that the external magnetic field (i.e. the geomagnetic field) does not vary appreciably over short distances when deriving the conductivity tensor of the cold plasma. In other words, magnetoionic theory places some requirements on the approximate homogeneity of the medium.

When deriving the ray equations, however, we have chosen not to approximate the magnetic field gradient as zero. Ray theory places additional restrictions separate to those of magnetoionic theory, and these will be discussed in more detail in section 4.

In this section, we set to establish the necessary theoretical framework needed to describe wave propagation, and ultimately, to perform ray tracing in magnetoplasmas. The fluid description of plasmas is used as a starting point to give the macroscopic equations, from which the conductivity- and dielectric tensors are derived under certain conditions. From here, the Appleton-Hartree refractive index is derived for an anisotropic and collisional plasma. Finally, the constitutive relations for electromagnetic fields are presented, which are shown to support only certain states of polarization in a magnetoplasma.

3.2 Macroscopic description of a cold magnetoplasma

Fluid descriptions of plasmas relate the macroscopic quantities (such as e.g. density, flow velocity and pressure) by sets of partial differential equations, and are formally acquired by taking velocity moments of the Boltzmann equation. A thorough treatment of this procedure is given by e.g. [2] and [11]. The zeroth and first moment give respectively the continuity and momentum equations, which are presented below for particle species α :

$$\frac{\partial N_\alpha}{\partial t} + \nabla \cdot (N_\alpha \mathbf{u}_\alpha) = 0, \quad (3.1)$$

$$\begin{aligned} N_\alpha m_\alpha \frac{\partial \mathbf{u}_\alpha}{\partial t} + N_\alpha m_\alpha \mathbf{u}_\alpha \cdot \nabla \mathbf{u}_\alpha - N_\alpha q_\alpha (\mathbf{E} + \mathbf{u}_\alpha \times \mathbf{B}) + \nabla \cdot \mathbf{P}_\alpha \\ = m_\alpha \int \mathbf{v} \left(\frac{\partial f_\alpha}{\partial t} \right)_c d^3 \mathbf{v}. \end{aligned} \quad (3.2)$$

Here \mathbf{u}_α denotes the flow velocity and \mathbf{P}_α the pressure tensor of species α at a given point. The momentum equation (3.2) now contains an unknown variable, namely the pressure tensor \mathbf{P}_α . The cold plasma approximation asserts that the temperature is too low to manifest significant pressure components so that $\nabla \cdot \mathbf{P}_\alpha = 0$.

The collision operator, i.e. the right hand side of eqn. (3.2), is generally a complicated function in velocity space, and practically necessitates some sort of approximation. A common approximation for the collision operator is the Krook model [11], which expresses the integral using estimated collision frequencies:

$$m_\alpha \int \mathbf{v} \left(\frac{\partial f_\alpha}{\partial t} \right)_c d^3 \mathbf{v} \approx - \sum_\beta m_\alpha N_\alpha (\mathbf{u}_\alpha - \mathbf{u}_\beta) \langle \nu_{\alpha\beta} \rangle, \quad (3.3)$$

where $\langle \nu_{\alpha\beta} \rangle$ is the average collision frequency between particle species α and β . In the ionosphere, plasma electrons collide predominantly with neutrals with an average collision frequency ν_{en} . Treating the neutrals as a stationary background,

the electron collision operator simplifies to

$$m_e \int \mathbf{v} \left(\frac{\partial f_e}{\partial t} \right)_c d^3\mathbf{v} \approx -m_e N_e \mathbf{u}_e \nu_{en}. \quad (3.4)$$

In the entirety of this thesis, we will assume all species except electrons to be immobile. The total electric current density then only consists of the electron current, and is given by

$$\mathbf{j} = \mathbf{j}_e = -e N_e \mathbf{u}_e. \quad (3.5)$$

No work on plasma physics would be complete without presenting Maxwell's equations. Plasmas, consisting of free charged particles in abundance, both interact with and act as sources for electromagnetic fields. The electromagnetic fields and their sources are related by Maxwell's equations, given by [11]

$$\nabla \cdot \mathbf{D} = \rho_f, \quad (3.6)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (3.7)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (3.8)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}. \quad (3.9)$$

In linear media, the fields \mathbf{E} and \mathbf{B} are related to \mathbf{D} and \mathbf{H} by the constitutive relations

$$\mathbf{D} = \varepsilon_0 \boldsymbol{\varepsilon}_r \cdot \mathbf{E}, \quad (3.10)$$

$$\mathbf{B} = \mu_0 \boldsymbol{\mu}_r \cdot \mathbf{H} \approx \mu_0 \mathbf{H}. \quad (3.11)$$

In eqn. (3.11), we have assumed the relative permeability to be close to unity, also another common approximation in magnetoionic theory [3]. However, the dielectric tensor $\boldsymbol{\varepsilon}_r$ is generally a tensor of the second order in a magnetoplasma, accounting for the anisotropic properties of the medium. In the following sections, we will derive the dielectric tensor of a cold electron magnetoplasma.

3.3 Plane wave approximation

Assume the electron density and magnetic field to respectively take time-invariant equilibrium values, which are perturbed slightly by a passing EM wave with fields \mathbf{E} and \mathbf{B}_1 . The electron density and magnetic field are decomposed into

$$N_e = N_{e0}(\mathbf{r}) + N_{e1}(\mathbf{r}, t), \quad (3.12)$$

$$\mathbf{B} = \mathbf{B}_0(\mathbf{r}) + \mathbf{B}_1(\mathbf{r}, t), \quad (3.13)$$

where N_{e0} and \mathbf{B}_0 denote in this thesis respectively the equilibrium electron density and the geomagnetic field in the ionosphere. Consider now small-amplitude perturbations of the components of N_{e1} , \mathbf{u}_e , \mathbf{B}_1 and \mathbf{E} of the harmonic form

$$g(\mathbf{r}, t) \propto e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}. \quad (3.14)$$

Equation (3.14) describes progressive plane waves of angular frequency ω that all propagate in the direction of the wave normal \mathbf{k} . The angular velocity is equivalently $\omega = 2\pi f$, where f is the frequency of the wave in Hertz. With the plane wave approximation, the partial derivatives of the first-order perturbations are simplified to the following operators:

$$\frac{\partial}{\partial t} \rightarrow -i\omega, \quad (3.15)$$

$$\nabla \cdot \rightarrow i\mathbf{k} \cdot, \quad (3.16)$$

$$\nabla \times \rightarrow i\mathbf{k} \times \quad (3.17)$$

Assuming that the harmonic perturbations are much smaller than the equilibrium values, e.g. $|\mathbf{B}_1| \ll |\mathbf{B}_0|$, the momentum equation (3.2) can be linearized by only including terms no higher than of the first order:

$$-i\omega N_{e0} m_e \mathbf{u}_e + N_{e0} e (\mathbf{E} + \mathbf{u}_e \times \mathbf{B}_0) = -m_e N_{e0} \mathbf{u}_e \nu_{en}. \quad (3.18)$$

In equation (3.18), the Krook model collision operator (3.4) has been used.

3.4 Conductivity tensor of a cold magnetoplasma

Multiplying eqn. (3.18) by e/m_e , and substituting the linearized total current $\mathbf{j} = -eN_{e0}\mathbf{u}_e$, one obtains the equation

$$(i\omega - \nu_{en})\mathbf{j} - \Omega_c \mathbf{j} \times \mathbf{b}_0 = -\omega_p^2 \varepsilon_0 \mathbf{E}, \quad (3.19)$$

where $\mathbf{b}_0 = \mathbf{B}_0/|\mathbf{B}_0|$. The symbols ω_p and Ω_c denote respectively the electron plasma frequency and gyrofrequency in radians per second, and are given by the expressions

$$\omega_p = \sqrt{\frac{N_{e0} e^2}{m_e \varepsilon_0}}, \quad (3.20)$$

$$\Omega_c = \frac{e |\mathbf{B}_0|}{m_e}. \quad (3.21)$$

It should be noted, that Ω_c is positive. Some authors (e.g. Stix, 1992 [18]) define the gyrofrequencies to be of the same sign as the charge of species α , which is a notation required for multicomponent plasmas to reflect the opposite directions of gyro-orbits.

Without loss of generality, the coordinate axes are chosen so that $\mathbf{b}_0 \parallel \hat{\mathbf{e}}_z$. In this coordinate system, the cross product in eqn. (3.19) becomes

$$\mathbf{j} \times \mathbf{b}_0 = \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ j_x & j_y & j_z \\ 0 & 0 & 1 \end{vmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} j_x \\ j_y \\ j_z \end{bmatrix}, \quad (3.22)$$

and the momentum equation is recast into matrix form:

$$\begin{bmatrix} i\omega - \nu_{en} & -\Omega_c & 0 \\ \Omega_c & i\omega - \nu_{en} & 0 \\ 0 & 0 & i\omega - \nu_{en} \end{bmatrix} \begin{bmatrix} j_x \\ j_y \\ j_z \end{bmatrix} = -\omega_p^2 \varepsilon_0 \mathbf{E}. \quad (3.23)$$

Dividing eqn. (3.23) by $i\omega$ yields

$$\begin{bmatrix} U & iY & 0 \\ -iY & U & 0 \\ 0 & 0 & U \end{bmatrix} \begin{bmatrix} j_x \\ j_y \\ j_z \end{bmatrix} = iX\omega\varepsilon_0 \mathbf{E}. \quad (3.24)$$

In eqn. (3.24) and in the remainder of this thesis, the following shorthand notation is used:

$$X = \frac{\omega_p^2}{\omega^2}, \quad (3.25)$$

$$Y = \frac{\Omega_c}{\omega}, \quad (3.26)$$

$$Z = \frac{\nu_{en}}{\omega}, \quad (3.27)$$

$$U = 1 + iZ. \quad (3.28)$$

Once again, an apparent deviation from the standard literature sticks out, namely the positive sign of $\text{Im}[U]$; Some authors (such as [3] and [26]) have given the equivalent expression $U = 1 - iZ$, which is a result of their usage of plane waves of the form $g \propto e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$.

Solving the inverse of the matrix in eqn. (3.24) gives Ohm's law

$$\mathbf{j} = \boldsymbol{\sigma} \cdot \mathbf{E}, \quad (3.29)$$

where $\boldsymbol{\sigma}$ is the sought conductivity tensor

$$\boldsymbol{\sigma} = i\omega\varepsilon_0 \frac{X}{U^2 - Y^2} \begin{bmatrix} U & -iY & 0 \\ iY & U & 0 \\ 0 & 0 & \frac{U^2 - Y^2}{U} \end{bmatrix}. \quad (3.30)$$

3.5 Dielectric tensor and the wave equation

We now wish to solve the radio wave fields \mathbf{B}_1 and \mathbf{E} self-consistently in a cold magnetoplasma. In what follows, we assume \mathbf{B}_0 to not vary appreciably over one wavelength $\lambda = 2\pi/|\mathbf{k}|$. Consequently, $\nabla \times \mathbf{B} \approx \nabla \times \mathbf{B}_1$, and the relevant Maxwell's equations (3.8)-(3.9) are written using the plane wave operators (3.15)-(3.17) as

$$\mathbf{k} \times \mathbf{E} = \omega \mathbf{B}_1, \quad (3.31)$$

$$i\mathbf{k} \times \mathbf{B}_1 = \mu_0 \boldsymbol{\sigma} \cdot \mathbf{E} - i\frac{\omega}{c^2} \mathbf{E}. \quad (3.32)$$

In eqn. (3.32), we have used the result (3.30) to relate \mathbf{j} and \mathbf{E} by Ohm's law. Inserting (3.31) into (3.32), and noting that $c^2 = \frac{1}{\epsilon_0\mu_0}$, one obtains

$$\frac{c^2}{\omega^2} \mathbf{k} \times (\mathbf{k} \times \mathbf{E}) + \boldsymbol{\epsilon}_r \cdot \mathbf{E} = \mathbf{0}, \quad (3.33)$$

where the dimensionless dielectric tensor has been defined as

$$\boldsymbol{\epsilon}_r \equiv \boldsymbol{\mathcal{I}} + \frac{i}{\omega\epsilon_0} \boldsymbol{\sigma} = \begin{bmatrix} S & iD & 0 \\ -iD & S & 0 \\ 0 & 0 & P \end{bmatrix}, \quad (3.34)$$

$$S = 1 - \frac{XU}{U^2 - Y^2}, \quad (3.35)$$

$$D = \frac{XY}{U^2 - Y^2}, \quad (3.36)$$

$$P = 1 - \frac{X}{U}. \quad (3.37)$$

The dielectric tensor (3.34) verifies the earlier assertion that magnetoplasmas are anisotropic media; the dielectric response of the magnetoplasma is sensitive to the orientation of the EM fields. What's more, the dielectric tensor is not even necessarily Hermitian — when collisions are allowed for, waves are attenuated by absorption as the wave energy is dissipated by Ohmic heating. Methods of evaluating absorption losses are presented in section 4 of this thesis.

At this point, we define the complex refractive index \mathbf{n} by

$$\mathbf{k} = \frac{\mathbf{n}\omega}{c}. \quad (3.38)$$

Moreover, utilizing the vector identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$, the wave equation (3.33) takes the form

$$\mathbf{n}(\mathbf{n} \cdot \mathbf{E}) - n^2 \mathbf{E} + \boldsymbol{\epsilon}_r \cdot \mathbf{E} = \mathbf{0}. \quad (3.39)$$

Without loss of generality, we may further choose the real wave vector $\text{Re}[\mathbf{k}]$ (and thus also $\text{Re}[\mathbf{n}]$) to lie in the xz -plane, so that the angle between \mathbf{B}_0 and $\text{Re}[\mathbf{k}]$ is θ :

$$n_x = n \sin \theta, \quad n_y = 0, \quad n_z = n \cos \theta. \quad (3.40)$$

Thus, in this coordinate system, the wave equation (3.39) can be written in matrix form as

$$\begin{bmatrix} S - n^2 \cos^2 \theta & iD & n^2 \cos \theta \sin \theta \\ -iD & S - n^2 & 0 \\ n^2 \cos \theta \sin \theta & 0 & P - n^2 \sin^2 \theta \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \equiv \mathbf{M} \cdot \mathbf{E} = \mathbf{0}. \quad (3.41)$$

The problem of finding possible modes of propagation, or equivalently, the refractive index n , has been reduced to a problem of finding the nontrivial solutions of \mathbf{E} from eqn. (3.41). Under the assumptions described above, such as restrictions on small-amplitude perturbations and nearly homogeneous \mathbf{B}_0 , eqn. (3.41) represents a self-consistent solution to EM wave propagation in a cold magnetoplasma. From here, we proceed towards deriving the all-important Appleton-Hartree refractive index and the corresponding polarization equations.

3.6 Dispersion relations: the Booker quartic and Appleton-Hartree formula

Nontrivial (i.e. nonzero) solutions of the radio wave field \mathbf{E} exist if and only if the determinant of the matrix \mathbf{M} in eqn. (3.41) is zero for some value of the refractive index n . The determinant condition yields

$$\det \mathbf{M} = An^4 - Bn^2 + C = 0, \quad (3.42)$$

where the shorthand notation below has been used:

$$A = S \sin^2 \theta + P \cos^2 \theta, \quad (3.43)$$

$$B = (S^2 - D^2) \sin^2 \theta + SP(1 + \cos^2 \theta), \quad (3.44)$$

$$C = (S^2 - D^2)P. \quad (3.45)$$

The determinant condition (3.42) is the general dispersion relation of an electron magnetoplasma, whose solution is the Appleton-Hartree refractive index. Before diving headfirst into solving it, we should process the expression a little to reveal the relative simplicity of an electron plasma treatment. Inserting the definition of the refractive index $n^2 = (ck/\omega)^2$ into (3.42), and substituting (3.43)-(3.45) with the definitions (3.35)-(3.37), and finally multiplying by $2\omega^4 U (U^2 - Y^2)$, one obtains the dispersion relation

$$\begin{aligned} & [U^2(U - X) - UY^2 + XY^2 \cos^2 \theta] c^4 k^4 \\ & + [Y^2(2U - X) - 2U(U - X)^2 - XY^2 \cos^2 \theta] c^2 k^2 \omega^2 \\ & + (U - X) [(U - X)^2 - Y^2] \omega^4 = 0, \end{aligned} \quad (3.46)$$

or equivalently, defining $\mathbf{Y} = Y\mathbf{b}_0$ and noting that $(\mathbf{k} \cdot \mathbf{Y}) = kY \cos \theta$, one obtains [26]

$$\begin{aligned} & [U^2(U - X) - UY^2] c^4 k^4 + X(\mathbf{k} \cdot \mathbf{Y})^2 c^4 k^2 \\ & + [Y^2(2U - X) - 2U(U - X)^2] c^2 k^2 \omega^2 - X(\mathbf{k} \cdot \mathbf{Y})^2 c^2 \omega^2 \\ & + (U - X) [(U - X)^2 - Y^2] \omega^4 = 0. \end{aligned} \quad (3.47)$$

The all-important dispersion relation (3.47) is sometimes referred to as the Booker quartic [26], and will prove to be of paramount importance for a versatile ray tracing algorithm when used as a so-called ray Hamiltonian.

Returning to the determinant condition (3.42), the values of n^2 are obtained as solutions of the common quadratic equation:

$$n^2 = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} \quad (3.48)$$

This gives, at least in principle, the sought refractive index of a wave propagating in a magnetoplasma. In an electron plasma, however, the expression (3.48) can be simplified into a more illuminating form. Adding An^2 to both sides of (3.42) and rearranging gives

$$n^2 = \frac{An^2 - C}{An^2 + A - B}, \quad (3.49)$$

and inserting (3.48) into the right hand side of (3.49) gives

$$n^2 = 1 - \frac{2(A - B + C)}{2A - B \pm \sqrt{B^2 - 4AC}}. \quad (3.50)$$

Finally, inserting (3.43)-(3.45) into (3.50) prompts a somewhat laborious but straightforward simplification. The result is the Appleton-Hartree refractive index [3]:

$$n^2 = 1 - \frac{X(U - X)}{U(U - X) - \frac{1}{2}Y^2 \sin^2 \theta + S_R}, \quad (3.51)$$

$$S_R = \pm \left[\frac{1}{4}Y^4 \sin^4 \theta + Y^2(U - X)^2 \cos^2 \theta \right]^{1/2}. \quad (3.52)$$

The refractive index will be used almost exclusively in the form given by eqns. (3.51)-(3.52), and contains the effects of an external geomagnetic field as well as electron-neutral collisions. Arguably the most conspicuous feature of n^2 (both theoretically and for practical radio engineering) is that the ionosphere is doubly refracting; In the ionosphere, n^2 can take two values for a given ω and $\hat{\mathbf{k}}$, corresponding to the choice of plus or minus sign of S_R (3.52). These two branches of the dispersion relation represent the two magnetoionic wave modes, commonly referred to as the ordinary (O) and extraordinary (X) modes, and is a property caused solely by the anisotropy of magnetoplasmas. In section 3.8, we will show that the O- and X-mode correspond to different polarization states (except for in coupling points where the polarizations are equal). As a result, a radio wave incident on the ionosphere from the ground will be split into two waves that propagate separately, a phenomenon known as magnetoionic splitting [3].

In the absence of a geomagnetic field, $Y = 0$ and the plasma is said to be isotropic. As a result, the refractive index simplifies to

$$n^2 = 1 - \frac{X}{U}. \quad (3.53)$$

In the isotropic case, no magnetoionic splitting takes place as only one value of n^2 is possible for a given ω and $\hat{\mathbf{k}}$. Furthermore, if the plasma is collisionless, $U = 1$ and the refractive index is given by the simple expression

$$n^2 = 1 - X. \quad (3.54)$$

The refractive index is a complex number, and is for practical purposes written as

$$n = \mu + i\chi \quad (3.55)$$

where μ and χ are both real and positive numbers. If $\mu = 0$, \mathbf{k} is imaginary and the EM wave appears as a non-propagating evanescent wave of the form

$$\mathbf{E} = \tilde{\mathbf{E}} e^{-\frac{\omega}{c}\chi \cdot \mathbf{r}} e^{-i\omega t}. \quad (3.56)$$

Although the evanescent wave demonstrates harmonic variation time, there is no harmonic variation in space. In fact, it can be shown that evanescent waves may

propagate some energy perpendicular to the wave normal [3]. However, the ray tracing formalism used in this thesis requires that the refractive index be mostly real with only a small imaginary part; This assumption is titled the approximation of pseudoreal rays, and is of practical use for estimating wave attenuation by absorption where χ is taken to be in the direction of the real wave vector. As a result, the ray tracing formalism in this thesis cannot (and should not) propagate evanescent waves.

3.7 Cutoffs, resonances and coupling points of the refractive index

For certain values of X , Y and U , the refractive index goes to zero and the wave is said to be at a *cutoff*. From eqn. (3.42) it can be seen that one (and only one) value of n^2 is zero when $C = 0$. This occurs when

$$n^2 = 0 \text{ if } \begin{cases} S^2 - D^2 = 0 & \Leftrightarrow X = U \pm Y \\ \text{or} \\ P = 0 & \Leftrightarrow X = U \end{cases}$$

Since X and Y real-valued quantities, these conditions cannot be satisfied in a collisional plasma where U is complex-valued. Moreover, it is worth noting that these cutoff conditions do not depend on θ ; in other words, they do not depend on the propagation direction of the wave. If the wave encounters a cutoff, it cannot propagate further and is reflected [11].

In an anisotropic plasma, the refractive index can be infinite and is said to be at *resonance*. From eqn. (3.48), it can be seen that one (and only one) value of n^2 is infinite if $A = 0$. Unless if $U = 1$ and $Y = 1$, this occurs when [3]

$$|n^2| \rightarrow \infty \text{ if } X = \frac{U(U^2 - Y^2)}{U^2 - Y^2 \cos^2 \theta}.$$

As an example, in the special case of perpendicular propagation ($\theta = \pi/2$) in a collisionless plasma ($U = 1$) one recovers the upper hybrid resonance [11]

$$X = 1 - Y^2 \Leftrightarrow \omega = \sqrt{\omega_p^2 + \Omega_c^2}. \quad (3.57)$$

Furthermore, in the case of propagation parallel to the magnetic field ($\theta = 0$ or $\theta = \pi$) in a collisionless plasma ($U = 1$), one value of n^2 is infinite when $Y = 1$. This is known as cyclotron resonance, as it occurs when the wave is oscillating at the gyrofrequency.

Resonances are efficient mechanisms of dissipating wave energy in the plasma as heat. One should, however, keep in mind that infinities of the refractive index inevitably signal a breakdown of the cold plasma approximation, and should be regarded as an incomplete description of the big picture. This is because the wave phase speed becomes smaller than the plasma thermal speed, which violates the fundamental assumption of cold plasma waves. When kinetic effects are accounted

for, the theory of warm plasmas (see e.g. Ginzburg 1964 [5]) shows that at resonance n^2 continuously becomes only very large, but not infinite. The cold plasma ray tracing formalism adopted in this thesis should thus maintain diagnostics to detect resonances. In the ray tracer developed for this thesis, the ray is stopped if $\text{Re}[n]$ becomes too large (e.g. 10^4), signaling that the refractive index becomes very large and is likely at resonance.

Points where the two refractive indices are equal in an anisotropic plasma are referred to as coupling points, and occur when S_R is zero. This is only possible if [3]

$$X = 1 \text{ and } Z = \frac{Y \sin^2 \theta}{2 |\cos \theta|} \equiv Z_t. \quad (3.58)$$

In a collisionless plasma therefore, the refractive indices may be equal only if $\mathbf{k} \parallel \mathbf{B}_0$, because then $\sin \theta = 0$ and $Z = Z_t = 0$. The point $X = 1$ is somewhat troublesome in ray tracing, because the refractive index may be indeterminate at this point. This is related to the phenomenon of the "Spitze": If a ray is propagating in the magnetic meridian from the ground, for high enough elevation angles the ray may reach an altitude where $X = 1$ and is reflected sharply. In this case, special numerical techniques are required and are discussed in section 4. An example of a Spitze is presented in figure 7.

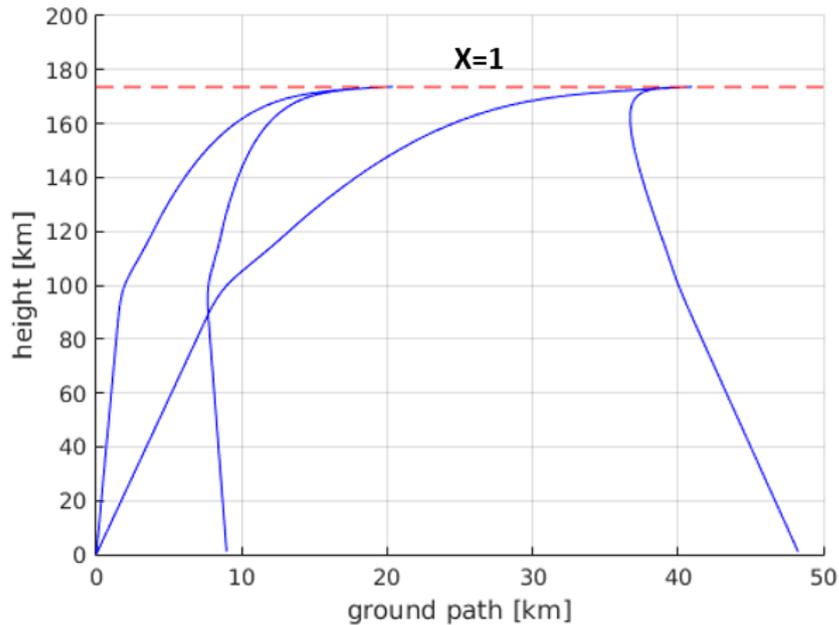


Figure 7: Two ordinary rays with launch elevation angles 85° and 87° passing through a Spitze in a vertically varying ionosphere. The altitude where $X = 1$ is drawn by a dashed line. At this altitude, the ray paths are reflected sharply and require special care in numerical integration. The ray paths have been solved using the developed ray tracing program.

3.8 Polarization of magnetoionic modes

3.8.1 Constitutive relations

Applying Maxwell's equations to the radio wave fields \mathbf{E} and $\mathbf{H}_1 = \mathbf{B}_1/\mu_0$, and using the plane wave operators (3.15)-(3.17), the following relations are recovered in a general magnetoplasma:

$$i\mathbf{k} \cdot \mathbf{D} = \rho_f, \quad (3.59)$$

$$i\mathbf{k} \cdot \mathbf{H}_1 = 0, \quad (3.60)$$

$$i\mathbf{k} \times \mathbf{E} = i\omega\mu_0\mathbf{H}_1, \quad (3.61)$$

$$i\mathbf{k} \times \mathbf{H}_1 = \mathbf{j}_f - i\omega\mathbf{D}. \quad (3.62)$$

Quasineutrality proves to be a valid approximation in the ionosphere, dictating that the free charge density $\rho_f \approx 0$ [11]. Furthermore, we assume there to be no free current \mathbf{j}_f to affect the propagation of radio waves; the current $\mathbf{j} = \boldsymbol{\sigma} \cdot \mathbf{E}$ used in eqn. (3.32) is assumed to only consist of the polarization current density, formed by the motion of free electrons in response to the radio wave electric field [3]. Consequently, we assume that $\mathbf{j}_f = \mathbf{0}$ in eqn. (3.62). Choosing the coordinate axes so that $\mathbf{k} \parallel \hat{\mathbf{e}}_z$, Maxwell's equations (3.59)-(3.62) for plane waves give [3]

$$D_z = 0, \quad (3.63)$$

$$H_{1z} = 0, \quad (3.64)$$

$$-kE_y = \omega\mu_0H_{1x}, \quad (3.65)$$

$$kE_x = \omega\mu_0H_{1y}, \quad (3.66)$$

$$kH_{1y} = \omega D_x, \quad (3.67)$$

$$kH_{1x} = -\omega D_y. \quad (3.68)$$

From eqns. (3.63)-(3.64), we conclude that \mathbf{D} , \mathbf{H}_1 and \mathbf{B}_1 are all perpendicular to the wave normal. The same does not generally hold for \mathbf{E} , which can exhibit a longitudinal component in anisotropic media. Moreover, dividing (3.65) by (3.66) and (3.68) by (3.67) shows that

$$\frac{D_y}{D_x} = \frac{E_y}{E_x} = -\frac{H_{1x}}{H_{1y}} \equiv \rho_\perp, \quad (3.69)$$

where ρ_\perp is defined as the transverse polarization of the wave. In the next section, we will derive an expression for ρ_\perp in anisotropic plasmas, as well as present an expression for the longitudinal polarization ρ_\parallel .

3.8.2 Polarization formulae

To begin the discussion on polarization, it is worth taking a quick look at what occurs in a simple isotropic plasma. It will be shown, that an isotropic plasma does not restrict the transverse polarization in any way, while an anisotropic plasma allows only two states of polarization.

In an isotropic plasma, the conductivity- (see eqn. (3.30)) and dielectric tensors (see eqn. (3.34)) are scalars, respectively

$$\sigma = i\omega\varepsilon_0 \frac{X}{U}, \quad \varepsilon_r = 1 + \frac{i}{\omega\varepsilon_0} \sigma = 1 - \frac{X}{U}.$$

Inserting the scalar ε_r into the general wave equation (3.39), and (without loss of generality) choosing the coordinate axes so that $\mathbf{n} \parallel \hat{\mathbf{e}}_z$, one obtains

$$\begin{bmatrix} \varepsilon_r - n^2 & 0 & 0 \\ 0 & \varepsilon_r - n^2 & 0 \\ 0 & 0 & \varepsilon_r \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \mathbf{0}. \quad (3.70)$$

Finding the nontrivial solutions from the determinant condition gives two solutions: One of them is $\varepsilon_r = 0$, or equivalently $X = U$. In a noncollisional plasma ($U = 1$) there exists a solution $\omega = \omega_p$, which describes a longitudinal ($E_z \neq 0$) and non-propagating electrostatic oscillation. It is well known that longitudinal polarizations ($\mathbf{k} \parallel \mathbf{E}$) cannot propagate in an isotropic medium or vacuum [13], so it is of little interest to study them for radio waves transmitted from the ground. The other solution is $n^2 = \varepsilon_r = 1 - X/U$, which describes propagating transverse waves. This solution in turn gives two linearly independent vectors for \mathbf{E} , for example

$$\mathbf{E} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{E} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \quad (3.71)$$

The choice of electric vectors in eqn. (3.71) is not unique. In isotropic media, the common choice of linearly independent polarizations are called the TE (transverse electric) and TM (transverse magnetic) waves [13].

Now we turn our attention to the more general anisotropic case. In principle, we could solve the eigenvectors \mathbf{E} from eqn. (3.41) as a function of n^2 and θ , yielding one distinct polarization for each of the two characteristic modes (O/X). We could then, in principle, express this polarization in any other coordinate system by performing a sequence of rotation operations. This however leads to rather cumbersome expressions that are not that easy to simplify. Here, a suitable choice of coordinate axes goes a long way with a bit of physical reasoning.

In eqn. (3.34), the dielectric tensor has been given in a coordinate system where $\mathbf{B}_0 \parallel \hat{\mathbf{e}}_z$. For studying transverse and longitudinal polarizations, one would be arguably better off in a coordinate system where the wave normal is parallel to one of the coordinate axes, e.g. $\text{Re}[\mathbf{k}] \parallel \hat{\mathbf{e}}_z$. This transformation can be simply performed by rotating the coordinate system by the angle θ around the y-axis; Such a right-hand rotation is given by the rotation matrix

$$R_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \quad (3.72)$$

and its inverse

$$R_y^{-1}(\theta) = R_y^T(\theta) = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}. \quad (3.73)$$

In this new coordinate system, in which $\text{Re}[\mathbf{k}] \parallel \hat{\mathbf{e}}_z$ and \mathbf{B}_0 lies in the xz -plane at an angle θ to $\text{Re}[\mathbf{k}]$, the dielectric tensor $\boldsymbol{\epsilon}'_r$ can be found by a rotation of (3.34) as follows:

$$\boldsymbol{\epsilon}'_r = R_y^{-1}(\theta) \boldsymbol{\epsilon}_r R_y(\theta) \quad (3.74)$$

$$= \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} S & iD & 0 \\ -iD & S & 0 \\ 0 & 0 & P \end{bmatrix} \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \quad (3.75)$$

$$= \begin{bmatrix} S \cos^2 \theta + P \sin^2 \theta & iD \cos \theta & (S - P) \cos \theta \sin \theta \\ -iD \cos \theta & S & -iD \sin \theta \\ (S - P) \cos \theta \sin \theta & iD \sin \theta & S \sin^2 \theta + P \cos^2 \theta \end{bmatrix}. \quad (3.76)$$

An equivalent form of the dielectric tensor (3.76) is given by Budden [3] (page 53, eqn. (3.54)).

The electric displacement of the wave in this system is given by the constitutive relation

$$\mathbf{D} = \varepsilon_0 \boldsymbol{\epsilon}'_r \cdot \mathbf{E}. \quad (3.77)$$

The components of this constitutive relation read

$$\begin{cases} D_x/\varepsilon_0 = (S \cos^2 \theta + P \sin^2 \theta)E_x + iD \cos \theta E_y + (S - P) \cos \theta \sin \theta E_z \\ D_y/\varepsilon_0 = -iD \cos \theta E_x + S E_y - iD \sin \theta E_z \\ D_z/\varepsilon_0 = (S - P) \cos \theta \sin \theta E_x + iD \sin \theta E_y + (S \sin^2 \theta + P \cos^2 \theta)E_z \end{cases} \quad (3.78)$$

In fact, Gauss's law (eqn. (3.59)) posits that $D_z = 0$ in this coordinate system. From this condition, one can solve E_z from the z -component of (3.78) as

$$E_z = \frac{(P - S) \cos \theta \sin \theta E_x - iD \sin \theta E_y}{S \sin^2 \theta + P \cos^2 \theta}. \quad (3.79)$$

Inserting (3.79) into the x - and y -components of (3.78), and dividing D_y by D_x gives

$$\frac{D_y}{D_x} = \frac{2iXY(X - U) \cos \theta E_x + [2U(U - X)^2 + Y^2(X - 2U) + XY^2(1 - 2 \sin^2 \theta)] E_y}{2(U - X)((U^2 - UX - Y^2)E_x + iXY \cos \theta E_y)} \quad (3.80)$$

From the definition of transverse polarization (3.69), substitution of $D_y/D_x = E_y/E_x = \rho_\perp$ into (3.80) gives after some simplification a quadratic equation for ρ_\perp [3]:

$$\rho_\perp^2 - i \frac{Y \sin^2 \theta}{(U - X) \cos \theta} \rho_\perp + 1 = 0. \quad (3.81)$$

The solution yields two possible states of polarization

$$\rho_{\perp} = i \frac{\left(\frac{1}{2}Y^2 \sin^2 \theta + S_R\right)}{Y(U - X) \cos \theta}, \quad (3.82)$$

where S_R is the same parameter as in the Appleton-Hartree formula and is defined by (3.52). The sign of S_R can be shown to correspond to its respective magnetoionic mode [3].

From eqn. (3.79) it is now possible to solve for the longitudinal polarization by dividing the equation by E_x and using the transverse polarization $E_y/E_x = \rho_{\perp}$. The result is given by [3] in the equivalent form

$$\rho_{\parallel} \equiv \frac{E_z}{E_x} = -i\rho_{\perp} \frac{Y \sin \theta (n^2 - 1)}{U - X}. \quad (3.83)$$

As can be seen, the longitudinal polarization is generally not zero in an anisotropic plasma, meaning that the electric field may oscillate in the direction of the wave normal \mathbf{k} . As a direct result of this, the energy carried by plasma EM waves is generally not propagated in the same direction as \mathbf{k} [3]. Furthermore, because n^2 and ρ_{\perp} are generally different between the two magnetoionic modes, ρ_{\parallel} is different as well which consequently leads to the wave energy being propagated in different directions depending on the mode.

3.8.3 Polarization coordinate system transformations

The polarization formulae (3.82)-(3.83) are by definition given in a coordinate system where $\mathbf{k} \parallel \hat{\mathbf{z}}$ and \mathbf{B}_0 is in the xz -plane. For future reference, this system will be referred to as the *local polarization system*. Since the polarization of the wave can be given in practically any basis vectors, e.g. the system defined by the spherical coordinates of the radiating antenna or the plane of incidence at ground reflection, it is prudent to present a recipe for converting the polarization to and from the local polarization system.

Assume that all vectors \mathbf{B}_0 , \mathbf{k} and \mathbf{E} are expressed by the fixed orthonormal basis vectors $(\hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\mathbf{z}}_0)$. For example, one such suitable basis is the convention presented in section 2 for the ECEF system: $\hat{\mathbf{x}}_0$ points from the center of the Earth to the Greenwich meridian ($0^\circ N, 0^\circ E$), $\hat{\mathbf{y}}_0$ points from the center of the Earth to ($0^\circ N, 90^\circ E$), and $\hat{\mathbf{z}}_0$ points from the center of the Earth to ($90^\circ N, 0^\circ E$). In the case that $\mathbf{k} \not\parallel \mathbf{B}_0$, the local polarization system $(\hat{\mathbf{x}}_1, \hat{\mathbf{y}}_1, \hat{\mathbf{z}}_1)$ is given by

$$\hat{\mathbf{z}}_1 = \frac{\mathbf{k}}{|\mathbf{k}|}, \quad \hat{\mathbf{y}}_1 = \frac{\mathbf{k} \times \mathbf{B}_0}{|\mathbf{k} \times \mathbf{B}_0|}, \quad \hat{\mathbf{x}}_1 = \hat{\mathbf{y}}_1 \times \hat{\mathbf{z}}_1. \quad (3.84)$$

In the singular case that $\mathbf{k} \parallel \mathbf{B}_0$, the local polarization system is not unique, and can be formed by finding any vector $\hat{\mathbf{u}} \in \{\hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\mathbf{z}}_0\}$ that is not parallel to \mathbf{k} and setting

$$\hat{\mathbf{z}}_1 = \frac{\mathbf{k}}{|\mathbf{k}|}, \quad \hat{\mathbf{y}}_1 = \frac{\mathbf{k} \times \hat{\mathbf{u}}}{|\mathbf{k} \times \hat{\mathbf{u}}|}, \quad \hat{\mathbf{x}}_1 = \hat{\mathbf{y}}_1 \times \hat{\mathbf{z}}_1. \quad (3.85)$$

Next, assume that the wave demonstrates a transverse polarization $\rho_{\perp,2}$ in some arbitrary orthonormal system $(\hat{\mathbf{x}}_2, \hat{\mathbf{y}}_2, \hat{\mathbf{z}}_2)$ so that $\mathbf{E} = E_{x2} (\hat{\mathbf{x}}_2 + \rho_{\perp,2} \hat{\mathbf{y}}_2)$ and $\mathbf{k} \parallel \hat{\mathbf{z}}_2$. We'd wish to express this electric vector in the form of $\mathbf{E} = E_{x1} \hat{\mathbf{x}}_1 + E_{y1} \hat{\mathbf{y}}_1$. Taking the dot product of the latter with $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{y}}_1$, one finds that

$$E_{x1} = E_{x2} (\hat{\mathbf{x}}_2 + \rho_{\perp,2} \hat{\mathbf{y}}_2) \cdot \hat{\mathbf{x}}_1, \quad E_{y1} = E_{x2} (\hat{\mathbf{x}}_2 + \rho_{\perp,2} \hat{\mathbf{y}}_2) \cdot \hat{\mathbf{y}}_1. \quad (3.86)$$

Consequently, the transverse polarization in the local polarization system is $\rho_{\perp,1} = E_{y1}/E_{x1} = (\hat{\mathbf{x}}_2 + \rho_{\perp,2} \hat{\mathbf{y}}_2) \cdot \hat{\mathbf{y}}_1 / (\hat{\mathbf{x}}_2 + \rho_{\perp,2} \hat{\mathbf{y}}_2) \cdot \hat{\mathbf{x}}_1$.

The procedure can be worked the other way to convert from the local polarization system to some other system. Given that $\mathbf{E} = E_{x1} (\hat{\mathbf{x}}_1 + \rho_{\perp,1} \hat{\mathbf{y}}_1)$ in the local polarization system, and given the basis vectors $(\hat{\mathbf{x}}_2, \hat{\mathbf{y}}_2, \hat{\mathbf{z}}_2)$ and $\hat{\mathbf{z}}_2 \parallel \mathbf{k}$ in some other system, we would wish to express the same field by $\mathbf{E} = E_{x2} \hat{\mathbf{x}}_2 + E_{y2} \hat{\mathbf{y}}_2$. Analogously to the formula (3.86), we find that

$$E_{x2} = E_{x1} (\hat{\mathbf{x}}_1 + \rho_{\perp,1} \hat{\mathbf{y}}_1) \cdot \hat{\mathbf{x}}_2, \quad E_{y2} = E_{x1} (\hat{\mathbf{x}}_1 + \rho_{\perp,1} \hat{\mathbf{y}}_1) \cdot \hat{\mathbf{y}}_2, \quad (3.87)$$

and furthermore, that $\rho_{\perp,2} = E_{y2}/E_{x2} = (\hat{\mathbf{x}}_1 + \rho_{\perp,1} \hat{\mathbf{y}}_1) \cdot \hat{\mathbf{y}}_2 / (\hat{\mathbf{x}}_1 + \rho_{\perp,1} \hat{\mathbf{y}}_1) \cdot \hat{\mathbf{x}}_2$.

3.8.4 Polarization ellipses

For an EM wave in a magnetoplasma, the electric field can be expressed using the polarizations as

$$\mathbf{E} = \tilde{E}_x \begin{bmatrix} 1 \\ \rho_{\perp} \\ \rho_{\parallel} \end{bmatrix} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}, \quad (3.88)$$

where \tilde{E}_x is the complex amplitude of E_x . The real part of the electric field vectors trace out ellipses, with the direction of rotation being opposite for the O- and X-mode. As an example, consider the special case of propagation parallel to the magnetic field $\theta = 0$. The polarizations (3.82) and (3.83) give the well-known result

$$\rho_{\perp} = \pm i, \quad \rho_{\parallel} = 0.$$

This outcome represents circularly polarized waves, with (+) corresponding to the O-mode and (−) to the X-mode if $X < 1$. If $X > 1$, the signs are reversed for the modes. In this thesis, the polarization handedness is defined from the view of an observer looking in the direction of \mathbf{k} . Hence, $\rho_{\perp} = +i$ is defined as being circularly polarized in the left-handed sense, while again $\rho_{\perp} = -i$ is circularly polarized in the right-handed sense.

With the definition above, the particular handedness of polarization in the xy-plane can be determined unambiguously from the phase delay of E_y with respect to E_x . From equation (3.88), the y-component of the complex electric field is written

$$E_y = \tilde{E}_x \rho_{\perp} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} \quad (3.89)$$

$$= \tilde{E}_x |\rho_{\perp}| e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t + \varphi)}, \quad (3.90)$$

where φ is the phase delay of E_y with respect to E_x , given by

$$|\rho_{\perp}| \cos \varphi = \operatorname{Re}[\rho_{\perp}] \quad (3.91)$$

$$|\rho_{\perp}| \sin \varphi = \operatorname{Im}[\rho_{\perp}] \quad (3.92)$$

$$\Rightarrow \varphi = \operatorname{atan2}(\operatorname{Im}[\rho_{\perp}], \operatorname{Re}[\rho_{\perp}]). \quad (3.93)$$

Here $\operatorname{atan2}(y, x)$ denotes the 2-argument arctangent function that is capable of identifying the correct quadrant of a point (x, y) . The $\operatorname{atan2}$ -function has been implemented in standard libraries of many modern programming languages. If $\varphi \in (0, \pi)$, the sense of polarization is left-handed, and if $\varphi \in (-\pi, 0)$, the sense of polarization is right-handed. If φ is zero or π (i.e. if ρ_{\perp} is purely real), or if $\rho_{\perp} = 0$, the electric field is said to be linearly polarized.

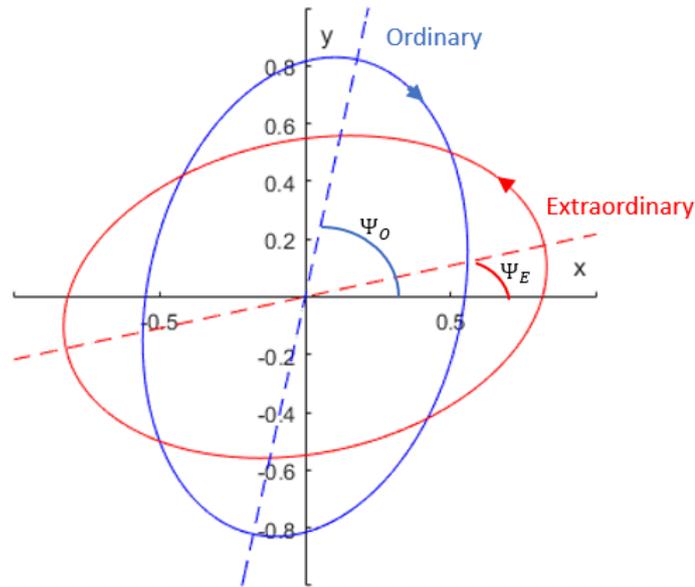


Figure 8: Polarization ellipses of the O- and X-mode using some set of representative parameters $X < 1$. If $X > 1$, the rotation directions are reversed. The wave normal is pointing towards the reader, i.e. in the positive z -direction, and hence the ordinary mode is polarized in the left-handed sense. The electric field vectors have been normalized to be of the same magnitude for clarity.

In general, the polarization ρ_{\perp} is complex, and as a result, the major/minor axes of the ellipses are not necessarily parallel to the coordinate axes. An illustration of polarization ellipses projected to the xy -plane is presented in figure 8. The symbol $\Psi_{O/E}$ denotes the tilt angle of the respective polarization ellipse, and represents the angle the ellipse major axis makes with the x -axis. The axis ratio $M_{O/E}$ in turn denotes the ratio of minor axis to major axis, and can be either positive or negative. The tilt angle and axis ratio of both ellipses can be evaluated from knowing the ordinary wave polarization $\rho_{\perp,O}$ as follows [3]: Let

$$\rho_{\perp,O} = \tan \gamma = u + iv, \quad (3.94)$$

and

$$\gamma = \Psi_O + i\phi_O. \quad (3.95)$$

The O-ellipse major axis makes an angle Ψ_O with the x-axis, while the X-ellipse major axis is at an angle $\Psi_E = \frac{\pi}{2} - \Psi_O$ [3]. The modulus of the axis ratio can be evaluated for both modes as

$$|M_{O/E}| = |\tanh \phi_O|. \quad (3.96)$$

Finally, the sign of the axis ratio is chosen as follows: For left-handed polarization M is taken as positive, and is taken to be negative for right-handed polarization. In the case that the polarization is linear, M is equal to zero.

The discussion presented here on polarization ellipses can be used to compute the coupling of incident waves to magnetoionic modes. This knowledge will then be used in section 4.7 to estimate how a wave with a given polarization will be split into O/X-waves upon entering the ionosphere.

3.9 Wave mode interchange

It is prudent to adopt some definition of the terms "ordinary" and "extraordinary" when denoting the corresponding refractive indices and polarizations. We chose to use the same definition as in [3], and it goes as follows:

$$\begin{cases} \text{Ordinary (O)} & \text{if } \text{Re}[S_R] > 0 \\ \text{Extraordinary (X)} & \text{if } \text{Re}[S_R] < 0 \end{cases}$$

In some special cases $\text{Re}[S_R]$ can be equal to zero, and an ambiguity presents itself in the above definition. This occurs only if one of the following conditions is met:

1. $Z > Z_t$ (3.58) and $X = 1$, or
2. $\sin \theta = 0$, $Z = 0$ and $X = 1$.

The developed C++-program presented in section 5 always selects the principal branch of the multivalued complex square root, for which $\text{Re}[\sqrt{\mathcal{Z}}] \geq 0$ for any complex number \mathcal{Z} . Except for the points where $\text{Re}[S_R] = 0$ described above, selecting (+) for the sign of S_R always corresponds to the ordinary mode and (-) to the extraordinary mode.

When a propagating wave crosses a point where $X = 1$, the wave may interchange its mode and continue propagating as the other mode if the above condition 1 is fulfilled. An illustration of this process is presented in figure 9. Consider that an ordinary wave is propagating from a region of $X < 1$ to a region of $X > 1$ for some representative constant values Y and Z . For the low value of $\theta = 6^\circ$ (leftmost column of graphs), it follows that $Z > Z_t$ at $X = 1$. For the wave to retain its polarization, the wave has to resume propagation as an extraordinary mode after the crossing $X = 1$. One can then observe, that μ , χ and φ change continuously at the crossing. It can be argued, that such a process should not be referred to as "mode

conversion" as there is no physical process responsible for it, and in particular, no change in polarization. All that changes is the adjective used to describe the wave, and consequently, the sign used for S_R [3].

On the other hand, for a value of $\theta = 11^\circ$ (rightmost column in figure 9), it follows that $Z < Z_t$ at the crossing $X = 1$. Now the ordinary wave must remain an ordinary mode to retain its polarization [3]. To determine whether interchange occurs at $X = 1$, it is thus necessary to compare Z to Z_t .

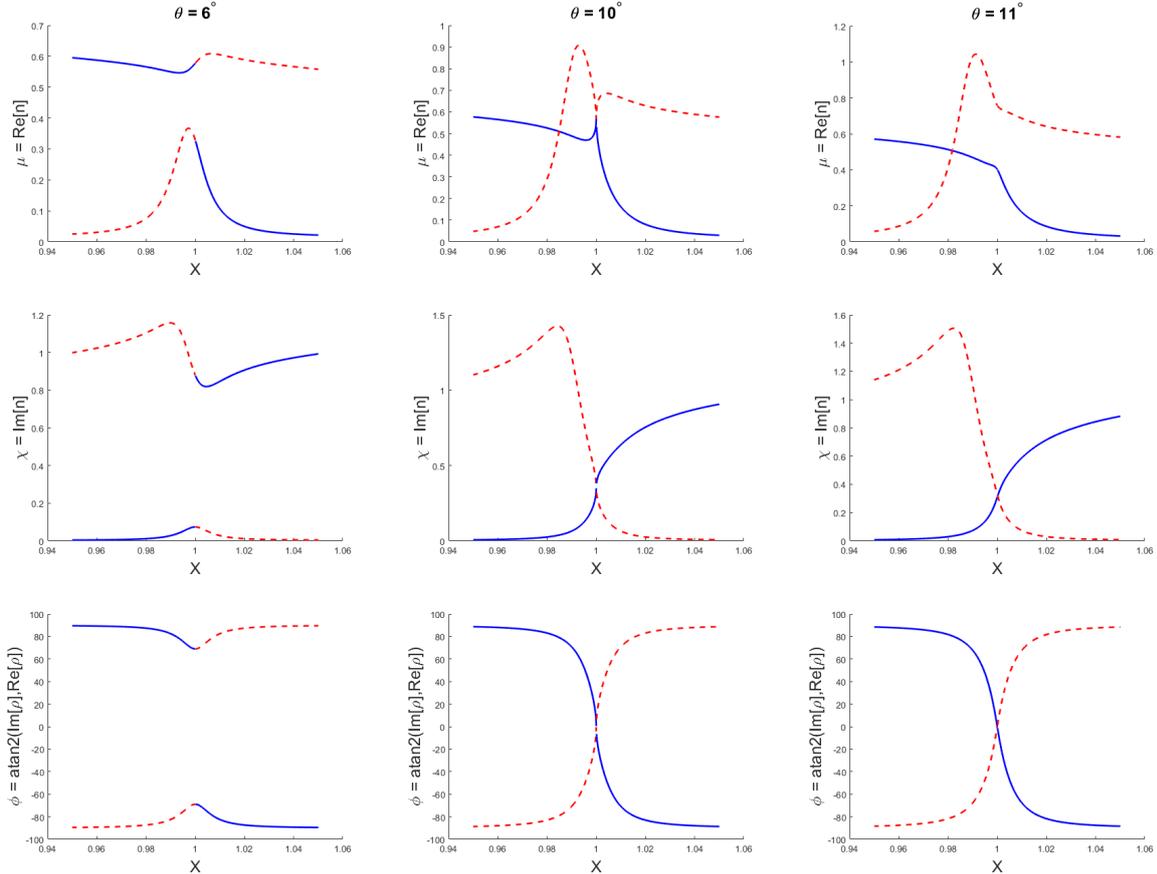


Figure 9: Illustration of mode interchange at $X = 1$ for different θ . The magnetic field and collision frequency are held constant so that $Y = 0.5$ and $Z = 0.007655$ are constant, while the wave is propagating from a low density region to a high density region so that X changes continuously over the value $X = 1$. The solid blue lines represents values of μ , χ and φ (see eqn. (3.93)) for the ordinary mode, while the dashed red lines are for the extraordinary mode. In the leftmost column of the graphs $Z > Z_t$, in the middle column $Z = Z_t$, and in the rightmost column $Z < Z_t$.

However, if the point $X = 1$ happens to be at a coupling point where $Z = Z_t$, n and ρ_\perp are identical for both modes, and it is not possible to determine mode interchange unambiguously without resorting to full wave theory. This case is presented in the middle column of figure 9. Full wave theory shows, that at such coupling points both modes can emerge [3], but in this thesis no attempt is made

to model this phenomenon. To ensure that rays may continue to propagate after coupling points (unless other reasons arise to terminate tracing, e.g. resonance or non-pseudoreal n), we have chosen to not interchange the mode at $Z = Z_t$. From a user's perspective, such a caveat should necessitate a warning to signal a compromise of ray tracing reliability.

A possible pseudocode implementation for detecting mode interchange is presented below. Suppose we denote the O-mode as $mode = 1$ and the X-mode as $mode = -1$, and store the value $X(t - \Delta t) \equiv X_{old}$ of the previous time step. The program then retrieves the current values $X(t) \equiv X_{new}$, $Y(t) \equiv Y$ and $Z(t) \equiv Z$ after the crossing. For a slowly varying ionosphere and small time step, only a small error is made in assuming that the values $Z(t)$ and $Y(t)$ remain constant at the crossing $X = 1$. The mode interchange algorithm proceeds then as given below:

```

if  $(X_{old} - 1)(X_{new} - 1) < 0$  then
  //Crossing  $X = 1$  detected
   $Z_t = CalcCriticalZ(Y, \theta)$ 
  if  $Z > Z_t$  then
    //Perform mode interchange
     $mode = -mode$ 
  else if  $Z == Z_t$  then
    //Coupling point: do not perform mode interchange
    //Warn user of possible physical breakdown
     $mode = mode$ 
  else if  $Z < Z_t$  then
    //No mode interchange: Do nothing
  end if
end if
//Store value of  $X$ 
 $X_{old} = X_{new}$ 

```

3.10 Energy transfer by EM waves

The flux of electromagnetic energy carried by the radio wave is given by the real-valued Poynting vector [3]

$$\mathbf{\Pi} = \text{Re} [\mathbf{E}] \times \text{Re} [\mathbf{H}_1]. \quad (3.97)$$

Given the radio wave fields $\mathbf{E} = \tilde{\mathbf{E}}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ and $\mathbf{H}_1 = \tilde{\mathbf{H}}_1e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$, the time averaged Poynting vector can be shown to equate to [3]

$$\mathbf{\Pi}_{av} = \frac{1}{4} \left(\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}_1^* + \tilde{\mathbf{E}}^* \times \tilde{\mathbf{H}}_1 \right) = \frac{1}{2} \text{Re} \left[\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}_1^* \right]. \quad (3.98)$$

Using the relations (3.65)-(3.66) for $\tilde{\mathbf{H}}_1$, the cross product in (3.98) is evaluated

as

$$\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}_1^* = \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ \tilde{E}_x & \tilde{E}_y & \tilde{E}_z \\ -\frac{n^* \tilde{E}_y^*}{c\mu_0} & \frac{n^* \tilde{E}_x^*}{c\mu_0} & 0 \end{vmatrix} = \frac{n^*}{c\mu_0} \begin{bmatrix} -\tilde{E}_x^* \tilde{E}_z \\ -\tilde{E}_y^* \tilde{E}_z \\ (\tilde{E}_x^* \tilde{E}_x + \tilde{E}_y^* \tilde{E}_y) \end{bmatrix}. \quad (3.99)$$

Using the definitions of the polarizations ρ_\perp and ρ_\parallel , and substituting $c = 1/\sqrt{\varepsilon_0\mu_0}$ with $n = \mu + i\chi$ into (3.99), the time averaged Poynting vector is expressed neatly by

$$\mathbf{\Pi}_{av} = \frac{|\tilde{E}_x|^2}{2Z_0} \begin{bmatrix} -\text{Re}[n^* \rho_\parallel] \\ -\text{Re}[n^* \rho_\perp^* \rho_\parallel] \\ \mu(1 + |\rho_\perp|^2) \end{bmatrix}, \quad (3.100)$$

where Z_0 denotes the wave impedance of free space. In deriving (3.100), the identities $|\mathcal{Z}|^2 = \mathcal{Z}^* \mathcal{Z}$ and $\text{Re}[\mathcal{Z}] = \frac{1}{2}(\mathcal{Z} + \mathcal{Z}^*)$ have been used for $\mathcal{Z} \in \mathbb{C}$. The result (3.100) demonstrates the earlier assertion that wave energy is generally not transported parallel to $\mathbf{k} \parallel \hat{\mathbf{e}}_z$ in an anisotropic plasma. Only in the special case that $\rho_\parallel = 0$ (which happens when e.g. $Y = 0$ or $n^2 = 1$) is $\mathbf{\Pi}_{av}$ parallel to \mathbf{k} . Thus in isotropic plasmas where $\rho_\parallel = 0$, one recovers the familiar result

$$\mathbf{\Pi}_{av} = \frac{\mu}{2Z_0} \left(|\tilde{E}_x|^2 + |\tilde{E}_y|^2 \right) \hat{\mathbf{e}}_z. \quad (3.101)$$

In this thesis, the signal amplitude \mathcal{A} is understood as the modulus of $\mathbf{\Pi}_{av}$, also known as the intensity or power per unit area transported by the EM wave. From this definition, it follows from the general result (3.100) that the signal amplitude is given by the expression [3]

$$\mathcal{A} = \frac{1}{2Z_0} \left| \boldsymbol{\mu} \left(\tilde{\mathbf{E}} \cdot \tilde{\mathbf{E}}^* \right) - \text{Re} \left[\tilde{\mathbf{E}}^* \left(\boldsymbol{\mu} \cdot \tilde{\mathbf{E}} \right) \right] \right|, \quad (3.102)$$

where $\boldsymbol{\mu} = \text{Re}[\mathbf{n}] = (c/\omega)\text{Re}[\mathbf{k}]$. From (3.102) it is deduced, that the signal amplitude increases or decreases quadratically with the magnitude of $\tilde{\mathbf{E}}$: if the electric field is transformed by $\tilde{\mathbf{E}} \rightarrow a\tilde{\mathbf{E}}$ where $a > 0$, the signal amplitude is transformed $\mathcal{A} \rightarrow a^2\mathcal{A}$. This result will be used in section 4 to estimate signal amplitude loss by absorption and magnetoionic splitting.

4 Ray tracing in the ionosphere

In previous sections we covered the physics of plasma EM wave propagation and the ionosphere. Having done this, the theory is finally put into use in this section.

Ray tracing refers to a collection of computational methods used to predict wave paths in slowly varying media. In much the same way as ray tracing is used in computer graphics to render realistic images, ray tracing in the ionosphere is a powerful tool for visualizing EM wave propagation from a transmitter on the basis of first principles. Ionospheric ray tracing has found several useful applications in e.g. OTH radar and skywave radio applications: For example, given prior knowledge of the ionosphere, predicting usable frequencies and coverage maps can be done efficiently using ray tracing. For the Suomi 100 satellite, ray tracing is proposed as a means of data analysis for its space radio instrument. The methods presented in this section allow estimation of e.g. coverage maps and signal amplitude attenuation, which are to be used in conjunction with in-situ space radio measurements.

Originally motivated by developments in both civilian and military radar technologies as well as skywave radio communications, ionospheric ray tracing techniques predate satellite communications by decades. In 1954, British physicist Jenifer Haselgrove invented her highly influential method of Hamiltonian ray tracing by applying Hamilton's principle to geometrical optics [7]. The idea of making an analogy between particle trajectories and ray paths using Hamiltonian mechanics was not completely new; In 1941, eminent Soviet physicist Lev Landau published the second volume of his legendary series on theoretical physics, in which a similar Hamiltonian ray formalism was presented [12]. It is widely accepted that Haselgrove's ionospheric ray tracing technique is most suitable for HF waves, and to some extent MF waves as well [3][26]. Lower frequencies necessitate more sophisticated complex ray tracing or full-wave methods.

Despite the excellent applicability of ray tracing in high frequency ionospheric radio, one should not be oblivious to the inherent limitations of ray theory. Being confined within the margins of geometrical optics, ray tracing is valid only for short wavelengths in media that are not very lossy. In addition to this, the theory of cold plasma waves in section 3 necessitates that the wave phase velocity greatly exceeds the plasma thermal velocity. It is not always easy to determine in advance whether a launched ray will violate any of these conditions. Fortunately, many of the common pitfalls have been studied extensively in established literature, and an abundance of versatile numerical plasma ray tracing techniques have been published for use in the ionosphere as well as laboratory plasmas (see e.g. [26], [23] and [20]).

4.1 Group velocity of wave packets

In section 3, the discussion on EM wave propagation considered only monochromatic plane waves. In reality, signals propagate as wave packets consisting of multiple frequencies, all of which may propagate at their own phase velocity. However, if the medium is not too dispersive and lossy, the wave packet propagates with its original form at a velocity known as the *group velocity*. The specific value of the group

velocity is determined at the dominant frequency component of the wave packet known as the carrier frequency. It is then the carrier frequency what is understood as the ray frequency that ultimately determines the path of the wave packet.

For simplicity, let's restrict our attention to a one-dimensional wave propagating in the z -direction. For a continuous distribution of frequencies, the electric field of the wave packet is written [2]

$$E(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} a(k) e^{i[kz - \omega(k)t]} dk. \quad (4.1)$$

If the spectrum of frequencies $a(k)$ is centered around the carrier frequency $\omega_0 = \omega(k_0)$, the wave frequency can be Taylor-expanded to the first order in the vicinity of the carrier frequency:

$$\omega(k) = \omega_0 + (k - k_0) \left. \frac{d\omega}{dk} \right|_{k=k_0} + \mathcal{O}((k - k_0)^2). \quad (4.2)$$

In simple terms, this entails that the information of the signal has been modulated to a relatively high-amplitude carrier wave using a narrow frequency bandwidth. The quantity $\left. \frac{d\omega}{dk} \right|_{k=k_0} \equiv v_g$ is known as the group velocity, and it will be shown to represent the velocity at which the wave packet transmits information. Inserting (4.2) into (4.1), we get

$$E(z, t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} a(k) e^{i[k(z - v_g t) + (k_0 v_g - \omega_0)t]} dk \quad (4.3)$$

$$= \frac{e^{i(k_0 v_g - \omega_0)t}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} a(k) e^{ik(z - v_g t)} dk \quad (4.4)$$

$$= E(z - v_g t, 0) e^{i(k_0 v_g - \omega_0)t}. \quad (4.5)$$

It can be seen, that the form of the wave packet at time t is simply the wave packet at an earlier time $t = 0$ after propagating a distance $v_g t$ [2].

The group velocity can be generalized to the three-dimensional case as

$$\mathbf{v}_g = \frac{\partial \omega}{\partial \mathbf{k}} = \hat{\mathbf{e}}_x \frac{\partial \omega}{\partial k_x} + \hat{\mathbf{e}}_y \frac{\partial \omega}{\partial k_y} + \hat{\mathbf{e}}_z \frac{\partial \omega}{\partial k_z}. \quad (4.6)$$

In the next section, we will see that the group velocity (4.6) is one of the ODEs used to determine the ray path in ray tracing. The second implication of the above discussion is that it is (in most practical cases) enough to trace the carrier frequency to determine the ray path.

4.2 Geometrical optics in slowly varying media

Ray theory is the application of EM wave propagation to slowly varying media, also known as the WKB (Wentzel-Kramers-Brillouin) approximation in mathematical physics. By slowly varying, it is understood that the wavelength of the traced wave

is much smaller than the spatial scale of inhomogeneities in the medium. Moreover, absorption in the HF and lower frequency bands mainly stems from electron collisions with neutrals in the D layer, with absorption per path length increasing with lower frequency. The ray formalism presented in this thesis requires that absorption is small enough to cause only a small imaginary component of n . As such, geometrical optics and ray tracing are then most suitable for waves of high frequency.

The electromagnetic components of the radio wave have been heretofore treated as monochromatic sinusoidal plane waves of the form

$$f = \tilde{f}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \quad (4.7)$$

where f could be any component of \mathbf{E} or \mathbf{H}_1 . Let us write the waves in the form

$$f = \tilde{f}e^{i\Psi}. \quad (4.8)$$

As the wave propagates, both the amplitude \tilde{f} and phase Ψ may take complicated functions of position and time. The phase Ψ is known as the *eikonal* [12], and is generally not given by a simple expression as in (4.7). However, it is worth noting that it is a large quantity; as the wave propagates a distance of one wavelength, the eikonal increases by 2π , bearing in mind that geometrical optics signifies the limit $\lambda \rightarrow 0$. Over a small displacement, the eikonal can be expressed as a first-order Taylor series expansion around some reference position ($\mathbf{r} = \mathbf{0}$, $t = 0$) as

$$\Psi = \Psi_0 + \mathbf{r} \cdot \frac{\partial\Psi}{\partial\mathbf{r}} + t \frac{\partial\Psi}{\partial t}. \quad (4.9)$$

Within this small displacement, the plane wave appears as one of the form (4.7) given that the wavelength is very small. The derivatives in (4.9) are then immediately identified as

$$\mathbf{k} \equiv \frac{\partial\Psi}{\partial\mathbf{r}}, \quad \omega \equiv -\frac{\partial\Psi}{\partial t}. \quad (4.10)$$

It is from here that a fascinating analogy between particle dynamics and wave propagation is discovered. Hamiltonian mechanics formulates the equations of motion for the canonical coordinates (\mathbf{q}, \mathbf{p}) using the particle Hamiltonian $\mathcal{H}(\mathbf{p}, \mathbf{q}, t)$ as [4]

$$\dot{\mathbf{q}} = \frac{\partial\mathcal{H}}{\partial\mathbf{p}}, \quad (\text{evolution of canonical position}) \quad (4.11)$$

$$\dot{\mathbf{p}} = -\frac{\partial\mathcal{H}}{\partial\mathbf{q}}. \quad (\text{evolution of canonical momentum}) \quad (4.12)$$

The Hamiltonian \mathcal{H} usually corresponds to the total energy of the particle, and is formally defined by the Hamilton-Jacobi equation [4]

$$\mathcal{H} = -\frac{\partial S}{\partial t}. \quad (4.13)$$

The quantity S is known as the action of the particle. Comparing (4.10) to (4.13) suggests, that the eikonal Ψ is analogous to the action S in particle dynamics, and by

extension, that the ray frequency $\omega(\mathbf{k}, \mathbf{r}, t)$ is analogous to the particle Hamiltonian $\mathcal{H}(\mathbf{p}, \mathbf{q}, t)$. The wave vector \mathbf{k} and position \mathbf{r} then give a pair of canonical coordinates corresponding to \mathbf{p} and \mathbf{q} respectively. By this analogy, the equations of motion (4.11)-(4.12) yield the corresponding ones for the ray as [12]

$$\dot{\mathbf{r}} = \frac{\partial \omega}{\partial \mathbf{k}}, \quad (4.14)$$

$$\dot{\mathbf{k}} = -\frac{\partial \omega}{\partial \mathbf{r}}. \quad (4.15)$$

We have already encountered the ray equation (4.14) earlier as the group velocity of a wave packet, lending credibility to the above Hamiltonian treatment of geometrical optics. Much like equations of motion that describe particle trajectories, the above equations are to be integrated in time to solve for the ray path. For a thorough derivation of (4.14)-(4.15) from first principles, the reader is referred to [7].

4.3 Ray equations

From the dispersion relation (3.47) or refractive index (3.51) it is, at least in principle, possible to solve for ω for given coordinates $(\mathbf{k}, \mathbf{r}, t)$ and perform the differentiations in eqns (4.14)-(4.15). However, it is well known that this is not necessary; given any implicit relation $H(\omega, \mathbf{k}, \mathbf{r}, t) = \text{constant}$ to be satisfied along the ray path, the eqns. (4.14)-(4.15) can be recast into a more useful alternative form. The total differential of H is given by

$$dH = \frac{\partial H}{\partial \omega} d\omega + \frac{\partial H}{\partial \mathbf{k}} \cdot d\mathbf{k} + \frac{\partial H}{\partial \mathbf{r}} \cdot d\mathbf{r} + \frac{\partial H}{\partial t} dt = 0, \quad (4.16)$$

and the total differential of ω as

$$d\omega = \frac{\partial \omega}{\partial \mathbf{k}} \cdot d\mathbf{k} + \frac{\partial \omega}{\partial \mathbf{r}} \cdot d\mathbf{r} + \frac{\partial \omega}{\partial t} dt. \quad (4.17)$$

Substitution of (4.17) into (4.16) gives

$$dH = \left(\frac{\partial H}{\partial \omega} \frac{\partial \omega}{\partial \mathbf{k}} + \frac{\partial H}{\partial \mathbf{k}} \right) \cdot d\mathbf{k} + \left(\frac{\partial H}{\partial \omega} \frac{\partial \omega}{\partial \mathbf{r}} + \frac{\partial H}{\partial \mathbf{r}} \right) \cdot d\mathbf{r} + \left(\frac{\partial H}{\partial \omega} \frac{\partial \omega}{\partial t} + \frac{\partial H}{\partial t} \right) dt = 0. \quad (4.18)$$

Requiring that $H = \text{constant}$ on the ray path, the total differential (4.18) has to be zero for all increments $d\mathbf{k}$, $d\mathbf{r}$ and dt , so the corresponding prefactors vanish identically. Requiring that the prefactors of $d\mathbf{k}$ and $d\mathbf{r}$ be zero, one can show that

$$\frac{\partial \omega}{\partial \mathbf{k}} = -\frac{\partial H}{\partial \mathbf{k}} \bigg/ \frac{\partial H}{\partial \omega}, \quad \frac{\partial \omega}{\partial \mathbf{r}} = -\frac{\partial H}{\partial \mathbf{r}} \bigg/ \frac{\partial H}{\partial \omega}. \quad (4.19)$$

Inserting (4.19) into (4.14)-(4.15), one obtains the well-known ray equations [23]

$$\dot{\mathbf{r}} = -\frac{\partial H}{\partial \mathbf{k}} \bigg/ \frac{\partial H}{\partial \omega}, \quad (4.20)$$

$$\dot{\mathbf{k}} = \frac{\partial H}{\partial \mathbf{r}} \bigg/ \frac{\partial H}{\partial \omega}. \quad (4.21)$$

The ray equations (4.20)-(4.21) give six coupled first-order ODEs that may be integrated in time to solve the ray path, and are used in this form by our ray tracer. The function $H(\omega, \mathbf{k}, \mathbf{r}, t)$ is often called the ray Hamiltonian [26], and will be referred to as such in this thesis. The particular choice of Hamiltonian is not unique; some H demonstrate superior numerical properties, while some H may be rendered physically invalid in certain plasma regions. For a versatile ray tracer, usually no single H is sufficient, and the program is required to switch between Hamiltonians in the middle of the path. These difficulties are studied in detail in section 4.4.

From eqns. (4.16)-(4.17), an additional interesting phenomenon can be identified. Dividing equation (4.17) by dt one obtains

$$\dot{\omega} = \frac{\partial \omega}{\partial \mathbf{k}} \cdot \dot{\mathbf{k}} + \frac{\partial \omega}{\partial \mathbf{r}} \cdot \dot{\mathbf{r}} + \frac{\partial \omega}{\partial t}, \quad (4.22)$$

into which we insert the ray equations (4.14)-(4.15) for $\dot{\mathbf{k}}$ and $\dot{\mathbf{r}}$. The result is

$$\dot{\omega} = \left[\frac{\partial \omega}{\partial \mathbf{k}} \cdot \frac{\partial \omega}{\partial \mathbf{r}} - \frac{\partial \omega}{\partial \mathbf{r}} \cdot \frac{\partial \omega}{\partial \mathbf{k}} \right] + \frac{\partial \omega}{\partial t} = \frac{\partial \omega}{\partial t}. \quad (4.23)$$

Furthermore, requiring that the prefactor of dt in the differential (4.18) vanishes identically, one can show that

$$\frac{\partial \omega}{\partial t} = - \frac{\partial H}{\partial t} \bigg/ \frac{\partial H}{\partial \omega}. \quad (4.24)$$

As a result, even the frequency may change as the ray propagates: The frequency shift of the ray can then be computed by integrating the equation

$$\dot{\omega} = - \frac{\partial H}{\partial t} \bigg/ \frac{\partial H}{\partial \omega}. \quad (4.25)$$

In static media, the Hamiltonian has no explicit time dependence, and as a result, $\dot{\omega} = 0$. Although the ionosphere is dynamic in nature due to natural causes (e.g. solar flares, diurnal variation etc.), it is argued in [26] that the frequency shifts are less than one part in 10^5 due to these slow variations, resulting in negligible frequency shifts to have an effect on propagation. The ray tracer developed for this thesis does not track the frequency shift of the ray. Nevertheless, we have chosen to include necessary formulae for evaluating $\dot{\omega}$, assuming that the electron density of the ionosphere may demonstrate time variation.

4.4 Choosing the Hamiltonian

The Hamiltonian $H(\omega, \mathbf{k}, \mathbf{r}, t) = \text{constant}$ physically signifies a condition to be satisfied along the ray path. In section 3, we have already encountered two such suitable conditions in the forms of dispersion relations: The wave equation determinant condition, also known as the Booker quartic (3.47), and the Appleton-Hartree refractive index $n^2 = (ck/w)^2$ (3.51). Knowing the analytic expression of the Hamiltonian,

the partial derivatives in (4.20)-(4.21) are evaluated as functions of $(\omega, \mathbf{k}, \mathbf{r}, t)$. The resulting expressions are rather lengthy, and are given in appendix A for the Hamiltonians used in this thesis.

In the theory of EM wave propagation in a collisional plasma, we saw that the dispersion relations are complex. The ray equations (4.20)-(4.21) would then formally result in complex \mathbf{r} and \mathbf{k} . Allowing the coordinates to take complex values is known as ray tracing in complex space [3], and is theoretically more intricate than the real-space ray formalism used in this thesis. For HF (and some MF) frequencies, the dispersion relation is mostly real when using collision frequencies encountered in the ionosphere, and it is then desirable to describe ray paths with real coordinates \mathbf{r} and \mathbf{k} . To this end, we adopt the method used by Jones and Stephenson [26] to only require the real part of the dispersion relation to be satisfied. The effects of collisions are then mainly to attenuate the signal.

The particular choices of Hamiltonians demonstrate their own computational strengths and weaknesses, and no single Hamiltonian is universally applicable for the developed ray tracer. Hence, the program should maintain diagnostics to switch between Hamiltonians when necessary. Depending on the choice of the ionosphere (e.g. isotropic or anisotropic), the ray tracer uses one or two Hamiltonians to integrate the ray path. The used Hamiltonians and their advantages are discussed in the following section. For further discussion, the interested reader is referred to [26] and [23].

4.4.1 Appleton-Hartree Hamiltonian

From the definition of the refractive index $n^2 = (ck/\omega)^2$, the Appleton-Hartree Hamiltonian used in this thesis is defined as

$$H = \text{Re} \left[\frac{c^2}{\omega^2} (k_x^2 + k_y^2 + k_z^2) - n^2 \right]. \quad (4.26)$$

The Appleton-Hartree refractive index (3.51) is used for n^2 , as well as S_R (3.52) to select the mode (O/X) of the ray. The Cartesian wave vector components (k_x, k_y, k_z) as well as ω are all used as real-valued quantities.

The adaptive Runge-Kutta integrator used to solve the path requires evaluation of n^2 not only on the ray position \mathbf{r} , but at points near the path. For near vertical incidence skywaves (NVIS), the refractive index has to be evaluated in an evanescent region near reflection. Unlike $\text{Re}[n]$, the function $\text{Re}[n^2]$ is well-behaved in or at the boundary of an evanescent region, providing more reliable computational performance. However, (4.26) will not work when passing through a Spitzer or coupling points since n^2 is indeterminate at these points. In anisotropic plasmas, we have decided to use the Hamiltonian (4.26) in regions where $X \leq 0.1$. In isotropic plasmas, no Spitzes nor coupling points are encountered, and the Hamiltonian (4.26) is expected to function without difficulty for all $X > 0$.

Depending on whether the geomagnetic field or collisions are included in ray tracing as options, the program provides four versions of the Appleton-Hartree Hamiltonian with the corresponding derivatives being documented in appendix A.1:

1. AHFWFC (Appleton-Hartree with field, with collisions)
2. AHWFNC (Appleton-Hartree with field, no collisions)
3. AHNFWC (Appleton-Hartree no field, with collisions)
4. AHNFNC (Appleton-Hartree no field, no collisions)

4.4.2 Booker quartic Hamiltonian

From the determinant condition of the cold plasma wave equation, the Booker quartic Hamiltonian is defined as [26]

$$\begin{aligned}
 H = \text{Re} \Big[& [U^2(U - X) - UY^2] c^4 k^4 + X(\mathbf{k} \cdot \mathbf{Y})^2 c^4 k^2 \\
 & + [Y^2(2U - X) - 2U(U - X)^2] c^2 k^2 \omega^2 - X(\mathbf{k} \cdot \mathbf{Y})^2 c^2 \omega^2 \\
 & + (U - X) [(U - X)^2 - Y^2] \omega^4 \Big]. \tag{4.27}
 \end{aligned}$$

As usual, \mathbf{k} and ω are taken as real quantities, with $k^2 = k_x^2 + k_y^2 + k_z^2$. The Hamiltonian (4.27) can be shown to work only in so-called non-degenerate regions of the plasma [23]: In degenerate regions, two or three of the eigenvalues of \mathbf{M} (3.41) vanish, and as a result, all of the derivatives of (4.27) are zero. Common examples of degenerate plasma regions include free space ($X = 0$) and isotropic plasmas ($Y = 0$).

In an anisotropic plasma, the ray tracer starts by using the Hamiltonian (4.26) until $X > 0.1$, at which point the Hamiltonian is switched to (4.27). If the ray later reaches a region where $X \leq 0.1$, the Hamiltonian is switched back to (4.26). The Booker quartic Hamiltonian cannot distinguish between the modes of the ray, and because of this, will work in Spitzes and coupling points where (4.26) fails [26]. Conversely, in isotropic plasmas (4.27) will not work, and (4.26) is always used as the Hamiltonian.

Since the Booker quartic Hamiltonian is only used in anisotropic plasmas, the program provides only two versions of (4.27) with derivatives in appendix A.2:

1. BQFWFC (Booker quartic with field, with collisions)
2. BQWFNC (Booker quartic with field, no collisions)

4.5 Adaptive numerical integration

The path of the ray is solved by numerically integrating the ray equations (4.20)-(4.21) in time using the Cash-Karp method [22]. The use of adaptive stepsize integrators is a well-established modern practice for various kinds of computational problems, and is justified for ionospheric ray tracing where the right-hand side of the ray equations varies considerably; Near NVIS reflection, very small steps are needed to tiptoe through regions where \mathbf{k} is turned around, whereas at the bottom of the ionosphere larger steps are sufficient when the ray direction changes more

slowly. Moreover, the right hand side of (4.20)-(4.21) can change rapidly if the ray is about to encounter a resonance, which suggests the use of integrators suitable for non-smooth problems.

The Cash-Karp method is in the family of Runge-Kutta ODE solvers, and has proved useful for solving some non-smooth and stiff problems. Similar to other adaptive stepsize integrators, the leading idea of the method is to compare the results of a fifth-order solution to an embedded fourth-order solution as an error estimate [14]. Let us define the vector \mathbf{y} as

$$\mathbf{y} = \begin{bmatrix} \mathbf{r} \\ \mathbf{k} \end{bmatrix} \in \mathcal{R}^6, \quad (4.28)$$

and the initial-value problem to be solved as

$$\dot{\mathbf{y}} = \begin{bmatrix} -\frac{\partial H}{\partial \mathbf{k}} / \frac{\partial H}{\partial \omega} \\ \frac{\partial H}{\partial \mathbf{r}} / \frac{\partial H}{\partial \omega} \end{bmatrix} \equiv \mathbf{f}(t, \mathbf{y}) \in \mathcal{R}^6, \quad \mathbf{y}(t_0) = \mathbf{y}_0. \quad (4.29)$$

The general explicit Runge-Kutta formula of fifth-order requires six function evaluations per step, given the stepsize Δt_n of step n as

$$\begin{aligned} \mathbf{s}_1 &= \Delta t_n \mathbf{f}(t_n, \mathbf{y}_n) \\ \mathbf{s}_2 &= \Delta t_n \mathbf{f}(t_n + a_2 \Delta t_n, \mathbf{y}_n + b_{21} \mathbf{s}_1) \\ &\vdots \\ \mathbf{s}_6 &= \Delta t_n \mathbf{f}(t_n + a_6 \Delta t_n, \mathbf{y}_n + b_{61} \mathbf{s}_1 + \cdots + b_{65} \mathbf{s}_5) \\ \mathbf{y}_{n+1} &= \mathbf{y}_n + \sum_{i=1}^6 c_i \mathbf{s}_i + \mathcal{O}((\Delta t_n)^6) \\ t_{n+1} &= t_n + \Delta t_n. \end{aligned} \quad (4.30)$$

The embedded fourth-order formula is

$$\mathbf{y}_{n+1}^* = \mathbf{y}_n + \sum_{i=1}^6 c_i^* \mathbf{s}_i + \mathcal{O}((\Delta t_n)^5). \quad (4.31)$$

The error of the fifth-order solution is then taken as the difference of the two:

$$\Delta \equiv \mathbf{y}_{n+1} - \mathbf{y}_{n+1}^* = \sum_{i=1}^6 (c_i - c_i^*) \mathbf{s}_i. \quad (4.32)$$

The constants of the Cash-Karp method are tabled in figure 10.

Cash-Karp Parameters for Embedded Runge-Kutta Method								
i	a_i	b_{ij}					c_i	c_i^*
1							$\frac{37}{378}$	$\frac{2825}{27648}$
2	$\frac{1}{5}$	$\frac{1}{5}$					0	0
3	$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$				$\frac{250}{621}$	$\frac{18575}{48384}$
4	$\frac{3}{5}$	$\frac{3}{10}$	$-\frac{9}{10}$	$\frac{6}{5}$			$\frac{125}{594}$	$\frac{13525}{55296}$
5	1	$-\frac{11}{54}$	$\frac{5}{2}$	$-\frac{70}{27}$	$\frac{35}{27}$		0	$\frac{277}{14336}$
6	$\frac{7}{8}$	$\frac{1631}{55296}$	$\frac{175}{512}$	$\frac{575}{13824}$	$\frac{44275}{110592}$	$\frac{253}{4096}$	$\frac{512}{1771}$	$\frac{1}{4}$
$j =$		1	2	3	4	5		

Figure 10: Tabled constants of the Cash-Karp method [14].

From the error estimate (4.32), the stepsize is adjusted using a predefined error tolerance. The method used in our ray tracer is paraphrased from [14]: We require that for each of the six components i of Δ , the error is smaller than a scale S_i :

$$|\Delta_i| \leq S_i = \epsilon_{atol} + \max(|\mathbf{y}_{n+1,i}|, |\mathbf{y}_{n,i}|)\epsilon_{rtol}, \quad (4.33)$$

where ϵ_{atol} and ϵ_{rtol} are respectively the absolute and relative tolerance. Because the moduli of \mathbf{r} and \mathbf{k} differ by several orders of magnitude, the tolerances should be adjusted separately for the integration of the position and the wave vector, denoted by the superscripts r and k respectively. We have found, that $\epsilon_{atol}^r = \epsilon_{rtol}^r = \epsilon_{atol}^k = \epsilon_{rtol}^k = 10^{-8}$ usually yields satisfactory convergence. The accuracy can be increased by decreasing the tolerances at the expense of longer computation time.

The stepsize is rescaled to the needs of the "worst-offender" using the Euclidean norm

$$err_n = \sqrt{\frac{1}{6} \sum_{i=1}^6 \left(\frac{\Delta_i}{S_i} \right)^2}. \quad (4.34)$$

If $err_n \leq 1$, the error is sufficiently small and the step is accepted. Otherwise, the step is rejected, and integration of the step is attempted again using a decreased stepsize. By default, the stepsize of the next step is adjusted using the formula

$$\Delta t_{n+1} = safe \times \Delta t_n \left(\frac{1}{err_n} \right)^\alpha, \quad (4.35)$$

where $safe \sim 0.9$ is a safety factor slightly less than unity and $\alpha = 0.2$. Furthermore, experience has shown that it is wise to not allow Δt_n to increase nor decrease too much during a step. In our work, we limit Δt_n to an increase by a factor of 10 and to a decrease by a factor of 5 per step. Moreover, the stepsize is not allowed to increase at all if the previous step was rejected.

When testing our ray tracer, we have concluded that it is reasonable to limit the maximum stepsize to e.g. $\Delta t_n = 10^{-6}$ s for reliability. The ionospheric models used to generate the background in our ray tracer typically evaluate altitude profiles with a resolution of circa one kilometer. A ray propagating with the maximum group velocity $|\dot{\mathbf{r}}| = c$ traverses in one step a distance less than 300 m if $\Delta t_n \leq 10^{-6}$ s.

4.6 Antenna radiation patterns

For tracing radio rays emanating from man-made sources, knowledge of the transmitting antenna type is required to determine the polarization and signal amplitude in a certain transmission direction. Real antennas are not isotropic: the polarization and radiation intensity are sensitive to the transmission direction, and is commonly visualized by a radiation pattern. A typical antenna radiation pattern is illustrated in figure 11. In this section, we briefly cover the basic theory of antennas to discuss radiation patterns, as well as present four antenna types as examples to implement in our ray tracer. For the antennas used as transmitters in the ray tracer, we are required to document at least three antenna-specific properties:

1. The geometry necessary to determine the orientation of the antenna in three dimensions. For instance, knowing the orientation of the wire axis of a dipole antenna is enough to determine the radiation pattern unambiguously.
2. The radiated electric field $\mathbf{E}(r, \theta, \phi)$, which is in turn used to determine the polarization of a wave transmitted in a particular direction.
3. The radiation intensity pattern $U(\theta, \phi)$, which physically represents the (average) power radiated in a certain direction per unit solid angle. This knowledge is needed to estimate the initial signal amplitude of the ray.

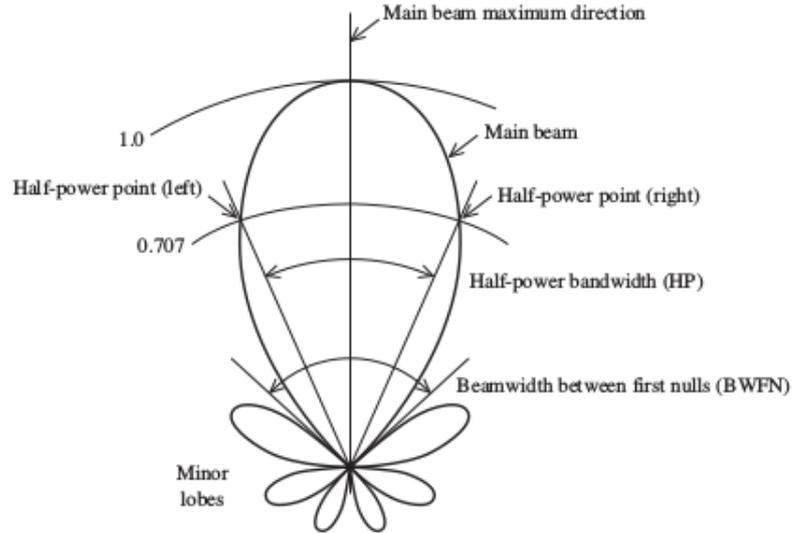


Figure 11: Typical radiation pattern as a polar plot, showing the field magnitude relative to the maximum of the main beam [19]

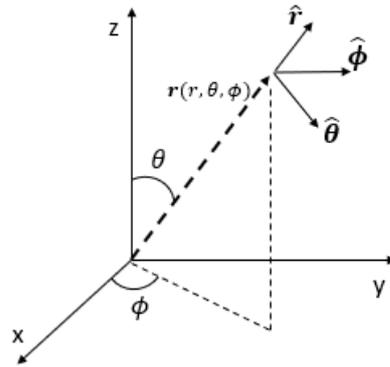


Figure 12: Spherical coordinates with the local orthonormal unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$. The polar angle θ in this figure is not to be confused with the angle θ earlier defined as the angle between \mathbf{k} and \mathbf{B}_0 .

It is convenient to express the radiated fields in spherical coordinates with the antenna placed at the origin $r = 0$. The spherical coordinate system as well as the local orthonormal unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ are illustrated in figure 12. These vectors are equivalently expressed using the fixed Cartesian unit vectors by

$$\begin{cases} \hat{\mathbf{r}} = \sin \theta \cos \phi \hat{\mathbf{x}} + \sin \theta \sin \phi \hat{\mathbf{y}} + \cos \theta \hat{\mathbf{z}}, \\ \hat{\boldsymbol{\theta}} = \cos \theta \cos \phi \hat{\mathbf{x}} + \cos \theta \sin \phi \hat{\mathbf{y}} - \sin \theta \hat{\mathbf{z}}, \\ \hat{\boldsymbol{\phi}} = -\sin \phi \hat{\mathbf{x}} + \cos \phi \hat{\mathbf{y}}. \end{cases} \quad (4.36)$$

For simplicity, let us consider a z -directed source. The electric field radiated from such an element has only a component in the direction of $\hat{\boldsymbol{\theta}}$, and demonstrates a

maximal field amplitude $E_{\theta,max}$ in some direction θ for some constant r much larger than λ or the antenna size d [19]. The approximations $r \gg \lambda$ and $r \gg d$ tell us that the fields are evaluated in the far-field region, and as a result, the fields \mathbf{E} and \mathbf{H}_1 decrease inversely with the distance. The normalized field pattern is then defined as

$$F(\theta) = \frac{E_\theta}{E_{\theta,max}}. \quad (4.37)$$

For a more general radiator, the normalized field pattern may contain an azimuthal dependence and is written $F(\theta, \phi)$. Generally speaking, the normalized field pattern may be complex-valued.

The result (3.102) shows that the radiated intensity is directly proportional to the square of the electric field magnitude, and consequently, that the intensity decreases inversely with the square of the distance r [19]. While it was seen in section 3 that the radiated energy flux is quantified by the Poynting vector $\mathbf{\Pi}(r, \theta, \phi)$, it is more common to indicate the power radiated in a certain direction by the so-called radiation intensity pattern $U(\theta, \phi) = |\mathbf{\Pi}_{av}(r, \theta, \phi)| r^2$. This quantity has the dimension of power per steradian, and physically represents the average power radiated in a certain direction per solid angle. The connection between the radiation intensity and normalized field pattern is simple [19]:

$$U(\theta, \phi) = U_m |F(\theta, \phi)|^2, \quad (4.38)$$

where U_m is the maximum power radiated in a certain direction. Given that the antenna radiates the total power P_{rad} , the factor U_m can be found by integrating $U(\theta, \phi)$ over all solid angles Ω :

$$P_{rad} = \int_{all \ \Omega} U(\theta, \phi) d\Omega = U_m \int_{all \ \Omega} |F(\theta, \phi)|^2 d\Omega \quad (4.39)$$

$$= U_m \int_{\theta=0}^{\theta=\pi} \int_{\phi=0}^{\phi=2\pi} |F(\theta, \phi)|^2 \sin \theta d\theta d\phi. \quad (4.40)$$

As an example, consider the theoretical isotropic radiator $F(\theta, \phi) = 1$. Performing the integrals in (4.40), one finds that for an isotropic radiator

$$P_{rad} = 4\pi U_m \Leftrightarrow U_m = \frac{P_{rad}}{4\pi} \Rightarrow U(\theta, \phi) = \frac{P_{rad}}{4\pi}.$$

Since the polarization and intensity of transmitted rays depend on the antenna type, it might be of interest to study the effects of different antennas on ray tracing. To this end, it was decided to include in addition to the isotropic antenna three common and simple antenna types, namely the ideal dipole, half-wave dipole and rhombic antenna. The characteristics and required inputs of these antenna types are documented below.

4.6.1 Isotropic antenna

The isotropic antenna is a hypothetical yet useful concept, as it can be used to model a transmitter that radiates equally in all directions. The isotropic antenna is

characterized by

$$F(\theta, \phi) = 1, \quad U(\theta, \phi) = \frac{P_{rad}}{4\pi}.$$

However, determining the polarization of the radiated field necessitates some caveats. A truly isotropic radiator would need to be invariant for all translations in θ and ϕ , so loosely speaking, the radiated electric field could have no $\hat{\theta}$ nor $\hat{\phi}$ components as such a polarization would prefer some spatial direction where no such asymmetry is present.

The developed ray tracing program contains a small number of subroutines that take the wave polarization as input, such as magnetoionic splitting and ground reflections presented in section 4.7. Since the isotropic antenna cannot exhibit a well-defined polarization, the developed ray tracing program should handle this type of radiator as an exception. For our ray tracing program, we have chosen the following:

- At magnetoionic splitting, the incident wave from an isotropic antenna is split by default into an O- and X-wave of equal amplitudes.
- At ground reflection, the reflection coefficients are computed using a circular polarization $\rho_{\perp} = +i$.

4.6.2 Ideal dipole

The ideal dipole is understood as an antenna consisting of two wire conductor elements of length $\Delta z \ll \lambda$ as in figure 13. For such an antenna, the current is uniform in both magnitude and phase over both elements, and the radiated fields can be solved analytically from Maxwell's equations [19]. In the far-field region in a vacuum, the radiated fields read

$$\mathbf{E} = \frac{I\Delta z}{4\pi} i\omega\mu_0 \frac{e^{i(kr-\omega t)}}{r} \sin\theta \hat{\theta}, \quad (4.41)$$

$$\mathbf{H}_1 = \frac{I\Delta z}{4\pi} ik \frac{e^{i(kr-\omega t)}}{r} \sin\theta \hat{\phi}. \quad (4.42)$$

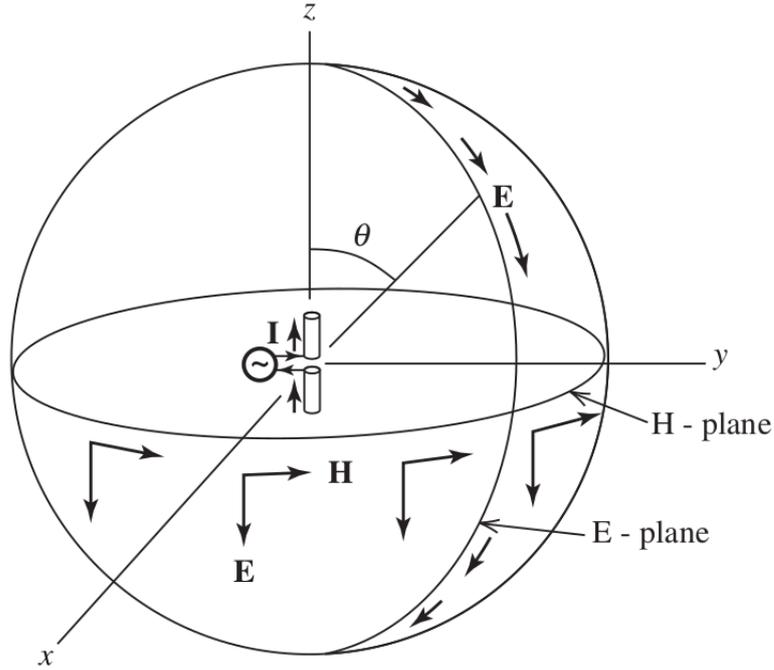


Figure 13: Illustration of the dipole antenna. The two wire elements of equal length are oriented parallel to the z -axis, connected to a feedline at $z = 0$ [19].

From (4.41) it can be seen, that the electric field is always polarized in the $\hat{\theta}$ -direction, and thus ultimately depends on the orientation of the wire axis. Furthermore, the dipole does not radiate at all in the direction of the wire axis, i.e when $\sin \theta = 0$. Equation (4.41) also reveals, that the normalized field pattern is [19]

$$F(\theta, \phi) = \sin \theta, \quad (4.43)$$

from which the radiation intensity pattern is evaluated

$$U(\theta, \phi) = U_m \sin^2 \theta. \quad (4.44)$$

Calculating the integral (4.40) for (4.44), the maximum intensity U_m is shown to be equal to $\frac{3P_{rad}}{8\pi}$, which results in the radiation intensity pattern

$$U(\theta, \phi) = \frac{3P_{rad}}{8\pi} \sin^2 \theta. \quad (4.45)$$

The field radiation pattern of the ideal dipole antenna is shown in figure 14. It is worth noting, that the dipole radiation pattern is azimuthally symmetric, and as a result, the pattern is symmetric about the z -axis with zero radiation parallel to the axis. Such an antenna is said to be omnidirectional.

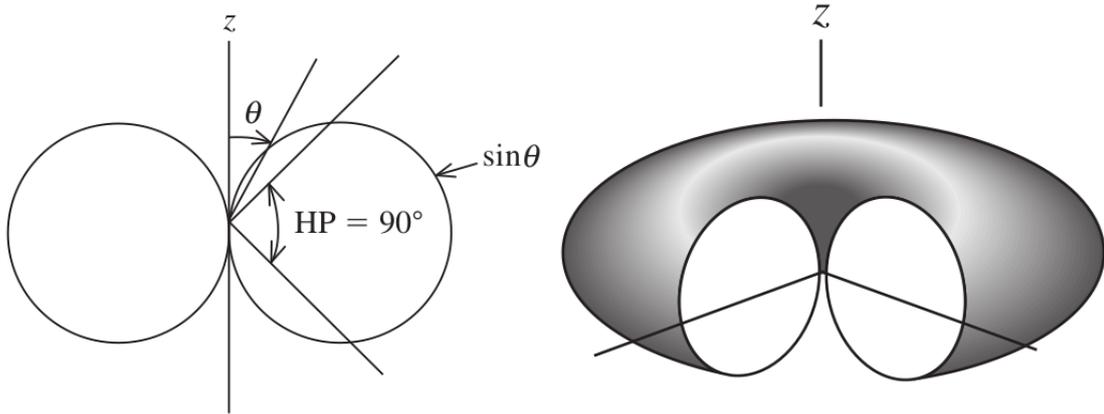


Figure 14: Field radiation patterns of the ideal dipole antenna as a polar plot (left) and in three dimensions (right). HP represents the half-power beamwidth [19].

4.6.3 Half-wave dipole

The half-wave dipole is characterized by its two wire conductors measured to one quarter of a wavelength each, making the total antenna length approximately half a wavelength. By tuning the wire ends, a zero input reactance can be achieved, and the current in the wires becomes that of a standing wave. This proves to be optimal for matching the input power to radiation, as the input impedance is then purely resistive and the radiation efficiency is close to 100% [19]. However, MF/HF half-wave dipole antennas need to be rather large; the Suomi 100 AM radio receiver is tuned to 1-10 MHz, so half-wave dipole antennas operating at these frequencies need to be 15-150 meters long.

The field radiated by a half-wave dipole bears many similarities to that of the ideal dipole. Once more, we orient the wire axis parallel to the z -axis. In the far field, the radiated electric field reads [19]

$$\mathbf{E} = i \frac{I_0 \omega \mu_0}{2\pi k r} \frac{\cos\left(\frac{\pi}{2} \cos \theta\right)}{\sin \theta} e^{(kr - \omega t)} \hat{\boldsymbol{\theta}}, \quad (4.46)$$

where I_0 is the peak current in the wire conductors. The normalized field pattern is then

$$F(\theta, \phi) = \frac{\cos\left(\frac{\pi}{2} \cos \theta\right)}{\sin \theta}, \quad (4.47)$$

and the radiation intensity pattern is given by

$$U(\theta, \phi) = U_m \frac{\cos^2\left(\frac{\pi}{2} \cos \theta\right)}{\sin^2 \theta}. \quad (4.48)$$

Numerical integration of (4.48) over all solid angles (equation (4.40)) gives

$$P_{rad} = U_m \int_{\theta=0}^{\theta=\pi} \int_{\phi=0}^{\phi=2\pi} \frac{\cos^2\left(\frac{\pi}{2} \cos \theta\right)}{\sin^2 \theta} \sin \theta d\theta d\phi \approx 7.6581 U_m. \quad (4.49)$$

$$\Rightarrow U(\theta, \phi) \approx \frac{P_{rad}}{7.6581} \frac{\cos^2\left(\frac{\pi}{2} \cos \theta\right)}{\sin^2 \theta}. \quad (4.50)$$

4.6.4 Ionosonde rhombic antenna

Ionosondes are specialized radars used widely in ionosphere research for bottomside sounding, and as such, it is of interest to include some representative ionosonde antenna type in this section. As a case example, the Sodankylä Geophysical Observatory in Northern Finland employs a rhombic antenna titled Alpha-Wolf [45]: It is raised in a 64-meter mast above the ground, and sweeps 0.5-16 MHz frequencies at a nominal input power of 5 W. The radiation pattern of a typical rhombic antenna is presented in figure 15. The main beam of the rhombic antenna is, of course, directed towards the sky, so the end with the terminating resistor is pointing skyward.

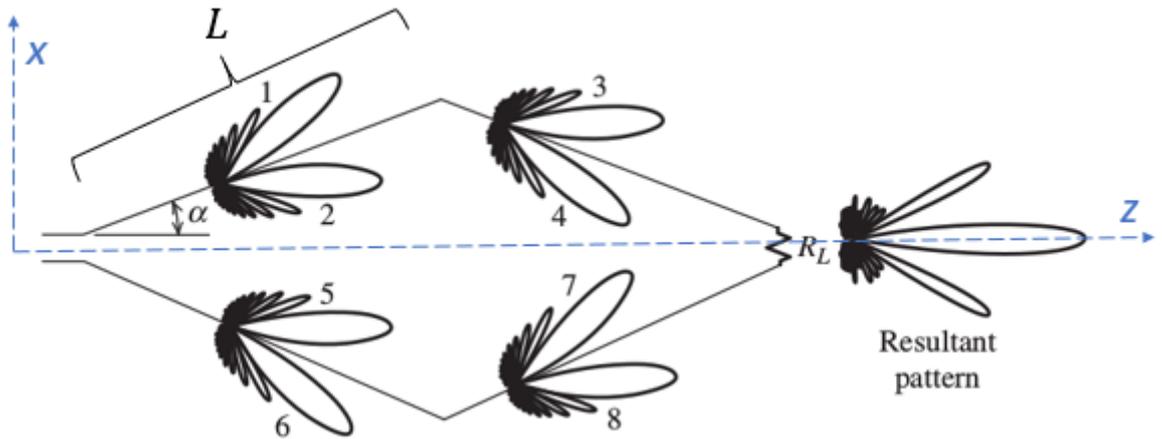


Figure 15: The rhombic antenna with resulting field radiation pattern. In this case, $L = 6\lambda$ and $\alpha = 16^\circ$. The terminating resistor used to prevent a reflected current is denoted as R_L [19]. The orthogonal axes z and x are drawn as dashed lines to demonstrate how the orientation of the rhombic antenna is defined in the ray tracer: For example, if z is oriented to point skyward and x to the west, the main beam is directed towards the sky with the broad side of the rhombus facing north/south.

Figure 15 illustrates, that a well-designed rhombic antenna is characterized by a main lobe in the desired direction of radiation with only small side lobes in other directions. As such, ionosondes employing rhombic antennas tend to orient the main beam towards the sky to result in a maximal vertical sounding intensity. Another conspicuous feature of the rhombic antenna is the terminating resistor R_L at the far end of the rhombus. The radiation efficiency of the antenna is significantly decreased as a result of input power being absorbed in the load – however, this loss of power would have manifested as a sizeable back lobe caused by a reflected current if no terminating resistor were present.

The total radiation pattern of the rhombic antenna can be reconstructed as a superposition of four traveling-wave antennas as illustrated in figure 15. In the

xz-plane, the total field pattern $F(\theta)$ is given by [19]:

$$F_1 = e^{i\gamma[\cos(\theta-\alpha)-1]} \sin(\theta - \alpha) \frac{\sin[\gamma(1 - \cos(\theta - \alpha))]}{\gamma(1 - \cos(\theta - \alpha))}, \quad (4.51)$$

$$F_2 = e^{i\gamma[\cos(\theta+\alpha)-1]} \sin(\theta + \alpha) \frac{\sin[\gamma(1 - \cos(\theta + \alpha))]}{\gamma(1 - \cos(\theta + \alpha))}, \quad (4.52)$$

$$F_3 = e^{2i\gamma \cos(\theta-\alpha)} F_2, \quad (4.53)$$

$$F_4 = e^{2i\gamma \cos(\theta+\alpha)} F_1, \quad (4.54)$$

$$F(\theta) = K_R [F_1 - F_2 + e^{-2i\gamma} (F_3 - F_4)]. \quad (4.55)$$

Here $\gamma \equiv \pi L/\lambda$, and K_R is a normalization constant that depends on the length L . Contrary to the dipole antennas discussed earlier, the rhombic antenna is not azimuthally symmetric, and the result (4.55) then is valid only in the xz-plane. However, the ray tracer developed in this thesis assumes that the result (4.55) is also valid outside of the xz-plane, or in other words,

$$F(\theta, \phi) \simeq F(\theta) = K_R [F_1 - F_2 + e^{-2i\gamma} (F_3 - F_4)]. \quad (4.56)$$

Moreover, the polarization of the field radiated by the rhombic antenna is assumed to be in the direction of $\hat{\theta}$ for all (θ, ϕ) . These assumptions are justified on the grounds that the resulting radiation pattern still exhibits the main desired characteristics of a rhombic antenna, namely that

1. the radiation pattern demonstrates a powerful main beam that allows the radiated power to be focused in a narrow desired direction, and
2. the polarization is linear.

As usual, the radiation intensity pattern is given by

$$U(\theta, \phi) = U_m |F(\theta, \phi)|^2. \quad (4.57)$$

In this thesis, the coefficients U_m and K_R have not been evaluated for a rhombic antenna of arbitrary length L .

4.7 Magnetoionic splitting and ground reflections

4.7.1 The limiting region of the ionosphere

Downwardly traveling rays emerging from the ionosphere exhibit a polarization ρ_\perp corresponding to one of the magnetoionic modes. At altitudes less than roughly 60 km above the ground, the electron density is of negligible magnitude, allowing altitudes lower than this to be treated as a vacuum. Here, $X = 0$, yet the effective collision frequency ν_{en} is rather large and varying with altitude h to result in $Z(h) > 0$. The polarization given by eqn. (3.82) then suggests, that the polarization of the downward ray is given by the expression

$$\rho_\perp(h) = i \frac{\left(\frac{1}{2} Y^2 \sin^2 \theta \pm \left[\frac{1}{4} Y^4 \sin^4 \theta + Y^2 U^2 \cos^2 \theta \right]^{1/2} \right)}{Y U \cos \theta}, \quad (4.58)$$

where $\rho_{\perp}(h)$ has been written to remind us that the polarization is essentially a function of altitude. However, at these altitudes the medium should act as a vacuum where the polarization of the downgoing ray cannot change with altitude.

To resolve this contradiction, a full-wave treatment is required, and doing so provides the following explanation [3]: The downward wave is constructively interfered by waves of the other characteristic mode generated by mode conversion. The interference is shown to be just enough for the wave traveling downwards to retain a constant polarization with which it emerges from the ionosphere. The polarization the wave emerges with is referred to as the *limiting polarization*, and can be shown to be determined at an altitude where X is in the range $0.01 - 0.0001$. Depending on wave frequency, this altitude occurs roughly in the range of $h = 60 - 80$ km. The altitude of limiting polarization is then known as the limiting region.

Conversely, a wave traveling upwards transmitted from the ground will initially exhibit a polarization determined by the transmitting antenna. When the wave reaches the limiting region, it is split into the characteristic O and X mode, both of which proceed to propagate independently from one another. This effect of magnetoionic splitting will be studied in more detail in the following section. The main takeaway here is that the limiting region at $h = 60 - 80$ km fixes the practical lower boundary of the ionosphere, and it is here that rays obtain a polarization corresponding to one of the magnetoionic modes they proceed to propagate with. Once excited to a polarization of one of the magnetoionic modes, the rays inside the ionosphere retain a polarization characteristic of their mode, except for some special cases at coupling points ($X = 1, Z = Z_t$). Hence, the ray tracer developed for this thesis assumes that rays of a given mode are described by a single polarization ρ_{\perp} inside the ionosphere.

4.7.2 Magnetoionic splitting formulae

Knowing the type and orientation of the transmitting antenna, it is in principle possible to determine the signal amplitude \mathcal{A} and polarization ρ_{\perp} of a ray the moment it enters the ionosphere. As noted above, the limiting region presents a valid boundary to denote as the entry point of the ionosphere in terms of radio propagation. Here, the wave is split into the characteristic O and X modes, which both propagate independently from one another. This is illustrated in figure 16, where both ray paths are solved from (4.20)-(4.21) using the corresponding branch of the dispersion relation. Let us denote the transmitted wave incident on the limiting region by the subscript I . The ray transmitted from the ground, exhibiting a single amplitude \mathcal{A}_I and polarization $\rho_{\perp,I}$, has been split into two rays with different polarizations and amplitudes.

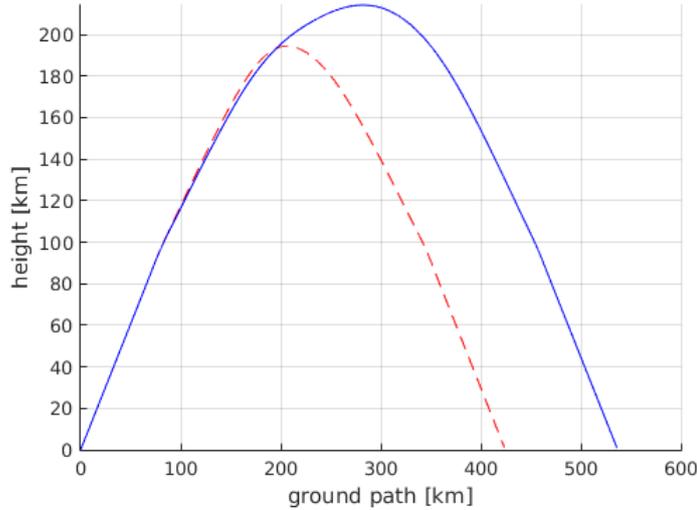


Figure 16: Ray tracing results for a single 5 MHz ray, with the O and X modes graphed as the blue solid line and red dashed line respectively. The two modes propagate independently after being split from the launched ray at a height of 60 km above the ground.

How is the signal amplitude of the incident wave then distributed between the magnetoionic modes? We denote the relevant quantities of the ordinary and extraordinary mode by subscripts O and E respectively. When the wave enters the ionosphere from the ground, $n_O \approx n_E \approx 1$, and consequently $\rho_{\parallel} \approx 0$. Using the result (3.102) with $\boldsymbol{\mu} \cdot \tilde{\mathbf{E}} = 0$ and $n \approx \mu \approx 1$, the signal amplitudes of the respective waves are given by

$$\mathcal{A} = \frac{1}{2Z_0} (1 + |\rho_{\perp}|^2) \left| \tilde{E}_x \right|^2. \quad (4.59)$$

In [30] a method of evaluating the ratio of magnetoionic mode amplitude to incident amplitude is presented using the tilt angle Ψ (eqn. 3.95) and axis ratio M (eqn. 3.96). Given the polarization $\rho_{\perp,I}$ and $\rho_{\perp,O/E}$ of the incident and magnetoionic mode respectively, the tilt angles Ψ_I and $\Psi_{O/E}$ as well as axis ratios M_I and $M_{O/E}$ are evaluated from eqns. (3.94)-(3.96). The intensity transmission coefficients, defined as the ratio of magnetoionic mode signal amplitude to incident signal amplitude, are given using the following formula presented in [30]:

$$T_{O/E} = \frac{\mathcal{A}_{O/E}}{\mathcal{A}_I} = \frac{(1 + M_I M_{O/E})^2 \cos^2(\Psi_I - \Psi_{O/E}) + (M_I + M_{O/E})^2 \sin^2(\Psi_I - \Psi_{O/E})}{(1 + M_I^2)(1 + M_{O/E}^2)}. \quad (4.60)$$

The intensity transmission coefficients (4.60) thus depend on the incident polarization as well as the ionospheric parameters at the point where the ray enters the ionosphere in the limiting region. If the incident polarization is chosen to correspond perfectly to one of the magnetoionic modes, only that mode is excited in the ionosphere with $T = 1$ for one mode and $T \approx 0$ for the other. This can be shown as follows: If the incident polarization is perfectly matched to one magnetoionic mode

– say, the ordinary mode, so that $\rho_{\perp,O} = \rho_{\perp,I}$ – then $M_I = M_O$ and $\Psi_I = \Psi_O$. The reader can then easily verify that $T_O = 1$ from eqn. (4.60). On the other hand, based on the discussion in section 3.8.4 and eqn. (3.96), it then must follow for the extraordinary mode that $M_E = -M_I$ and $\Psi_E = \pi/2 - \Psi_I$. As a consequence, $(M_I + M_E) = 0$, and the sine term in the numerator of (4.60) is zero identically for the extraordinary mode. However, the cosine term in (4.60) is not identically zero for the extraordinary mode; Since $\Psi_I - \Psi_E = 2\Psi_I - \pi/2$ for an incident polarization matched to the ordinary mode, it follows that $\cos^2(\Psi_I - \Psi_E) = \sin^2(2\Psi_I)$. Consequently, the numerator of eqn. (4.60) is $(1 - M_I^2) \sin^2(2\Psi_I)$, which is not necessarily zero. Hence, $T_E > 0$ in this case. The disturbing consequence of this result is that sometimes $T_O + T_E > 1$ when using equation (4.60); it appears, that energy has been created at magnetoionic splitting, a result that seems to violate the conservation of energy. If, however, the incident wave is matched to a circularly polarized ordinary mode, then the axis ratio is $M_I = \pm 1$ and $(1 + M_I M_E) = 0$, resulting in $T_E = 0$, and more importantly, $T_O + T_E = 1$.

In practice, however, the transmitted ray polarization is typically linearly polarized for common antenna types [19], and the amplitudes of the two modes are nearly equal [3]. This is because the high collision frequency in the limiting region causes Z to be typically larger than unity while $X \approx 0$, and as a result, $\rho_{\perp,O/E} \approx \pm i$ (cf. equation (4.58)). Consequently, both magnetoionic modes are circularly polarized and $M_{O/E} = \pm 1$ while $M_I = 0$ for any incident real-valued (i.e. linear) polarization $\rho_{\perp,I}$. With these values, it can be seen that eqn. (4.60) takes the value $T_{O/E} = 0.5$, that is, the incident wave is divided into two magnetoionic modes of equal amplitude. The polarization phase angles φ (eqn. (3.93)) of both magnetoionic modes are presented in figure 17 for typical values in the limiting region.

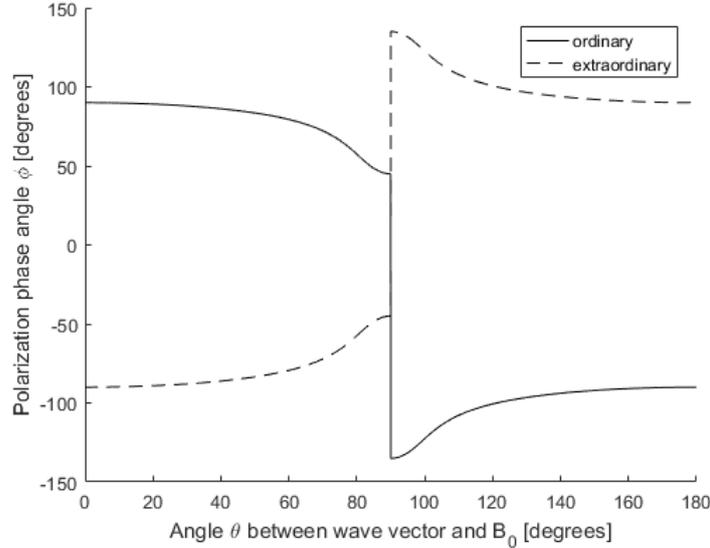


Figure 17: Limiting polarization (see section 4.7.1) of magnetoionic modes as a function of the angle θ between \mathbf{k} and \mathbf{B}_0 , using $X = 0.001$, $Y = 0.5$ and $Z = 1$. The limiting polarization is nearly circular for most angles θ , as evidenced by the polarization phase angle being close to $\pm 90^\circ$. A notable exception occurs at $\theta = 90^\circ$, at which point both the ordinary and extraordinary mode is linearly polarized [3].

The results presented here leave room for further investigations. An LEO satellite transmitting HF waves from 500 km altitude is practically transmitting from inside the ionosphere, and the treatment above is no longer valid. Moreover, a multi-hop skywave undergoing sequential ground reflections after emerging from the ionosphere can be interpreted to give rise to multiple magnetoionic splitting events each time the wave is incident on the ionosphere from below. This way, a single transmitted ray can theoretically give rise to $2^{(m+1)}$ rays after m ground reflections. The implications for ray tracing certainly leave room for judgement. Should the ray tracer follow just one of the modes after splitting for computational efficiency, assuming that the transmission coefficient of one mode $T \approx 1$ for a circularly polarized wave reflected from the ground? But what if the transmission coefficients are close to each other? In the current version of the ray tracer, the program selects the mode with highest transmission coefficient after ground reflection, discarding the other (weaker) mode. The consequence of this strategy is that the computation time is reduced, whilst a small fraction of the transmission power is lost with the discarded mode. This design choice is further discussed in section 7.

4.7.3 Signal loss and polarization shift at ground reflection

Once a downwardly traveling ray emerges from the limiting region of the ionosphere, it exhibits a near-circular polarization corresponding to one of the magnetoionic modes in the limiting region (cf. section 4.7.2). As the ray hits the surface of the Earth, it is reflected and continues to propagate skyward once more. Success-

ive reflections from the ionosphere and the ground can allow ray paths to extend up to thousands of kilometers from the transmitter – a well-known fact utilized by radio operators for long-distance communication for almost a century now [17]. Unfortunately, depending on the electrical properties of the reflecting surface, each ground reflection causes a loss in signal strength and a change in polarization. For this reason, the ray tracer needs to be fitted with physics to study reflections from conducting surfaces.

The surface of the Earth is not only a dielectric, but somewhat electrically conductive as well at MF/HF frequencies. Saline sea water exhibits high conductivity and permittivity to result in a reflection loss that is near to zero, while dry soil, on the other hand, is poorly conductive and less electrically permissive, resulting in a higher reflection loss. In the frequency range of 1-10 MHz, the conductivity σ and relative permittivity ε_r of common ground materials are largely frequency-independent, and are presented in table 1 for a selected set of materials.

Table 1: Electrical characteristics of the surface of the Earth for frequencies between 1-10 MHz, tabled for sea water of average salinity, wet soil and medium dry soil [25]

	Sea water	Wet soil	Dry soil
Conductivity σ [S/m]	5	0.01	0.001
Relative permittivity ε_r	70	30	15

The nonzero conductivity σ allows the wave electric field to drive a current $\mathbf{j}_f = \sigma \mathbf{E}$ in the conducting surface, as per Ohm's law [13]. As a result, the refractive index in the conductor is complex, accounting for the absorption in the medium. Taking the cross product $\mathbf{k} \times$ of equation (3.61), and inserting (3.62) into the right-hand side gives

$$i\mathbf{k} \times (\mathbf{k} \times \mathbf{E}) = i(\mathbf{k}(\mathbf{k} \cdot \mathbf{E}) - k^2 \mathbf{E}) = \omega \mu_0 \mathbf{j}_f - i\omega^2 \mu_0 \mathbf{D}. \quad (4.61)$$

Assuming the medium to be linear, it holds that $\mathbf{D} = \varepsilon_r \varepsilon_0 \mathbf{E}$. Restricting ourselves to the steady-state solution without transients, $\rho_f = 0$ and eqn. (3.59) gives $\mathbf{k} \cdot \mathbf{E} = 0$. Moreover, using Ohm's law $\mathbf{j}_f = \sigma \mathbf{E}$, the wave equation (4.61) becomes

$$\left(k^2 - \frac{\omega^2 \varepsilon_r}{c^2} - i \frac{\omega \sigma}{\varepsilon_0 c^2} \right) \mathbf{E} = \mathbf{0}. \quad (4.62)$$

Like in section 3.5, the nontrivial solutions of the wave equation (4.62) are found by solving the zeroes of the expression in the parentheses. Defining the refractive index of the surface n_g by $k^2 = (n_g \omega / c)^2$, we find that [13]

$$n_g^2 = \varepsilon_r + i \frac{\sigma}{\omega \varepsilon_0}. \quad (4.63)$$

When a downwardly traveling ray is incident on the ground at an angle of incidence θ_i , the ray is split into a reflected and refracted ray, both of which continue propagation in the plane of incidence. This situation is illustrated in figure 18. The

direction of the refracted ray is acquired from the refracted angle θ_t and Snell's law [13]

$$n_1 \sin \theta_i = n_2 \sin \theta_t. \quad (4.64)$$

When the ray incident on the ground is propagating in air, $n_1 \approx 1$. However, $n_2 = n_g$ is complex, and subsequently Snell's law yields a complex angle θ_t for the refracted ray. A formal complex solution to Snell's law is obtained as follows:

$$\sin \theta_t = \frac{\sin \theta_i}{n_g}, \quad \cos \theta_t = \sqrt{1 - \sin^2 \theta_t} = \sqrt{1 - \left(\frac{\sin \theta_i}{n_g}\right)^2}. \quad (4.65)$$

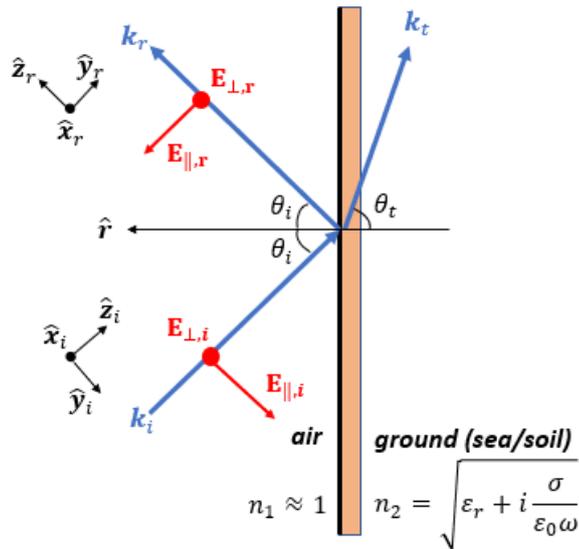


Figure 18: Oblique incidence of plane waves at the air/ground interface. The incident ray (subscript i) emerging from the ionosphere is reflected (subscript r) at the angle of incidence θ_i . The coordinate systems used in determining the polarizations of both waves are illustrated next to the rays. The vector $\hat{\mathbf{r}}$ (eqn. (2.9)) denotes the surface normal. The plane of incidence is defined to contain the vectors $\hat{\mathbf{r}}$ and \mathbf{k}_i .

The incident ray of generally elliptical polarization is split into electric fields that are perpendicular and parallel to the plane of incidence as $\tilde{\mathbf{E}}_i = \tilde{E}_{\perp,i} \hat{\mathbf{x}}_i + \tilde{E}_{\parallel,i} \hat{\mathbf{y}}_i = \tilde{E}_{\perp,i} (\hat{\mathbf{x}}_i + \rho_i \hat{\mathbf{y}}_i)$. The limiting polarization ρ_{\perp} has been converted to the coordinate system in figure 18 using the methods in section 3.8.3 to give a polarization ρ_i in the plane of incidence. The ratio of reflected to incident field needs to be solved separately for the perpendicular and parallel components, commonly referred to as the TE (transverse electric) and TM (transverse magnetic) modes. Using the coordinate convention in figure 18, the solution is given by the Fresnel coefficients

[13]:

$$r_{\perp} = \frac{\tilde{E}_{\perp,r}}{\tilde{E}_{\perp,i}} = \frac{\cos \theta_i - n_g \cos \theta_t}{\cos \theta_i + n_g \cos \theta_t}, \quad (4.66)$$

$$r_{\parallel} = \frac{\tilde{E}_{\parallel,r}}{\tilde{E}_{\parallel,i}} = \frac{\cos \theta_t - n_g \cos \theta_i}{\cos \theta_t + n_g \cos \theta_i}. \quad (4.67)$$

Using the basis vectors for the reflected ray in figure 18, the electric field of the reflected ray is given by $\tilde{\mathbf{E}}_r = \tilde{E}_{\perp,r} \hat{\mathbf{x}}_r - \tilde{E}_{\parallel,r} \hat{\mathbf{y}}_r = \tilde{E}_{\perp,r} (\hat{\mathbf{x}}_r + \rho_r \hat{\mathbf{y}}_r)$. Using the Fresnel coefficients (4.66)-(4.67), the polarization of the reflected ray in the system $(\hat{\mathbf{x}}_r, \hat{\mathbf{y}}_r, \hat{\mathbf{z}}_r)$ is then given by

$$\rho_r = \frac{-\tilde{E}_{\parallel,r}}{\tilde{E}_{\perp,r}} = \rho_i \left(-\frac{r_{\parallel}}{r_{\perp}} \right). \quad (4.68)$$

Furthermore, using the result (3.101) for the amplitudes of the incident and reflected rays, the ratio of reflected intensity to incident intensity is

$$R = \frac{|\tilde{E}_{\perp,r}|^2 + |\tilde{E}_{\parallel,r}|^2}{|\tilde{E}_{\perp,i}|^2 + |\tilde{E}_{\parallel,i}|^2} = \frac{|r_{\perp}|^2 + |\rho_i r_{\parallel}|^2}{1 + |\rho_i|^2}. \quad (4.69)$$

One should note, that the expression (4.69) gives the intensity reflection coefficient for the entire incident ray of polarization ρ_i , or in other words, contains the intensities reflected by both the TE and TM modes. In the singular case $\rho_i \rightarrow \pm\infty$, i.e. if only the TM mode is present in the incident ray, $R = |r_{\parallel}|^2$.

Observing the result (4.68), one can see that the polarization of the ray is shifted at reflection. In fact, ground reflection usually results in the handedness of polarization to be reversed: Sea water, for instance, being a good conductor reflects the ray almost perfectly for most incidence angles, and consequently, $r_{\parallel} = r_{\perp} \approx -1$ so that $\rho_r = -\rho_i$. An initially left-handed circular polarization $\rho_i = +i$ is thus reflected as a right-handed circular polarization $\rho_r = -i$ and vice versa. In the general case of a non-ideal reflector, the ellipticity of the polarization is shifted as well. As a direct result of this reversal of polarization handedness, the reflected wave traveling upwards may couple to a different mode than the initial downward wave. As an example, imagine that a wave traveling downwards emerges with a left-handed circular polarization corresponding to the O-mode of the limiting region. When the wave is reflected from the ground, the handedness tends to be reversed, resulting in an upwardly traveling wave with a right-handed circular polarization. Depending on the orientation of \mathbf{B}_0 in the limiting region where this upward ray enters the ionosphere, the X-mode may be right-handedly polarized, and the ray couples mostly to the X-mode. Quantitatively, the transmission coefficient T (4.60) is then greater for the X-mode than for the O-mode.

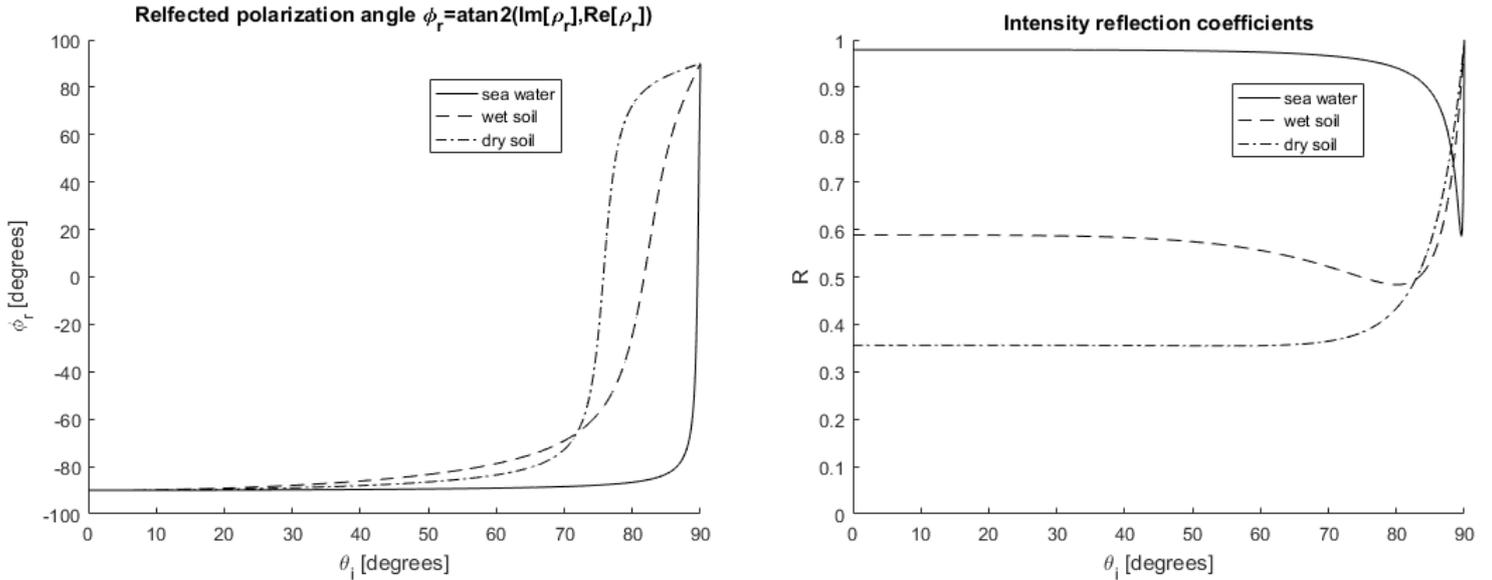


Figure 19: Values of the reflected polarization phase angle ϕ_r (eqn. (3.93)) and reflection coefficient R (4.69) computed for different angles of incidence θ_i , using $f = 5$ MHz and an incident left-handed circular polarization $\rho_i = +i$. The results are computed for common surface materials, namely saline sea water, wet soil, and dry soil.

The values of ρ_r (4.68) and R (4.69) have been computed and illustrated in figure 19 for different angles of incidence using $\rho_i = +i$ and $f = 5$ MHz. A number of observations can be made from these results. First of all, the polarization handedness is indeed reversed for most incidence angles θ_i , evidenced by the polarization phase angle φ (eqn. (3.93)) to be near -90° after reflection. For the dielectric surface materials, i.e. wet and dry soil, the phase angle deviates more visibly from -90° , indicating that the ellipticity of the polarization is changed while still undergoing handedness reversal for most θ_i . However, at an angle known as Brewster's angle θ_B , $\varphi = 0$ and the reflected ray emerges as perfectly linearly polarized along the Earth's surface [13]. For angles of incidence $\theta_i > \theta_B$, the handedness is not reversed, while the polarization is still elliptical. Secondly, the computed values of R corroborate the earlier assertion that sea water is a good reflector: Except for θ_i near Brewster's angle, the reflection coefficient $R \approx 1$, resulting in minimal signal strength loss. However, when reflected from wet or dry ground, the signal strength loss is much greater ($R \approx 0.5$), except for near-grazing angles.

4.8 Signal attenuation

The signal amplitude \mathcal{A} of a radio transmission may change for two reasons:

1. Geometric attenuation: Neighboring rays may diverge or converge, so the energy flux either decreases or increases. In a vacuum, this corresponds to the free space path loss of \mathcal{A} that obeys the inverse square law. For the sake of brevity, we shall refer to this effect as geometric attenuation.

2. Absorption attenuation: The signal amplitude may be attenuated by absorption stemming from collisions in the ionosphere. If $n = \mu + i\chi$, then the value of χ determines the rate of attenuation by absorption.

Computing the geometric attenuation in the ionosphere is generally not as simple as the common inverse square law treatment. This is because adjacent rays are not generally refracted equally to result in a $1/r^2$ -spread of their cross section. The adjacent rays may even converge tangentially to a single point to result in an apparent infinite intensity. Such points are referred to as caustics [12]. A method presented in [34] is to formulate 12 additional equations that are integrated parallel to the six ray equations (4.20-4.21), which in turn determines the signal intensity of the reference ray. The method necessitates rather laborious derivation of the 12 additional differential equations, and in this thesis no attempt has been made to evaluate geometric attenuation in this way. Instead, we devise a method similar to photon mapping in the field of computer graphics for computing global illumination [9]. This method is presented in detail in section 4.8.2.

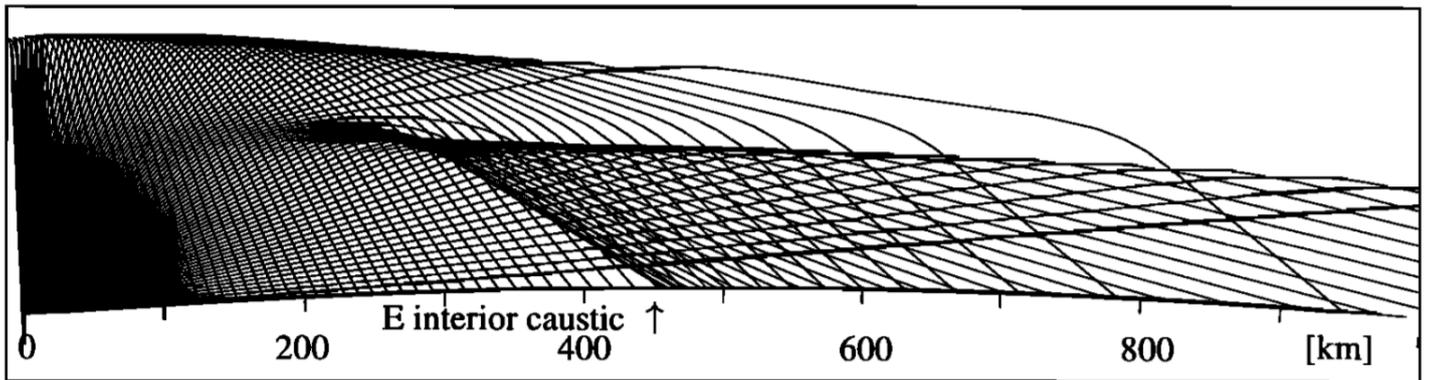


Figure 20: Ray tracing results for 3 MHz O-mode rays launched with various elevation angles [34]. The horizontal axis denotes the range along the Earth's surface. It can be seen that some adjacent rays reflected from the E layer converge to caustics on the ground.

4.8.1 Attenuation by absorption

For the ray tracing method used in this thesis to be valid, the absorption per wavelength has to be small. The plane waves (3.14) take the form

$$g(\mathbf{r}, t) \propto e^{-\frac{\omega}{c}\boldsymbol{\chi}\cdot\mathbf{r}} e^{i(\frac{\omega}{c}\boldsymbol{\mu}\cdot\mathbf{r}-\omega t)}, \quad (4.70)$$

which describes waves that are attenuated by $\omega\chi/c$ per unit path length in the direction of $\boldsymbol{\chi}$. Given that $\chi \ll \mu$, it is assumed that $\boldsymbol{\chi}$ is parallel to the real-valued wave vector $\mathbf{k} \approx \boldsymbol{\mu}\omega/c$. The electric field attenuation along the ray path is then given approximately by [3]

$$\left| \tilde{E} \right| = \left| \tilde{E}_0 \right| \exp \left[- \int_{\mathbf{r}_1}^{\mathbf{r}_2} \frac{\omega}{c} \boldsymbol{\chi} \hat{\mathbf{k}} \cdot d\mathbf{r} \right] = \left| \tilde{E}_0 \right| \exp \left[- \int_{t_1}^{t_2} \frac{\chi}{\mu} \mathbf{k} \cdot \dot{\mathbf{r}} dt \right]. \quad (4.71)$$

From the result (3.102), and assuming that the change in refractive index can be neglected in computing the signal amplitude, we define the attenuation in decibels as A_{dB} :

$$A_{dB} = 10 \log_{10} \left[\frac{\mathcal{A}}{\mathcal{A}_0} \right] \approx 20 \log_{10} \left[\frac{|\tilde{E}|}{|\tilde{E}_0|} \right] = -\frac{20}{\log_e 10} \int_{t_1}^{t_2} \frac{\chi}{\mu} \mathbf{k} \cdot \dot{\mathbf{r}} dt. \quad (4.72)$$

Evidently, the absorption can be evaluated by integration along the ray path. Substituting $\text{Im}[n^2] = 2\mu\chi$, $k^2 = (\mu\omega/c)^2$ and $\dot{\mathbf{r}}$ from (4.20), the attenuation by absorption can be evaluated by integrating the following equation in parallel to (4.20)-(4.21) [26]:

$$\frac{dA_{dB}}{dt} = \frac{10}{\log_e 10} \frac{\omega^2 \text{Im}[n^2]}{c^2 k^2} \frac{(\mathbf{k} \cdot \frac{\partial H}{\partial \mathbf{k}})}{\frac{\partial H}{\partial \omega}}. \quad (4.73)$$

4.8.2 Photon mapping to estimate geometric attenuation

This section is aimed at better understanding how far the skywaves from a certain transmitter can be heard. The signal strength of transmissions typically decreases quickly with an increasing path length, and at some point the signal becomes too weak to be received by conventional systems. This effect is referred to as geometric attenuation in this thesis, and is an attenuation effect separate from absorption. Published methods for directly calculating geometric attenuation for single rays in magnetoplasmas are found to rely on the concept of ray tubes, i.e. the volume defined by a principal ray and its three or four adjacent rays. The method presented in [3] elaborates on rather complex equations for solving the cross section of initially adjacent rays being transmitted in slightly differing initial directions. Similarly, [34] formulates 12 additional differential equations for tracing the paths of the adjacent rays in a ray tube. In this thesis, we attempt something a little different.

The method of photon mapping is a novel method in computer graphics for evaluating global illumination from a light source, and was originally proposed by Henrik Wann Jensen in the 1990s. To put it very briefly, photon mapping divides the power transmitted from a source between a large number of photons launched in random directions, the paths of which are followed by ray tracing. When the path intersects a surface, the strike-point is stored in the photon map, and the photon is subsequently either absorbed, reflected or refracted with some probability. After all photons are traced, the radiance of every surface is constructed. For a proper explanation worthy of reference, the reader is directed to [9]. The ray tracer developed for this thesis does not follow the standard photon mapping method of [9] to complete precision. Instead, we have opted to make some simplifications to provide a rough estimate of geometric attenuation. The work presented here should not be regarded as being a complete description of geometric attenuation, and ample room for improvement is left for later iterations.

In photon mapping mode, the ray tracer launches a large number of N photons in random directions. For example, the analysis in section 6 uses $N = 10^5$ photons. The launch directions, taken as the initial unit wave vectors of the rays, are randomly

generated on the unit sphere by generating two pseudorandom numbers u and $v \in [0, 1]$ from a continuous uniform distribution. The unit wave vector $\hat{\mathbf{k}}_{rand}$ is then given by [47]

$$\phi_{rand} = 2\pi u, \quad (4.74)$$

$$\theta_{rand} = \arccos(2v - 1), \quad (4.75)$$

$$\hat{\mathbf{k}}_{rand} = \begin{bmatrix} \sin \theta_{rand} \cos \phi_{rand} \\ \sin \theta_{rand} \sin \phi_{rand} \\ \cos \theta_{rand} \end{bmatrix}. \quad (4.76)$$

The polarization of the launched photon is evaluated from the antenna polarization, and the non-normalized power of the photon is computed from the radiation intensity pattern $U(\theta, \phi)$ of the antenna by setting the normalization factor U_m to unity. Repeating the process N times, one obtains a vector of N photons with non-normalized initial powers $\mathbf{U} = (U_1, U_2, \dots, U_N) \in \mathcal{R}^N$. The values U_k reproduce the radiation properties of the antenna, with photons in favorable directions demonstrating a higher power than those in unfavorable directions (e.g. along or near the axis of a dipole antenna). Finally, the initial powers U_k are properly normalized to the values U'_k by evaluating the sum of all U_k :

$$P_{sum} = \sum_{k=1}^N U_k \Rightarrow U'_k = \frac{U_k}{P_{sum}}. \quad (4.77)$$

The values U'_k , physically representing the power radiated per unit solid angle in a given direction from a radiator with total power $P_{rad} = 1$ W, is then taken as the initial signal amplitude \mathcal{A} of the photon ray. Then, the initial decibel attenuation of the photon ray is taken as $A_{dB}(0) = 10 \log_{10} U'_k$. All of the N photons are then launched in respective random directions and traced through the ionosphere. If traced in an anisotropic plasma, each photon is launched twice, with one being initially split into an O-wave and the other into an X-wave in the limiting region. When the photons are traced, their signal amplitude is attenuated by either absorption (eqn. (4.73)), ground reflections, or magnetoionic splitting.

The photon rays are traced until one of three stop conditions is met:

1. The ray altitude may exceed a predefined maximum altitude (e.g. 600 km),
2. the maximum number of ground reflections may be reached, or
3. the maximum number of integration steps (e.g. 10^6) may be exceeded.

These stop conditions are in addition to the exceptions of non-pseudoreal rays, evanescent regions and resonances, which cause a ray to be stopped prematurely. The detailed strategy of exception handling is described in section 5.4. Since we are realistically only interested in the signal amplitude at the ground or at the satellite altitude, the photon state is saved and written to a file if an intersection with the ground or some predefined target altitude is detected. The output file then

constitutes the photon map. Each line of the photon map stores the coordinates \mathbf{r}_k of intersection, the group velocity $\hat{\mathbf{r}}_k$, and the signal attenuation $A_{dB,k}$ in decibels.

After all photons have been traced and the photon map is constructed, the map is postprocessed to estimate the resulting intensity on the target surface (i.e. the ground or at the satellite altitude). The spherical target surface is divided into $n_\lambda \times n_\phi$ mesh elements m_{ij} with $i = (0, 1, 2, \dots, n_\lambda - 1)$ and $j = (0, 1, 2, \dots, n_\phi - 1)$. Defining the boundaries of the mesh by limiting the respective geographical latitude and longitude in the range $\lambda \in [\lambda_{min}, \lambda_{max}]$ and $\phi \in [\phi_{min}, \phi_{max}]$ in radians, the mesh elements can be (for example) defined by the linearly spaced coordinate intervals as follows:

$$\begin{aligned} \Delta\lambda &= \frac{\lambda_{max} - \lambda_{min}}{n_\lambda}, & \Delta\phi &= \frac{\phi_{max} - \phi_{min}}{n_\phi} \\ m_{ij} &= \{ \lambda_{min} + i\Delta\lambda \leq \lambda < \lambda_{min} + (i+1)\Delta\lambda, \\ & \phi_{min} + j\Delta\phi \leq \phi < \phi_{min} + (j+1)\Delta\phi \} \\ & i = (0, 1, 2, \dots, n_\lambda - 1), \quad j = (0, 1, 2, \dots, n_\phi - 1). \end{aligned} \quad (4.78)$$

A problem with the meshing (4.78) is that the element areas are not equal. Given that the target surface is at a radius r from the center of the Earth, the area S_{ij} of element m_{ij} is

$$\begin{aligned} S_{ij} &= r^2 \int_{\phi_{min} + j\Delta\phi}^{\phi_{min} + (j+1)\Delta\phi} \int_{\lambda_{min} + i\Delta\lambda}^{\lambda_{min} + (i+1)\Delta\lambda} \cos \lambda d\lambda d\phi \\ &= r^2 \Delta\phi [\sin(\lambda_{min} + (i+1)\Delta\lambda) - \sin(\lambda_{min} + i\Delta\lambda)] \\ &\approx r^2 \Delta\phi \Delta\lambda \cos(\lambda_{min} + i\Delta\lambda). \end{aligned} \quad (4.79)$$

In deriving (4.79), the first-order approximation $\sin(x + \Delta x) \approx \sin x + \cos x \Delta x$ has been used for small Δx . Once the mesh is constructed, the program searches the photon map and assigns each intersection to one mesh element at most. Let us denote the intersection positions of the photons as $\mathbf{r}_k \in \{m_{ij}\}$ for some (i, j) . Given the total radiated power P_{rad} in Watts and the decibel attenuation $A_{dB,k}$ of photon k stored in the photon map, the power P_{ij} incident on element m_{ij} is incremented as the omnidirectional flux

$$P_{ij} \rightarrow P_{ij} + 10^{A_{db,k}/10} P_{rad}. \quad (\text{omnidirectional flux}) \quad (4.80)$$

Another possible increment is that of the vector flux, given the unit group velocity $\hat{\mathbf{v}}_k$ and the outward unit normal $\hat{\mathbf{n}}$ of the target surface:

$$P_{ij} \rightarrow P_{ij} + 10^{A_{db,k}/10} P_{rad} (\hat{\mathbf{v}}_k \cdot \hat{\mathbf{n}}). \quad (\text{vector flux}) \quad (4.81)$$

Finally, the intensity I_{ij} incident on element m_{ij} is estimated by dividing the incident power P_{ij} by the element area S_{ij} :

$$I_{ij} = \frac{P_{ij}}{S_{ij}}. \quad (4.82)$$

For practical purposes, it might be advisable to limit the launch elevation of photons to within a certain range. For instance, performing photon mapping for an antenna directly above the ground, around half of the launched photons are immediately incident on the ground while the other half propagate towards the sky. Physically, the rays here incident on the ground are in the near field of the transmitter, invalidating some of our formulae for $U(\theta, \phi)$ which are usually evaluated in the far field. Moreover, rays launched with very low elevation angles would propagate to the sky unobstructed on a perfectly flat sphere, but not in the real world where buildings, trees and other terrain features would diffract the waves to cause them to propagate as ground waves [17].

The procedure presented above certainly necessitates a few caveats. For example, in Henrik Wann Jensen's original method [9], at surface intersection the photons are either reflected, absorbed or refracted without power loss with a certain probability to save computation time. Moreover, in the meshing strategy above to estimate the intensity, a single launched photon can intersect with many elements on the same surface if tracing is allowed to proceed for more than one ground reflection. Furthermore, the polarization and phase of the incident photon rays are not stored in the photon map. This means, that the incident power (4.82) cannot take into account destructive multipath interference. While the method above can be described as rather crude in many aspects, it should only be regarded as an estimate for visualizing the audibility of signals far away from the transmitter. In section 6, we present some tentative results of photon mapping in an anisotropic ionosphere using $N = 10^5$ photons.

4.9 Reversibility and reciprocity of radio waves in the ionosphere

The idea of using ray tracing in data analysis of Suomi 100 radio measurements was originally motivated by the possibility of tracing received signals back to the transmitter on the ground. In other words, if a satellite receives a signal of known carrier frequency from some known direction, can this information be used to deduce where the signal originated from? Here we are already faced with a first caveat: the Suomi 100 MF/HF antennas exhibit a nearly omnidirectional radiation pattern [27], and hence, the instrument is not capable of determining the direction of received signals. In reality, we are required to consider signals coming from all directions, and deduce *a posteriori* which of them are plausible. Whatever the case may be, deducing the nature of transmitted rays from measured received rays begs the question: If a ray transmitted from A is received at B with wave vector \mathbf{k}_B , is the same ray transmitted from B with wave vector $-\mathbf{k}_B$ received at A?

The problem should be separated into two separate aspects of EM wave propagation. The first part concerns the reversibility of the ray paths, or in other words, whether a ray path $\mathbf{r}(t)$ can be traced back to the initial position $\mathbf{r}(0)$ from any later time $t > 0$. The second part concerns that of EM field reciprocity, and is a property of the pair of antennas transmitting through the ionosphere. In fact, the ray paths turn out to be reversible in nearly all practical situations, while the

fields themselves are reciprocal only in special cases when propagating through the anisotropic ionosphere [3]. In what follows, we set to summarize reversibility and reciprocity with practical examples. For a thorough treatment of reversibility and reciprocity of EM fields, the reader is referred to [1].

4.9.1 Physical reversibility of ray paths

Consider a ray of angular frequency ω that begins propagation with initial phase space coordinates $(\mathbf{r}_0, \mathbf{k}_0)$ at time $t = 0$. After some time $T > 0$, the wave vector $\mathbf{k}(T)$ is turned around, and the ray is allowed to propagate from the new initial condition $(\mathbf{r}(T), -\mathbf{k}(T))$ for an additional time T . This is equivalent to a time reversal transformation $t \rightarrow -t$. If the ray traces its path back to the initial position, the ray path is said to be reversible. The implications of time reversal symmetry (or lack thereof) have been studied extensively in numerous fields of physics, and as such, is of interest in a broader perspective than just ray tracing [28].

In ionospheric radio physics, reversibility of the WKB approximation used in deriving the ray equations (4.20)-(4.21) has been thoroughly established in a static ionosphere [3] [32]. This is because in static media the Hamiltonian is symmetric under the transformation $t \rightarrow -t$, or equivalently, $H(\omega, \mathbf{k}, \mathbf{r}, t) = H(\omega, -\mathbf{k}, \mathbf{r}, -t)$. The proof for a general dynamical system being described by Hamiltonian mechanics has been documented in [28], and is adapted for the ray equations in what follows. Consider a ray with initial phase space coordinates $(\mathbf{r}_0, \mathbf{k}_0)$ that is let to propagate forward in time using the familiar ray equations (4.20)-(4.21):

$$\frac{d\mathbf{r}}{dt} = - \frac{\partial H}{\partial \mathbf{k}} \bigg/ \frac{\partial H}{\partial \omega}, \quad (4.83)$$

$$\frac{d\mathbf{k}}{dt} = \frac{\partial H}{\partial \mathbf{r}} \bigg/ \frac{\partial H}{\partial \omega}. \quad (4.84)$$

Then, imagine that the process is allowed to progress in reverse under the transformation $(\mathbf{r}, \mathbf{k}, t) \rightarrow (\mathbf{r}, -\mathbf{k}, -t)$. Under this transformation, the ray equations are transformed into

$$\frac{d\mathbf{r}}{d(-t)} = - \frac{\partial H}{\partial(-\mathbf{k})} \bigg/ \frac{\partial H}{\partial \omega} \Leftrightarrow \frac{d\mathbf{r}}{dt} = - \frac{\partial H}{\partial \mathbf{k}} \bigg/ \frac{\partial H}{\partial \omega}, \quad (4.85)$$

$$\frac{d(-\mathbf{k})}{d(-t)} = \frac{\partial H}{\partial \mathbf{r}} \bigg/ \frac{\partial H}{\partial \omega} \Leftrightarrow \frac{d\mathbf{k}}{dt} = \frac{\partial H}{\partial \mathbf{r}} \bigg/ \frac{\partial H}{\partial \omega}. \quad (4.86)$$

The resulting time-reversed ray equations (4.85)-(4.86) are identical to those in forward time (4.83)-(4.84). This in turn implies, that if $(\mathbf{r}(t), \mathbf{k}(t))$ is a solution to the ray path with initial conditions $(\mathbf{r}_0, \mathbf{k}_0)$, then $(\mathbf{r}(-t), -\mathbf{k}(-t))$ is a solution with initial conditions $(\mathbf{r}_0, -\mathbf{k}_0)$. Ultimately, the ray traces its path back to the transmitter if the wave vector is turned around on any point of the ray path. One should note, that reversibility is valid even if the ionosphere is allowed to be anisotropic and collisional.

However, physical irreversibility of ray paths is possible in cases that the WKB approximation underlying the reversible ray equations is no longer valid. If the

ionosphere varies significantly over one wavelength, a case that would be typical of low radio frequencies, the eikonal Ψ is no longer given by the simple expression (4.9). If so, the eikonal needs to be amended with an additional memory term, which can be shown to introduce physically irreversible ray paths [3] [32]. In [32] it has been shown, that ray irreversibility is apparent only when the wave frequency is lower than the electron gyrofrequency (which is in the order of ~ 1 MHz in the ionosphere). Thus, for the frequency range of 1-10 MHz considered in this thesis, we expect that ray paths solved from (4.20)-(4.21) can be taken to be physically reversible.

Furthermore, the possibility of a time-varying ionosphere introduces ray irreversibility. The cause of such behavior is simple to understand: A ray that is traced in the time interval $0 \leq t < T$ sees a different medium than a reversed ray propagating in the time interval $T \leq t < 2T$. Not only does the ionosphere change during this time, but so does the carrier frequency as given by (4.25). However, in [26] it is argued that the effects of a dynamic ionosphere are practically negligible for ray paths, given that the rays are not traced over a very long timespan where such dynamic changes would become apparent. All things considered, it seems justified to posit that ray paths are reversible in the frequency range of interest.

4.9.2 Irreversibility caused by numerics and model approximations

In the previous section, we established that a hypothetical exact solution to the ray path is reversible when solved from the ray equations (4.20)-(4.21). However, we are advised to keep in mind that the ray tracer solves the path by numerical integration, and as such, the acquired solution may exhibit (unphysical) irreversibility brought on by sole numerics. Moreover, we have found that certain reasonable and common approximations used in evaluating the derivatives of the Hamiltonian introduce a different kind of unwarranted irreversibility.

It is known that Runge-Kutta integrators may introduce a small (but nevertheless existent) irreversibility by each time step [6]. Subsequently, if the ray is reversed at some point, the ray does not return to the exact initial position after numerical integration. The deviation from exact reversibility does, however, decrease with a smaller step size. As an example, the Runge-Kutta fourth-order method has a global truncation error $\mathcal{O}((\Delta t)^4)$ when using a constant step size Δt [6]. In other words, using a numerical integrator with constant stepsize should manifest a solution, which when reversed, asymptotically approaches the initial position as the step size Δt is decreased.

In fact, some types of numerical integrators can be shown to produce *exact* reversibility (not taking into account floating point rounding errors). This class of integrators are commonly known as symplectic integrators [6], and are typically encountered in problems where certain invariants of motion (like energy) are required to be conserved in the numerical solution. Although the usage of symplectic integrators may reproduce reversibility, they are not necessarily any more accurate than e.g. the Cash-Karp method. Hence, we have decided to not use symplectic integrators in favor of more accurate adaptive stepsize integrators, and choose to

accept a small deviation from reversibility.

Finally, we have found that some approximations used in evaluating the derivatives of the Hamiltonian appear to result in a type of irreversibility that persists even when the stepsize Δt is decreased. One such reasonable approximation used in [33] is to approximate $\frac{\partial \mathbf{Y}}{\partial \mathbf{r}} \approx \mathbf{0}$ on the grounds that the spatial variations in the geomagnetic field are small compared to the gradients of the electron density. However, if the value of $\mathbf{Y}(\mathbf{r})$ is re-evaluated each time step on the ray's current position (i.e. the geomagnetic field is nonuniform) while $\frac{\partial \mathbf{Y}}{\partial \mathbf{r}}$ is assumed to be zero anyway, we have found that reversed rays do not return to the transmitter even if the integrator tolerances and stepsizes are decreased. These effects are studied further in section 6, where the results of the developed ray tracer are presented.

The study of reversibility in the acquired numerical solution may serve as a reasonable method of code validation. Any developed ray tracer program should be suspected of harboring a systematic error, if the ray tracer

- (a) produces irreversible paths, and
- (b) reversed paths do not converge closer to the initial condition after a decrease in step size.

Apart from convergence of the solution, approximate reversibility may then be used to fine tune the parameters of the adaptive stepsize controller.

4.9.3 Reciprocity of EM fields

It has now been established, that the ray paths $\mathbf{r}(t)$ are reversible: A signal transmitted from antenna A to antenna B will take the same path back to A, if the wave vector \mathbf{k} is turned around at B. In the fundamental electromagnetic picture, antenna A transmits a wave of polarization $\rho_{\perp,A}$, and as the wave propagates through the ionosphere, the polarization is gradually changed to $\rho_{\perp,B}$ at antenna B. If antenna B then transmits a reversed wave of the exact same received polarization $-\rho_{\perp,B}$ (given in the coordinate system where the z-axis is reversed, i.e. in the direction of the reversed \mathbf{k}), will then the polarization at A be $-\rho_{\perp,A}$?

For waves propagating through the anisotropic ionosphere, the answer is no – it turns out, that the external geomagnetic field \mathbf{B}_0 introduces a rotation of the plane of polarization that is always the same in the sense of the direction of \mathbf{B}_0 , irrespective of the direction of propagation [13]. This nonreciprocal magneto-optic effect is known as Faraday rotation. For us to better understand what is going on, consider a linearly polarized wave propagating parallel to \mathbf{B}_0 in the ionosphere with $X < 1$. As usual, let us denote the ordinary and extraordinary modes by respective subscripts O and E . Much like what was done in section 4.7.2, the linearly polarized wave \mathbf{E}_I can be written as a superposition of two circularly polarized waves which, of course, correspond to the O-mode and X-mode of polarizations $\rho_{\perp,O} = +i$ and $\rho_{\perp,E} = -i$ respectively if $\theta = 0$. In other words,

$$\mathbf{E}_I = \mathbf{E}_O + \mathbf{E}_E \tag{4.87}$$

$$\Leftrightarrow \tilde{\mathbf{E}}_{x,I} \begin{bmatrix} 1 \\ \rho_{\perp}(z) \\ 0 \end{bmatrix} e^{i(kz-\omega t)} = \tilde{\mathbf{E}}_{x,O} \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix} e^{i(k_O z - \omega t)} + \tilde{\mathbf{E}}_{x,E} \begin{bmatrix} 1 \\ -i \\ 0 \end{bmatrix} e^{i(k_E z - \omega t)} \quad (4.88)$$

$$\Leftrightarrow \mathbf{E}_I = \begin{bmatrix} 1 + \left(\frac{\tilde{\mathbf{E}}_{x,E}}{\tilde{\mathbf{E}}_{x,O}} \right) e^{i(k_E - k_O)z} \\ i \left[1 - \left(\frac{\tilde{\mathbf{E}}_{x,E}}{\tilde{\mathbf{E}}_{x,O}} \right) e^{i(k_E - k_O)z} \right] \\ 0 \end{bmatrix} \tilde{\mathbf{E}}_{x,O} e^{i(k_O z - \omega t)}. \quad (4.89)$$

Since $\mathbf{k} \parallel \mathbf{B}_0$, we may without loss of generality choose the coordinate axes so that $\rho_{\perp}(z=0) = 0$. The wave is then initially polarized along the x-axis, and knowing that $\tilde{\mathbf{E}}_{x,E}/\tilde{\mathbf{E}}_{x,O} = 1$ from the discussion in section 4.7.2, the polarization of the composite wave \mathbf{E}_I is solved

$$\rho_{\perp}(z) = \frac{E_{I,y}(z)}{E_{I,x}(z)} = i \frac{1 - e^{i(k_E - k_O)z}}{1 + e^{i(k_E - k_O)z}} = \frac{\sin[(k_E - k_O)z]}{1 + \cos[(k_E - k_O)z]}. \quad (4.90)$$

Because the O- and X-modes propagate at different phase velocities, the polarization of the wave changes. From the result (4.90), it can be seen that $\rho_{\perp}(z)$ is real. The wave polarization then remains linear, but the plane of polarization (or equivalently, the tilt angle Ψ , see eqn. (3.95)) changes as the wave propagates over a distance z . The tilt angle Ψ_I of the wave is then given by

$$\tan \Psi_I(z) = \rho_{\perp}(z) \Leftrightarrow \Psi_I(z) = \arctan(\rho_{\perp}(z)). \quad (4.91)$$

Let us denote the phase difference between the O- and X-mode as $\varphi \equiv (k_E - k_O)z$. The derivative of the tilt angle with respect to travel distance z angle is then

$$\begin{aligned} \frac{d\Psi_I}{dz} &= \frac{1}{1 + \left(\frac{\sin \varphi}{1 + \cos \varphi} \right)^2} \frac{d}{dz} \left(\frac{\sin \varphi}{1 + \cos \varphi} \right) \\ &= \frac{1}{1 + \left(\frac{\sin \varphi}{1 + \cos \varphi} \right)^2} \frac{k_E - k_O}{1 + \cos \varphi} = \frac{k_E - k_O}{2} \\ &= \frac{\omega}{2c} (n_E - n_O). \end{aligned} \quad (4.92)$$

In the case that $\omega \gg \omega_p$, or equivalently $X \ll 1$, the refractive index is approximated by a Taylor series expansion to the first order:

$$n = \left(1 - \frac{X}{1 \pm Y} \right)^{1/2} \approx 1 - \frac{X}{2(1 \pm Y)}, \quad (4.93)$$

and the difference in refractive indices is approximately

$$n_E - n_O \approx -\frac{XY}{1 - Y^2} \approx -XY. \quad (4.94)$$

Using the definitions (3.25)-(3.26), the derivative of the tilt angle is approximately [11]

$$\frac{d\Psi_I}{dz} = \frac{-e^3 N_{e0}}{2m_e^2 \varepsilon_0 c \omega^2} B_0 \equiv V B_0, \quad (4.95)$$

where the coefficient V is called Verdet's constant [13]. The result (4.95) is known as Faraday rotation, and is a first-order magneto-optic effect that causes a non-reciprocal shift in the wave's plane of polarization. Using representative values $f = \omega/2\pi = 10$ MHz, $N_{e0} = 10^{11}$ m⁻³ and $B_0 = 30 \cdot 10^{-6}$ T in the ionosphere, an order-of-magnitude estimate for Verdet's constant and the Faraday rotation is

$$V \sim -10 \text{ rad}/(\text{m}\cdot\text{T}), \quad \frac{d\Psi_I}{dz} \sim -10^{-3} \text{ rad}/\text{m}. \quad (4.96)$$

From the result (4.92) and the definition of the tilt angle Ψ (cf. section 3.8.4), one can see that the tilt angle is rotated clockwise in the sense of \mathbf{B}_0 . For propagation in the opposite direction (i.e. antiparallel to \mathbf{B}_0), the positive z -axis reverses direction and the mode polarizations are reversed to $\rho_{\perp,O} = -i$ and $\rho_{\perp,E} = +i$ (cf. eqn. (3.82)). Deriving the change in tilt angle for backward propagation is identical to the procedure presented above for forward propagation, except that the reversed polarizations result in the tilt angle derivative

$$\frac{d\Psi_I}{dz} = -\frac{\omega}{2c} (n_E - n_O). \quad (4.97)$$

The result for backward propagation (4.97) is equal but opposite to the result (4.92) for forward propagation. However, the reversal of the propagation direction must result in either the x - or y -axis to reverse direction to retain a right-handed local polarization coordinate system. Consequently, the tilt angle Ψ_I is again rotated clockwise in the sense of \mathbf{B}_0 . Hence, the rotation of the tilt angle is irrespective to the propagation direction of the wave: a wave propagating in the positive z -direction experiences the same rotation as one propagating in the negative z -direction, as seen in the sense of \mathbf{B}_0 . The net result of the forward and backward propagation is then twice the effect of forward propagation. Although we have only treated propagation parallel to \mathbf{B}_0 , the results can be extended for a general propagation direction θ . In theory, such nonreciprocity can isolate one antenna from the other, although the signal is audible in one direction. In fact, the Faraday effect is used to construct optical isolators that pass through light in only one direction [13]. To clarify this claim, consider the following thought experiment of transmission between two dipole antennas A and B, illustrated in figure 21:

The dipole axis of A is in the x -direction, while the dipole axis of B is in the xy -plane so that B is at a 45° angle to the x -axis. The antennas are then separated by a distance z perpendicular to both axes. Assume that the antennas transmit through a plasma between them, where $\mathbf{B}_0 \parallel \hat{\mathbf{e}}_z$. Antenna A then transmits a wave \mathbf{E}_A towards B in the z -direction that is initially linearly polarized in the x -direction. As the wave propagates in the plasma between A and B, it undergoes a Faraday rotation that we take to be exactly 45° . As the wave reaches B, it is then polarized

exactly in the direction of B, and $\mathbf{E}_{A \rightarrow B}$ can then be received by the linearly polarized antenna [19]. Say then, that B transmits a wave \mathbf{E}_B of identical frequency towards A, so the initial polarization is then 45° . Now, on the path back to A, the wave undergoes again a rotation of 45° , so that the wave $\mathbf{E}_{B \rightarrow A}$ is polarized along the y-axis when it reaches A. Since the wave polarization is perpendicular to the dipole axis, antenna A cannot receive any signal due to a perfect polarization mismatch. As a consequence, although B hears A, A does not hear B.

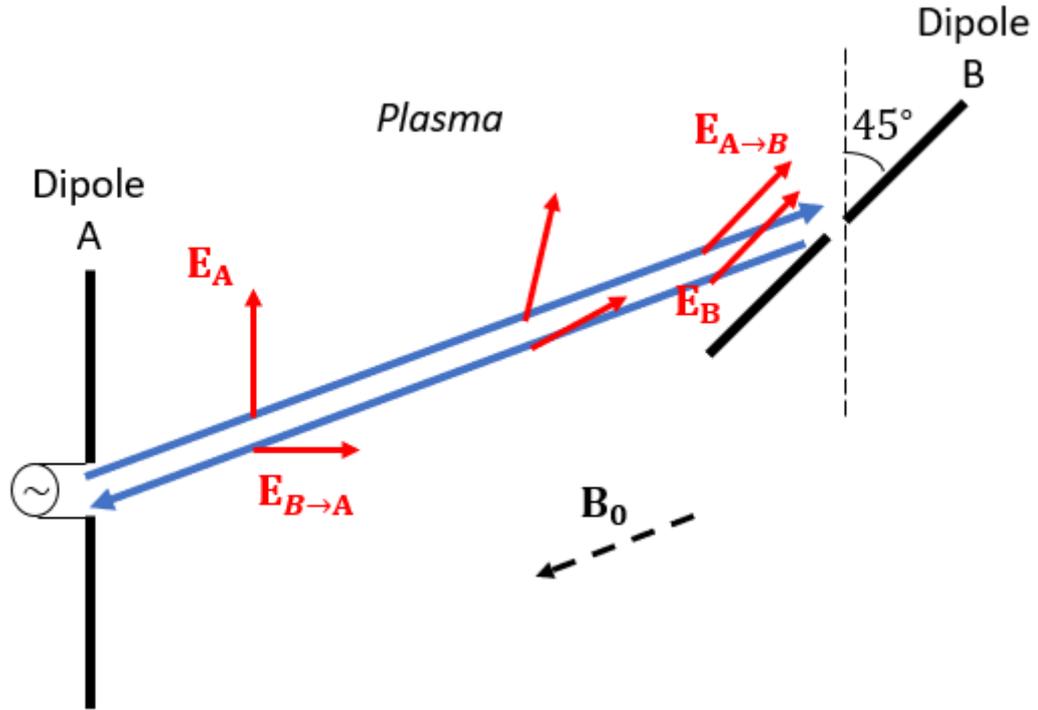


Figure 21: Illustration of signal nonreciprocity caused by the Faraday effect. The axes of two dipole antennas A and B are separated by the distance parallel to the z-axis, and the dipole axes are rotated by 45° with respect to each other in the xy-plane. The linearly polarized electric field transmitted by A undergoes Faraday rotation by 45° , and is conveniently polarized along B at the receiving end. On the other hand, the field transmitted by B is rotated by another 45° , resulting in $\mathbf{E}_{B \rightarrow A}$ to be polarized perpendicular to the dipole A axis.

5 Ray tracing program

This section is concerned with the developed ray tracing program from a software engineering perspective. Throughout sections 2 to 4, the necessary physics of ionospheric ray tracing has been discussed with the intent of developing an efficient ray tracing program. Knowledge of the nature and peculiarities of MF/HF radio wave propagation in the ionosphere is a bare necessity for this goal. If we wish to extract any practical use of this theoretical know-how, we need to translate this theory into source code for efficient and versatile software to function properly.

The developed ray tracer is written in the 2011 standard of the C++ language [43], and is developed for the Linux operating system. The current version of the software can be accessed at [48] using an Aalto University IT account. From the very early phases of development, the use of compiled languages (such as C++) for ray tracing were favored for fast performance. Experience shows, that ionospheric ray tracing is computationally expensive, with the numerical integration of eqns. (4.20)-(4.21) constituting the most resource-hungry part during program runtime. The problem becomes even greater when running the program in photon mapping mode, where several hundreds of thousands of rays are required to be traced through a few ground reflections to render a well-resolved intensity pattern. The choice of C++ as the compiled language for ray tracing was made for a few reasons, namely

1. Compatibility with other software written by the Aalto Space Physics research group, as the majority of our simulation models are written in C++, and
2. Established compatibility with various parallel computing APIs, such as OpenMP for the CPU [44] and CUDA for the GPU [42].

In addition to the core of the ray tracer being written in C++, the program is supported by ancillary software written in C, Fortran, and Python. The complete list of dependencies required by the ray tracing program is documented in section 5.1.

The compiled executable is called by providing a configuration text file, which includes all of the case-specific options and directives to be run by the ray tracer. In addition to the configuration file, the user is allowed to provide custom ionospheric altitude profiles of the electron density and collision frequency to constitute the background ionosphere medium. The required inputs and options are elaborated in section 5.2.

Sections 5.3-5.4 walk through the process of execution, presenting some central program subroutines using flowcharts. During execution, the ray tracer is required to check for certain exceptions related to problematic points of the refractive index. While the user is allowed to relax some of the interruption conditions (such as the maximal value of $|\text{Im}[n]/\text{Re}[n]|$), some of the conditions cannot be modified. For instance, if the ray is found to be evanescent at the beginning of a time step, tracing of the ray is interrupted immediately. In addition to documenting these design choices, it is prudent to give advice on when such exceptions might be encountered.

The last two sections 5.5-5.6 are concerned with the efficiency of the developed ray tracer. We present the utilized techniques of CPU acceleration by parallelization.

Ultimately, we document the achieved benchmark performance of the developed ray tracing program in some representative cases.

5.1 Dependencies and software architecture

The ray tracing program specifically developed for this thesis in C++ functions as an entry point executable. Apart from this ray tracing program, the execution of two ancillary third-party programs is required during runtime:

1. The IRI-2016 code written in Fortran 77 [40], and
2. the NRLMSISE-00 code written in the 1999 standard of the C language [39].

The IRI-2016 and NRLMSISE-00 source codes have been slightly modified for our use to facilitate compatibility with the ray tracer. The physics of both source codes are left untouched; the modifications that have been made only touch upon the I/O instructions in order to print variables useful to us. The modified source code for IRI-2016 and NRLMSISE-00 is provided with the developed ray tracer at [48], complete with makefiles for compiling the program. For compiling the ray tracer on Linux, the following dependencies are required:

- g++, i.e. the GNU C++ compiler, including the 2011 standard of C++ for compiling the ray tracer.
- OpenMP, used for CPU multithread parallelization of ray tracing.
- gcc, i.e. the GNU C compiler, including the 1999 standard of C for compiling NRLMSISE-00.
- f77, i.e. the Fortran 77 compiler for compiling the IRI-2016 standard.

Instructions for compiling the program are given in a README file provided with the ray tracer. The installation starts by building the IRI-2016 executable, and moves on to building the NRLMSISE-00 object files. Finally, the ray tracer object files are built and externally linked to the NRLMSISE-00 object files, and ultimately linked to the ray tracer executable for a program entry point. A high-level illustration of the installed program with dependencies is given in figure 22.

The compiled ray tracing program constitutes the most computationally expensive unit of the software product: It computes the radio ray paths by integrating eqns. (4.20)-(4.21), and prints the output to text files for later use in post-processing data analysis. The possible outputs of the program are described in detail in the next section. For post-processing, three Python scripts have been provided:

1. `plot_IRIoutput.py`, used for visualizing the ionosphere altitude profiles produced by NRLMSISE-00 and IRI-2016.
2. `plotrays3D.py`, which plots the three-dimensional ray paths $\mathbf{r}(t)$ on a geographic map projection.

3. `photonmap.py`, which reads the strike-points from the produced photon map and visualizes the calculated intensity on the predefined mesh (see section 4.8.2).

The provided post-processing Python scripts have been written in the 2.7 standard. In addition, the Python scripts require the following packages as dependencies:

- NumPy for processing multidimensional arrays.
- Matplotlib for plotting the results.
- Matplotlib Basemap for visualizing the results on geographic maps.

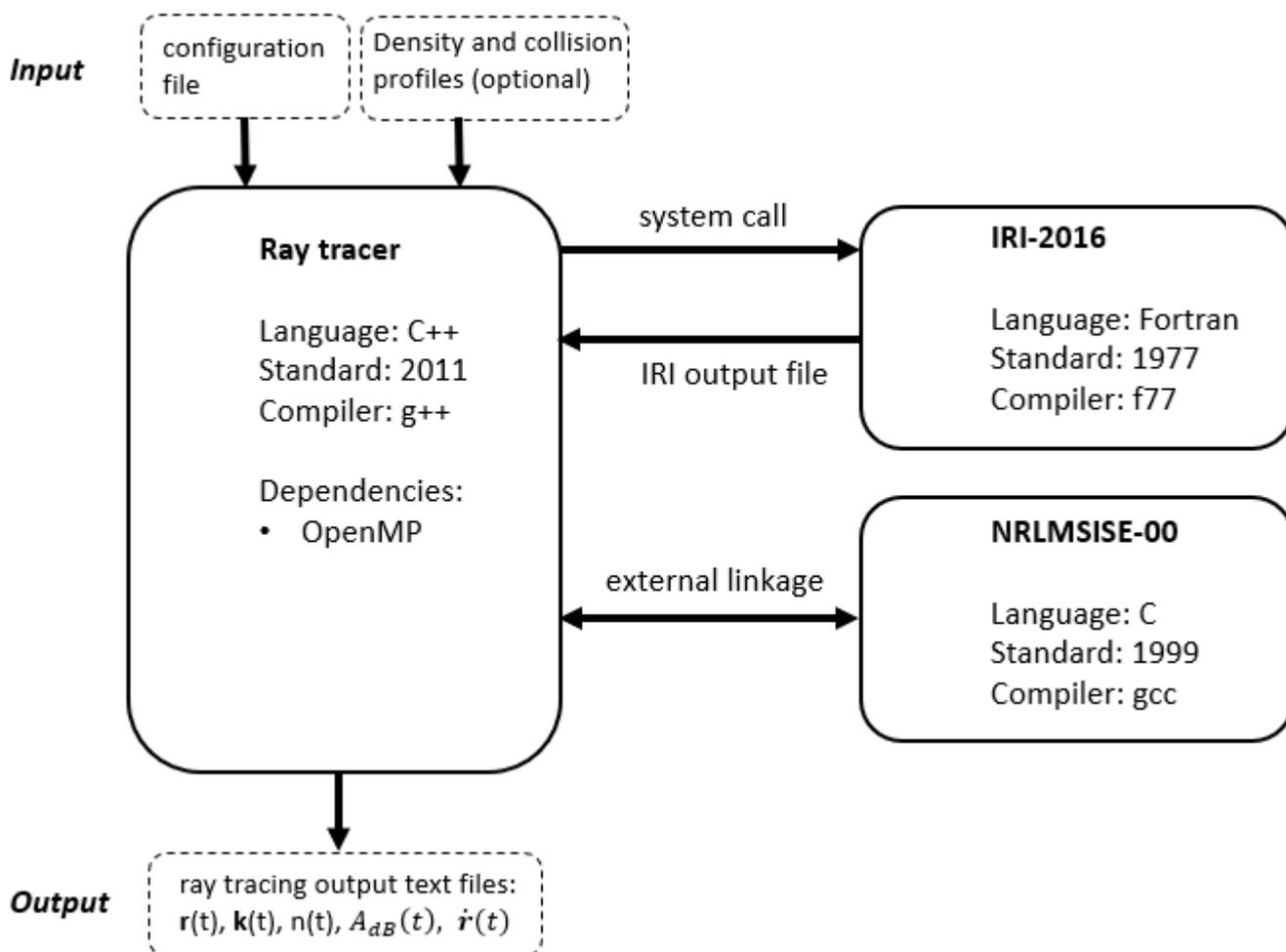


Figure 22: High-level illustration of the ray tracing program with required dependencies. The ray tracer (C++) and NRLMSISE-00 (C) modules are externally linked, while the IRI-2016 code (Fortran) is executed by system calls from the ray tracer. IRI-2016 produces an output text file, which is consequently read by the ray tracer to generate ionospheric altitude profiles of the electron density and collision frequency together with NRLMSISE-00.

5.2 Inputs, outputs and options

The ray tracer is controlled by providing a configuration text file as the single command line argument. The configuration file specifies all of the options required for the ray tracing session, such as

- The transmitter antenna location, type and orientation, and the transmission frequency,
- the geomagnetic field type, strength and dipole orientation,
- the magnetoionic mode (O/X) to be traced after the first magnetoionic splitting event,
- the permittivity and conductivity of the Earth's surface. The electrical properties are assumed to be the same for the entire surface.
- The path to the user-defined electron density and collision profiles, or alternatively, the options used by IRI and NRLMSISE-00 to generate profiles,
- the Cash-Karp 4 5 integrator parameters, such as the relative and absolute tolerances, initial and maximum stepsize, maximal number of integration steps etcetera,
- the choice whether collisions and/or the geomagnetic field are included in the ray equations. This choice ultimately determines the Hamiltonians to be used in ray tracing (see section 4.4).
- The choice of program mode, i.e. photon mapping or ray tracing. In ray tracing mode, the user defines the launch elevation and azimuth angles to be swept, while in photon mapping mode only the number of launched photons (and optionally the allowed elevation range) is defined.
- The maximum number of threads used by OpenMP for multithread parallelization, and
- the types of output to be written, such as ray position, wave vector, attenuation or refractive index for each time step.

A complete configuration file template is presented in appendix C, along with documentation about the possible options with explanations. At the beginning of execution, the ray tracer runs the subroutine `readConfigFile()`, which opens the provided configuration file and parses the contents into containers. Each option is given on a separate line in the configuration file; the input is case sensitive, but the options need not be in any particular order. If any of the required options are missing, or if the given option is not recognized, the program indicates the erroneous option and aborts execution. Furthermore, the program checks that the given numeric values in the configuration file are in their permitted domain: for instance,

since the integrator tolerance must be a strictly positive number, the program gives a warning and aborts if the given tolerance is smaller than or equal to zero.

The user is given the possibility to either provide ionospheric altitude profiles themselves as text files, or to generate them procedurally using IRI and NRLMSISE-00. The user-defined electron density and collision profiles list the ionospheric value (N_{e0} in m^{-3} or ν_{en} in Hertz) for each altitude (in kilometers) above the ground. As an example, the results in section 6 use profiles generated from 60 to 600 km with a step of one kilometer. The geometry of the provided profiles needs to be perfectly matched: the subroutine `readConfigFile()` checks the profiles line by line, and if the altitude on a given line differs between the electron and collision profiles, the program warns of a mismatch and aborts.

On the other hand, if the user has opted to generate the ionospheric background using IRI and NRLMSISE-00, the subroutine `readConfigFile()` performs a system call to the compiled IRI-2016 executable using the arguments provided in the configuration file. The IRI-2016 code is run using the standard options; this means, that the solar and magnetic indices as well as the F10.7 cm flux value are read from tabled data provided with the source code. The standard options used for running IRI-2016 – as well as the generated ionospheric profiles – are tabled in the generated text file *fort.7*, allowing the user to inspect the details of the system call. The generated *fort.7*-file is then opened and parsed by the `readConfigFile()`-subroutine, after which the NRLMSISE-00 subroutines are called by external linkage to generate corresponding profiles for the neutral densities. Together with the retrieved IRI-2016 profiles, the NRLMSISE-00 profiles are used to generate an altitude profile for the electron-neutral collision density $\nu_{en}(h)$ using eqns. (2.2)-(2.4). The generated altitude profiles are then saved to text files, allowing the user to inspect them using e.g. the provided Python script `plot_IRIoutput.py`

The ionospheric profiles used by this version of the developed ray tracer are one-dimensional: the values N_{e0} and ν_{en} are taken to vary only with altitude, which is a reasonable approximation within a marginal latitudinal and longitudinal deviation. The ionospheric profiles are generated above a given geographic location defined in the configuration file, and hence, the results of ray tracing are not expected to be very accurate for rays propagating far away from this location. For newer versions of the ray tracer, it is advised to implement three-dimensional profiles that vary with latitude and longitude as well. The three-dimensional variation of the ionosphere is illustrated in figure 23, where the F region critical frequency is drawn on isocontours above Northern Europe. It can be seen at the specified time of day, that the F critical frequency varies roughly one MHz in value over a latitudinal deviation of a thousand kilometers along the ground. The longitudinal change is however much less severe at these high latitudes, and cannot be distinguished without drawing more isocontours. However, although the program uses one-dimensional profiles of N_{e0} and ν_{en} , the geomagnetic field $\mathbf{B}_0(\mathbf{r})$ is taken to be truly three-dimensional, and is evaluated from eqn. (2.12) on all points of the ray path.

Using IRI-2016 and NRLMSISE-00, the one-dimensional ionosphere profiles are generated over a geographic location, spanning altitudes $h_j \in [h_{min}, h_{max}]$ where h_{min} and h_{max} are some defined minimum and maximum profile generation altitudes.

It should be noted, that although IRI-2016 is limited to altitudes in the range from 50 km to 1500 km, the ray tracing program appends the generated profiles to also include a value at the altitude $h = 0$ km. This way, interpolation can be performed safely for all rays with altitudes h within $h \in [0, h_{max}]$. Given that the ray altitude h is situated between two generated layers j and $(j + 1)$ so that $h_j \leq h < h_{j+1}$, the ionospheric value $f(h)$ (be it $N_{e0}(h)$ or $\nu_{en}(h)$) is estimated by linear interpolation:

$$f(h) \approx \frac{f(h_{j+1}) - f(h_j)}{h_{j+1} - h_j}(h - h_j) + f(h_j). \quad (5.1)$$

Moreover, the derivative with respect to altitude is approximated as being constant between the said two layers:

$$\frac{\partial f}{\partial h} \approx \frac{f(h_{j+1}) - f(h_j)}{h_{j+1} - h_j}. \quad (5.2)$$

From the definition (2.6) of h , the partial derivative of f with respect to Cartesian coordinate x_i can be evaluated by using the chain rule:

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial h} \frac{\partial h}{\partial x_i} = \frac{\partial f}{\partial h} \frac{x_i}{r}, \quad (5.3)$$

$$\Rightarrow \frac{\partial f}{\partial \mathbf{r}} = \frac{\partial f}{\partial h} \hat{\mathbf{r}}. \quad (5.4)$$

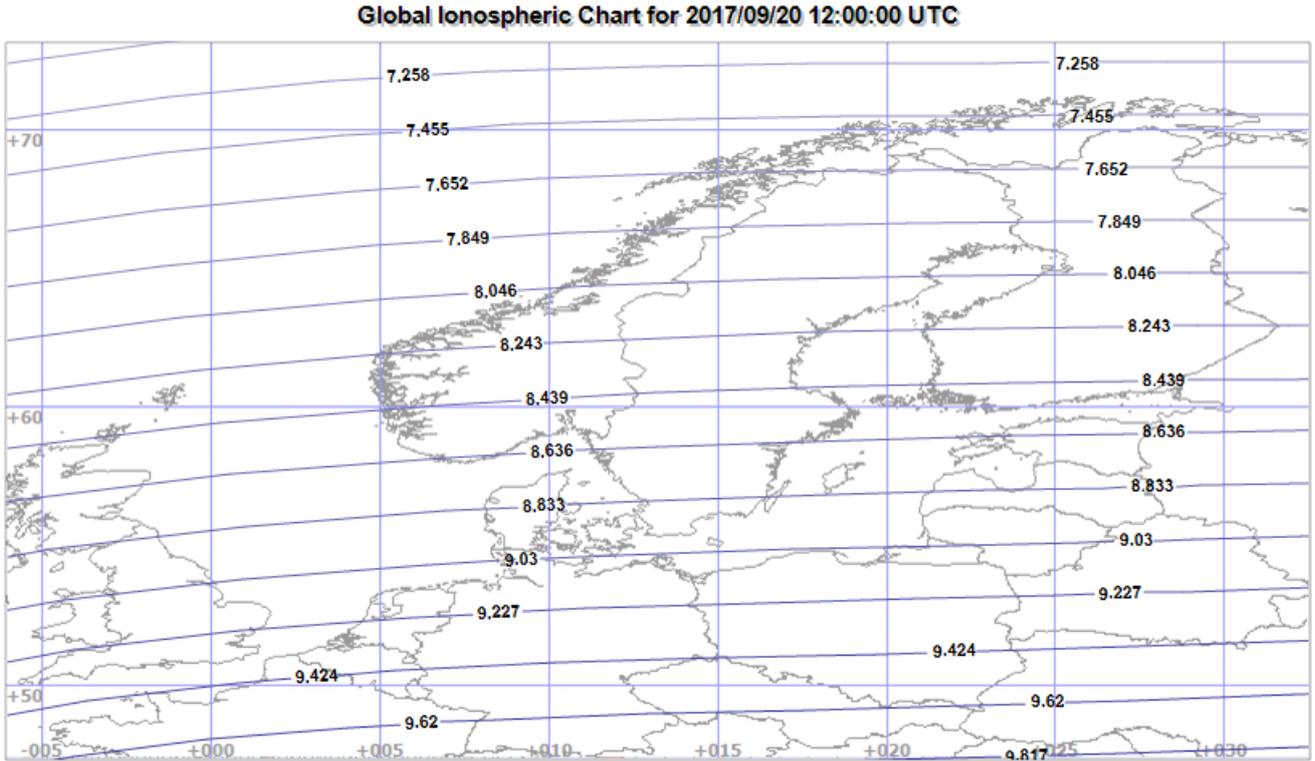


Figure 23: Isocontours of the F critical frequency in MHz above Northern Europe, generated using IRI by Proplab Pro 3. It can be seen, that the latitudinal variation is roughly one MHz over a ground path of a thousand kilometers, while the longitudinal variation cannot even be distinguished at this time of day in these high latitudes without drawing more isocontours. These results illuminate the problem when ionosphere profiles are assumed to be one-dimensional, unless ray tracing is limited to a small geographic area.

Finally, the subroutine `readConfigFile()` reads the desired outputs of the ray tracing session. The ray tracer allows for the following output to be saved:

- Ray position $\mathbf{r}(t)$
- Ray wave vector $\mathbf{k}(t)$
- Propagation time t since transmission
- Total attenuation $A_{dB}(t)$ in decibels
- Ray refractive index $n(t)$ evaluated from (3.51)-(3.52) on the path
- Ray group velocity $\dot{\mathbf{r}}(t)$ on the path
- Strike-points \mathbf{r}_k , $\dot{\mathbf{r}}_k$ and $A_{dB,k}$ on the ground $h = 0$
- Strike-points \mathbf{r}_k , $\dot{\mathbf{r}}_k$ and $A_{dB,k}$ on the surface defined by the maximum altitude $h = h_{max}$

The outputs of all rays are collected into a single text file, where the individual rays are separated by a header to facilitate post-processing. The output is evaluated for each time step of integration, except for the strike-point outputs which only save single points \mathbf{r}_k if the ray happens to intersect the target surface. The strike-point output files not only record the strike position \mathbf{r}_k , but the group velocity $\dot{\mathbf{r}}_k$ and attenuation $A_{dB,k}$ at the moment of intersection. This way, photon mapping can be performed as described in section 4.8.2. If the ray tracer is being run in photon mapping mode, all output options except for the strike-points are disabled: This safety precaution is hard-coded to save significant amounts of computation time and memory. Using moderate integration tolerances, the position output of a single ray requires a few hundred kilobytes of memory. In photon mapping mode using a few hundred thousand photons, the memory needed to save the positions $\mathbf{r}(t)$ of all rays would tally up to several tens of gigabytes.

5.3 Execution of ray tracing

The ray tracing program can be run in one of two modes:

1. Ray tracing mode: the user defines the launch elevation and azimuth angles to be swept, as well as the characteristic mode (ordinary/extraordinary) to be magnetoionically split into the first time the ray enters the ionosphere.
2. Photon mapping mode: the user defines the number of photons to be launched in random directions, as well as the allowed elevation range of launched photons.

Once the subroutine `readConfigFile()` has successfully parsed the contents of the configuration file, the program evaluates the position and orientation of the transmitter in the ECEF coordinate system presented in section 2. Given the transmitter latitude λ , longitude ϕ and altitude h above the ground, the position \mathbf{r} of the transmitter is determined unambiguously by using eqn. (2.5). Furthermore, given the antenna elevation α and bearing β , the orientation of the antenna is determined as illustrated in figure 24.

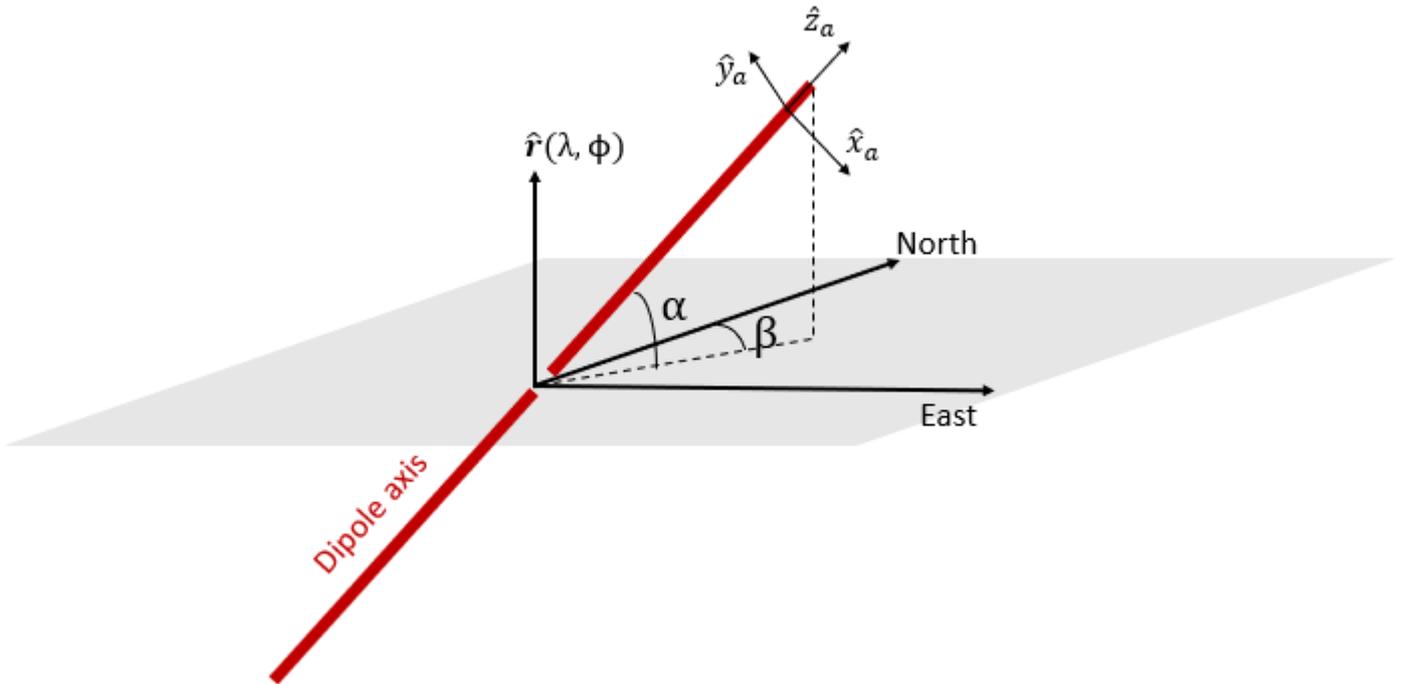


Figure 24: The antenna orientation as determined by the elevation angle α and bearing angle β . The figure depicts a dipole antenna for clarity, with the unit vector \hat{z}_a oriented parallel to the antenna axis. Since all the antenna types supported by this version of the ray tracer are assumed to be azimuthally symmetric, only two angles are required to orient the antennas unambiguously. If the antennas were not azimuthally symmetric, a third rotation angle would be necessary to determine the orientation unambiguously.

In ray tracing mode, the program loops through all the launch azimuth and elevation angles to determine the ray launch directions. The initial wave vector \mathbf{k}_0 of the ray is determined similarly as in figure 24, except that \hat{z}_a is taken to be in the direction of \mathbf{k}_0 and β is taken to be the azimuth value. One should thus note, that the ray tracer does not specify the initial group velocity of the ray, but the initial wave vector, which is parallel to the group velocity in a vacuum and isotropic plasmas. From the initial wave vector \mathbf{k}_0 , the program evaluates the initial polarization ρ_\perp and intensity per solid angle U from the specific antenna radiation pattern. An initial value for the decibel attenuation A_{dB} is obtained by setting the normalization factor U_m (eqn. (4.38)) to unity in the intensity pattern, and evaluating

$$A_{dB}(0) = 10 \log_{10} U. \quad (5.5)$$

Once the ray is initialized, the program calls the subroutine `integrate_ck45()`, which performs the Cash-Karp adaptive step size integration of eqns. (4.20)-(4.21) and (4.73) as detailed in section 4.5. If the ray altitude is initially below or above the ionosphere, i.e. if $h < h_{min}$ or $h > h_{max}$ as given in the configuration file, the subroutine propagates the ray into the ionosphere in a single time step without

altering \mathbf{k} or A_{dB} . Subsequently, a subroutine for performing magnetoionic splitting is called if the geomagnetic field is included in ray tracing. For the first magnetoionic splitting event, the subroutine selects the branch given in the configuration file to be followed. For subsequent splitting events (e.g. if the ray is later reflected from the ground back to the ionosphere), the subroutine selects the branch corresponding to the lowest splitting loss (i.e. for which eqn. (4.60) yields the highest value). After each magnetoionic splitting event, the attenuation of the ray is updated as

$$A_{dB} \rightarrow A_{dB} + 10 \log_{10} T_{O/E}. \quad (5.6)$$

Next, the subroutine `integrate_ck45()` proceeds to integrate the ray equations (4.20)-(4.21) and (4.73) as long as it holds that $h_{min} \leq h \leq h_{max}$ at the beginning of the time step. Once the subroutine has checked that the ray is still within the boundaries of the ionosphere, the program checks whether the mode of the ray needs to be interchanged due to a crossing of $X = 1$ (see section 3.9). Before performing an integration step, the subroutine still checks for whether the refractive index n is in a domain acceptable for the form of ray tracing described in this thesis. These exceptions, as well as their possible causes, are described in the next section. If all these checks are passed, the subroutine proceeds to propagate the ray further one time step using the Cash-Karp integrator.

As was detailed in section 4.4, the ray tracer uses the Booker quartic Hamiltonian in anisotropic plasmas if $X > 0.1$, and the Appleton-Hartree Hamiltonian in other cases. After computing the candidate for the next step $(\mathbf{r}, \mathbf{k}, A_{dB})(t + \Delta t)$, the subroutine `step_controller()` is called to enforce that the step is within the dictated error tolerances. `step_controller()` adjusts the step size accordingly, until a satisfactory step within error tolerances is found. This block of the subroutine is by far the most computationally expensive, as the Cash-Karp integrator requires six function evaluations per step. Moreover, from appendix A it can be seen that the derived analytic expressions of the ray equations are rather lengthy in anisotropic media, making it necessary to

- (a) avoid unnecessary function evaluations by prudently adjusting the step size, and
- (b) evaluate the expressions in appendix A in a way that utilizes the hardware efficiently to achieve high throughput. Such methods are discussed in section 5.5.

After an integration step has been computed successfully, the subroutine `integrate_ck45()` saves the current values of $(\mathbf{r}, \mathbf{k}, t, n, \dot{\mathbf{r}})$ to memory if the corresponding output is enabled. Finally, if the number of computed integration steps does not exceed the defined maximum limit, the integrator proceeds to compute the next integration step if the ray is still inside the ionosphere. However, if the ray is detected to be situated below or above the ionosphere, the program branches to one of two subroutines: If the ray is above the defined maximum altitude, the program stores the strike-point to memory if required and finishes tracing the ray. On the other hand, if the ray is situated below the predefined limiting region h_{min} ,

the program first computes the limiting polarization ρ_{\perp} corresponding to the mode of the downward ray. Next, the ray is propagated to the ground in a single time step without altering \mathbf{k} or A_{dB} , and the strike-point as well as the current ray values are saved to memory if required. If the ray has undergone ground reflections less than the defined maximum number of hops, the ray is reflected from the ground with a corresponding shift of polarization and reflection loss as detailed in section 4.7.3. The decibel attenuation is updated from the reflection loss R (4.69) by evaluating

$$A_{dB} \rightarrow A_{dB} + 10 \log_{10} R. \quad (5.7)$$

Finally, the ray values after reflection are saved to memory if required, and the subroutine `integrate_ck45()` is called again to resume propagation. The entire process of execution in ray tracing mode is illustrated in figure 25.

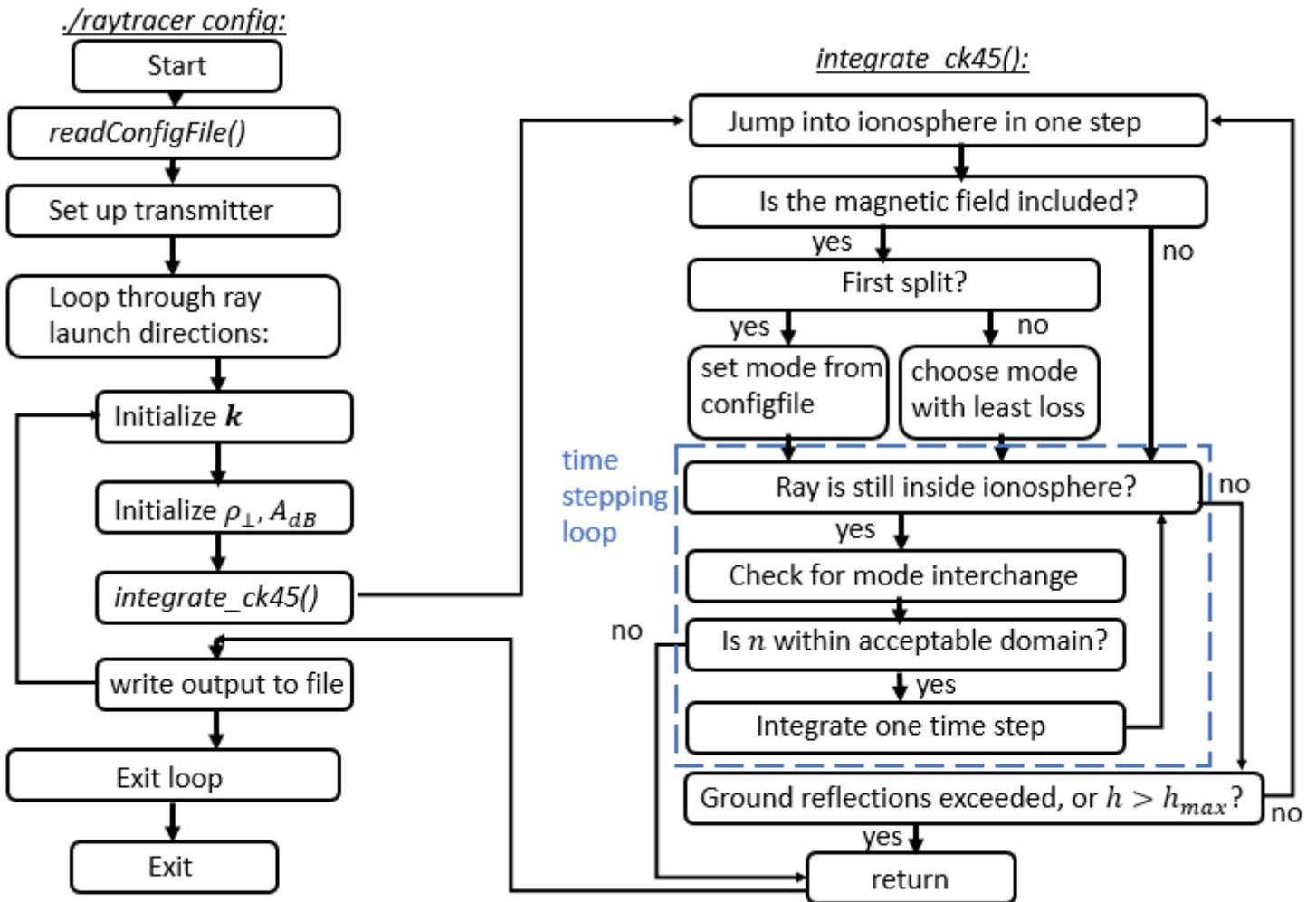


Figure 25: Flowchart of the developed program being run in ray tracing mode. The main time-stepping loop is denoted by a blue dashed rectangle, and constitutes the most computationally expensive unit of the program. At the end of each time step, the ray values are stored to memory if the corresponding output is enabled in the configuration file.

When running the program in photon mapping mode, the procedure is almost the same as in figure 25, except that the initial wave vectors are generated randomly on the unit sphere as described in section 4.8.2. Moreover, in ray tracing mode the branch of the first splitting event is defined in the configuration file, whereas in photon mapping mode the subroutine `integrate_ck45()` is called using both modes for a photon launched in a single random direction. Furthermore, in photon mapping mode only the strike-points are saved to memory to generate the photon map.

5.4 Exception handling

During ray tracing, the program maintains a number of diagnostics to detect whether the rays are about to encounter problematic regions of parameter space that could compromise the validity of ray tracing. Before each integration step, the subroutine `integrate_ck45()` evaluates the refractive index $n(\mathbf{k}, \mathbf{r})$ using the Appleton-Hartree formula (3.51) and checks that

1. $\text{Re}[n] > 0$, i.e. that the ray is not evanescent.
2. $\text{Re}[n] < \epsilon_{res}$, where $\epsilon_{res} \gg 1$ is the resonance tolerance defined in the configuration file. This check ascertains that the ray is not about to encounter a resonance, a condition required for the cold plasma approximation to be valid.
3. $|\text{Im}[n]/\text{Re}[n]| < \epsilon_{pr}$, where $\epsilon_{pr} \sim 0.1$ is the pseudoreal tolerance defined in the configuration file. This check in turn ascertains that the traced ray is pseudoreal, a condition required for the ray tracing formalism in section 4 to be valid in lossy media.

If the refractive index passes all of the above checks, the program proceeds to integrate the ray one step further as described in section 5.3. In any other case, tracing of the ray is stopped, the user is warned that the check was not passed, and the program returns to write the requested ray path variables to file. Although the tolerance of check number 1 in the list above is hard-coded, the user is allowed to tweak the tolerances of checks 2 and 3 at will. However, we should remind ourselves that these checks are in place primarily to ensure the approximate physical validity of the solved ray paths, and only secondarily to ensure that the program does not crash when integrating eqns. (4.20)-(4.21). In [3], it is argued that ray tracing in real space yields reliable results as long as $\arctan(\text{Im}[n]/\text{Re}[n]) < 10^\circ$, equating to $\epsilon_{pr} \approx 0.18$ in check number 3. Particularly when tracing rays of medium frequency, i.e. less than 3 MHz, the ratio $|\text{Im}[n]/\text{Re}[n]|$ may exceed this limit, although the integration of (4.20)-(4.21) can be carried out without difficulty to yield some solution for the ray path. If the user has defined a large value of ϵ_{pr} , the solved ray path may then not be physically valid.

Rays of medium frequency are most likely to encounter evanescence when reflected at near-vertical incidence. In many such cases, decreasing the integration tolerances and maximum step size will ensure that check number 1 is passed on the

entire integrated ray path, as the ray is forced to tiptoe through reflection without ever finding itself inside an evanescent region. In general, we have found that decreasing the integration stepsize is likely to help with passing checks 1-3 without the need of altering ϵ_{res} nor ϵ_{pr} .

So far, when ray tracing in the frequency range 1-10 MHz, we have never encountered a ray that had to be stopped for not passing check number 2 using $\epsilon_{res} = 10^3$. Based on these results, it simply appears that radio waves in the frequency range of our interest are unlikely to encounter resonance. It could be hypothesized, that resonance is more likely to occur if the frequency is lowered beyond 1 MHz, as the wave frequency then approaches the gyrofrequency in the ionosphere. For such low frequencies however, the use of full-wave methods are usually necessary [3], rendering the ray tracing formalism used in this thesis essentially useless.

One last design choice in exception handling deserves to be mentioned. If the downward traveling ray emerges from the limiting region (defined at the altitude h_{min}) at a very steep angle, the wave vector is nearly parallel to the Earth's surface. The program attempts as usual to propagate the ray to intersect the Earth's surface in a single time step, but the line extending from the limiting region in the direction of \mathbf{k} does not intersect the surface of the Earth. If such a case is detected, the ray is stopped in the limiting region, and the user is warned of the occurrence before writing the requested output to file. In practice, this case is encountered only if the ray is launched at a very low elevation angle (like $\alpha < 1^\circ$), in which case the wave should not be treated as a skywave but a ground wave instead. For this reason, the user may define an elevation range $\alpha \in [\alpha_{min}, \alpha_{max}]$ of the launched photons when performing photon mapping. However, if the problem persists, the user is advised to lower the altitude h_{min} of the limiting region. This fix may allow the rays to be refracted for a sufficient duration to result in downgoing rays to always intersect the Earth's surface.

5.5 Parallelization on the CPU

In this section, we briefly describe the measures taken to increase the program throughput on the CPU. Since the rays can be propagated completely independently from one another in the ray tracing formalism adopted for this thesis, the obvious possibility of multithread parallelization presents itself for hardware acceleration. In the developed program, the ray launch directions are distributed between the available threads allotted to ray tracing using OpenMP. Each thread then initializes the ray in the given launch direction, and calls `integrate_ck45()` on the initialized ray. Furthermore, the threads are instructed to write their output to separate blocks of memory: This way, no critical junctions are present inside the parallel block where the threads would have to wait for each other to finish. In essence, the for-loop parallelized using OpenMP appears as follows in the developed program:

```
#pragma omp parallel for
for direction in launch_directions:
    thread=omp_get_thread_num()
    ray=initialize_ray(direction)
```

```

integrate_ck45(ray)
write_output(ray,thread)

```

After all rays have been traced, a single thread is instructed to gather the output from all the other threads into a single output text file for the user to access. Reading and parsing the configuration file, as well as generation of ionospheric profiles, is not parallelized in this version of the ray tracer. Benchmark performances for single-thread and multithread execution are presented in the following section.

The time-stepping loop inside the subroutine `integrate_ck45()` (see figure 25) constitutes the most computationally expensive block when tracing a single ray. Here the methods of hardware acceleration become particularly crucial. For this block to achieve high throughput (i.e. high rate of floating point operations per second), the right-hand side of eqns. (4.20)-(4.21) need to be evaluated in a way that conforms to good programming practice. The common methods to achieve this are [15]

1. Reusage of data in registers and caches, saving the processor the trouble of fetching already evaluated variables from the slow main memory.
2. Instruction-level parallelism (ILP), i.e. writing a series of independently executable instructions that can be pipelined.
3. SIMD (single instruction, multiple data) instructions, commonly in the form of vector instructions.

In evaluating the derivatives of the Hamiltonian (appendix A), and subsequently, the right-hand side of eqns. (4.20)-(4.21), the developed software should be written in a way that utilizes the above methods. As an example of instruction-level parallelism, consider the following two versions of a function tasked with evaluating the sum of four numbers $\{a_1, a_2, a_3, a_4\}$. Assume, that the sum of two numbers of this data type can be evaluated in time T . The first version of the function, which does not incorporate ILP, would calculate all of the sums at once:

```

define sum_four_version1(a1, a2, a3, a4):
    return(a1 + a2 + a3 + a4)

```

The first version of the function has to wait for the sum $a_1 + a_2$ before evaluating its sum with a_3 and finally with a_4 , requiring time $3T$ to finish. However, the second version of the same function evaluates the partial sums separately, which the hardware sees as two independent operations by branch prediction:

```

define sum_four_version2(a1, a2, a3, a4):
    b1 = a1 + a2
    b2 = a3 + a4
    return(b1 + b2)

```

In this second version utilizing ILP, the partial sums b_1 and b_2 are evaluated by the processor in parallel in total time T , after which the total sum $b_1 + b_2$ is evaluated in time T to result in the total time $2T$ to finish. By writing the same operation in a way that has many independent instructions ready for execution, the program has increased its throughput by up to 50 percent.

Vector instructions allow the SIMD-capabilities of modern CPU designs to be utilized to perform multiple floating-point instructions in a single instruction. For instance, the sum of two three-dimensional vectors \mathbf{v}_1 and \mathbf{v}_2 could be performed in a single instruction, which would simultaneously evaluate three sums corresponding to the three vector components. Unfortunately, the hardware available to us presents a problem with vectorizing the arithmetic of double-precision complex numbers. The developed ray tracing program uses double floating-point precision in evaluating the right-hand side of eqns. (4.20)-(4.21), and a single variable of C++ type `double` takes 64 bits of memory on our system. The terms constituting the right hand side of (4.20)-(4.21) are complex numbers, making each complex-valued scalar term occupy 128 bits of memory (64 bits for the real and imaginary parts individually). By extension, each three-dimensional vector term then occupies 384 bits of memory. Alas, the vector registers are only 256 bits wide on our CPU, necessitating us to e.g. store two of the three scalar components in a single vector register and the remaining component in a second vector register. However, such vectorization would obfuscate the code and possibly jeopardize the readability of the program. For this reason, we have opted to not utilize vector instructions in the current version of the ray tracing program.

5.6 Benchmark performance

With the developed ray tracing program exhibiting the features and specifications as described in the previous sections, we set to benchmark the performance of the software. In this section, we only aim to measure the program execution time on the CPU in some representative cases. Of course, the execution time does not only depend on the particular system the program is being run on, but on the rays being traced as well. However, we are able to cast more objective judgement on the software design by studying scalability. Hence, we set to study how the execution time is affected when we

1. increase the number of rays to be traced, and
2. increase the number of threads available to the program.

The benchmark performance is measured on an Intel Xeon® E5-2687W v2 CPU, running on a Linux Ubuntu operating system. The key technical specifications of the CPU are

- Processor base frequency: 3.40 GHz
- Cache: 25 MB SmartCache

- Random access memory: 125.9 GB
- Number of cores: 8
- Number of threads: 16

We start by breaking down the execution time of a representative ray tracing session. The execution time of the program can be divided into three main sections, namely:

1. Parsing the configuration file
2. Generating the ionospheric profiles using IRI and NRLMSISE-00 (if IRI is enabled in the configuration file)
3. Ray tracing or photon mapping, including writing the output to file

Let us measure how these subtasks perform in a simple representative run when tracing a single ordinary ray of frequency 5 MHz. For the integrator, we use tolerances $\epsilon = 10^{-8}$, an initial stepsize 10^{-8} s, a maximum stepsize 10^{-6} s, and we include both the geomagnetic field and collisions in the ray equations. The ionospheric background is generated for midnight in midsummer of 2018 using IRI-2016, and the ray is transmitted northward (azimuth= 0°) at an elevation angle of 40° from Otaniemi. The ray path solved by the developed program is presented in figure 26.

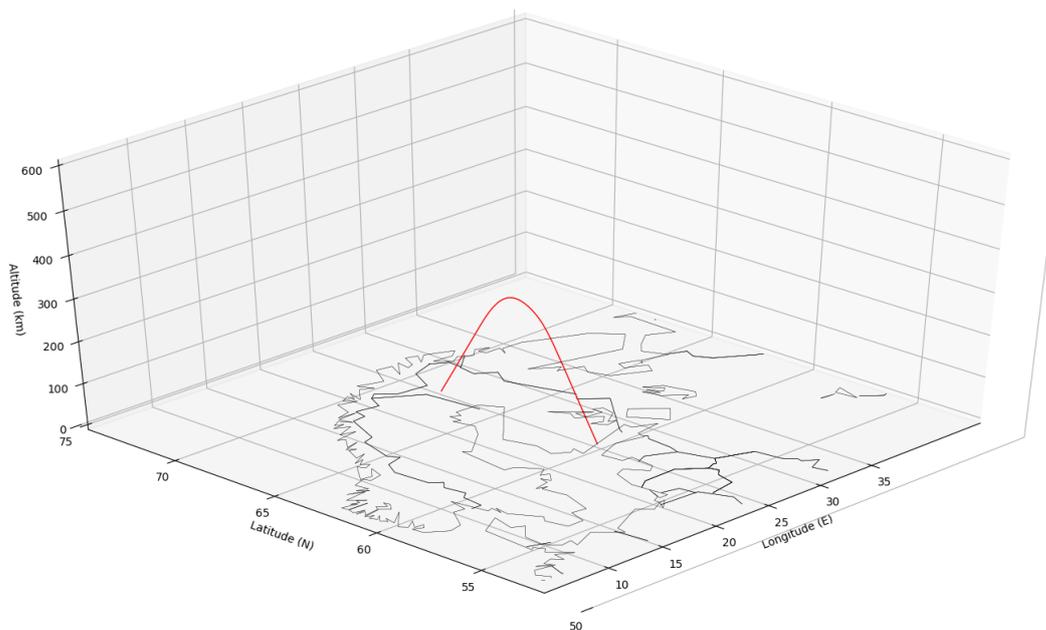


Figure 26: Ray tracing result of a single 5 MHz ordinary ray transmitted from Otaniemi. The ray reaches a maximum altitude of 270 km before being reflected back towards the ground, and reaches a ground path distance of roughly 1000 km from the transmitter before being reflected from the ground in Northern Finland. The total ray tracing time (not including ionospheric background generation) was measured to 93 milliseconds wall clock time, averaged over five runs. The axes of the three-dimensional plot are not to scale.

Reading and parsing of the configuration file, including importing user-defined ionospheric profiles, was performed in less than a millisecond. However, generating the ionospheric background from scratch using IRI-2016 and NRLMSISE-00 was clocked to 204 milliseconds wall clock time (averaged over five runs), a significant slowdown from the former option. This is because the program needs to perform a system call to the IRI executable, wait for the ionosphere to be generated, and subsequently read the produced output file. This process could surely be made faster by externally linking C++ and Fortran during compilation, which could prove to be a necessary improvement if generation of a three-dimensional ionosphere is attempted in future versions of the program. The actual tracing of the transmitted ray, including writing of output, was clocked to 93 milliseconds when all the output options were enabled. However, when only the strike-point outputs were enabled, the tracing was clocked at a mere 22 milliseconds. If the path output options are enabled, the program is required to store the ray values after each time step, whereas it is only required to store and write the strike-points if no path output is enabled. Thus, if the sought application allows, a sizeable speedup can be achieved by disabling the path output options.

Depending primarily on the ray frequency and launch elevation, the time required to trace one single ray varies. To study the scalability of the program, we inspect

the execution time of single-threaded photon mapping using an increasing number of photons N . The results for an anisotropic and collisional ionosphere are presented in figure 27. As could be expected, the execution time scales linearly for large N .

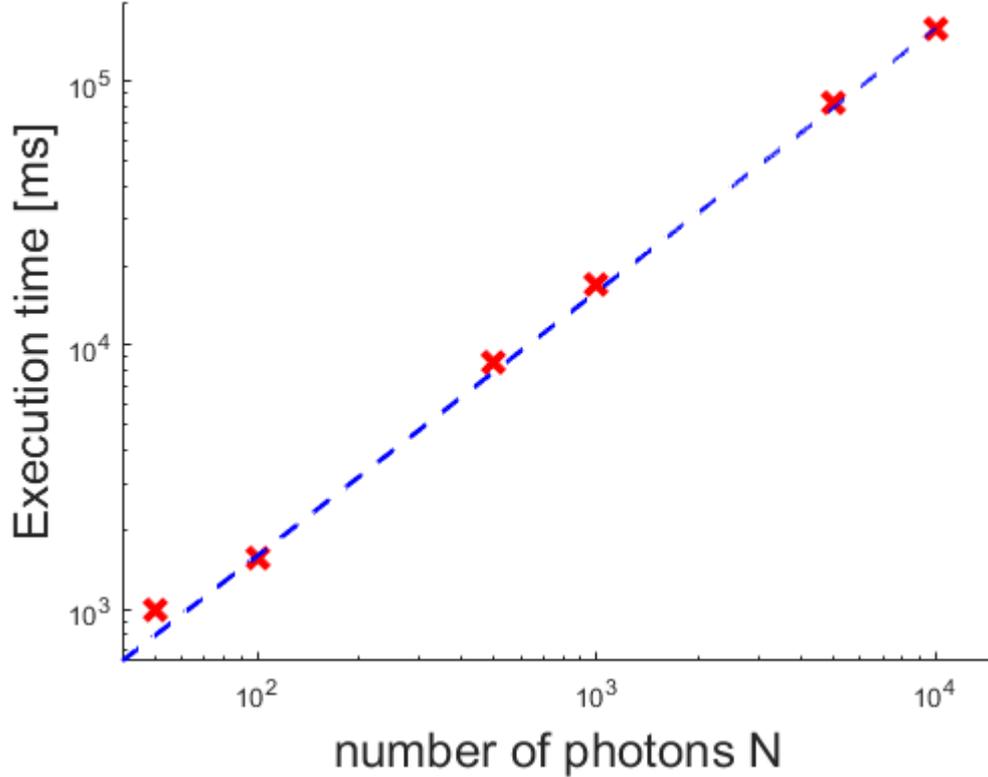


Figure 27: Scaling of single-threaded photon mapping, measured for an increasing number of launched photons N . The measured execution times are denoted by the red markers, while the blue dashed curve represents a linear least squares fit to the data. Note, that both axes are logarithmic. It can be seen, that the measured execution time of photon mapping scales linearly for large N .

For the last benchmark, we investigate the speedup from multithread parallelization. Performing photon mapping using $N = 10^4$ photons, we measure the execution time when using an increasing number of threads. The results are presented in figure 28. Although our system contains only 8 physical cores and 16 threads, our CPU is able to run multiple threads per core using hyper-threading, extending the number of maximum threads to 32. However, from the results in figure 28, we see that the program throughput does not increase significantly when using more than 16 threads, rendering the benefits of hyper-threading to be marginal at best.

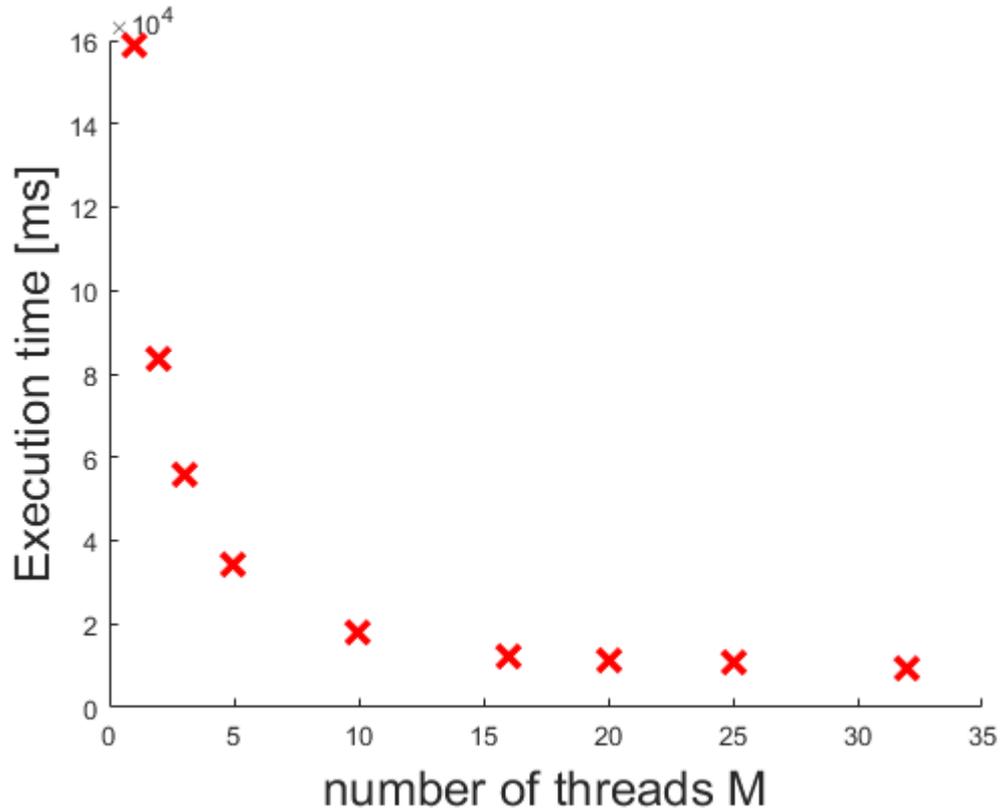


Figure 28: Benefits of multithread parallelization when running the program in photon mapping mode with $N = 10^4$. Launching and tracing the paths of randomly directed photons constitute the greatest computational expense of the program by far, providing ample opportunities for multithread parallelization. However, hyper-threading ($M > 16$) can be seen to yield marginal speed increase on our system, as the execution time is not decreased significantly beyond this point.

6 Ray tracing results

This penultimate section is tasked with validating the developed ray tracing program. As with any software in the field of computational physics, we are urged to investigate whether the program produces results that conform with reality or not, and to this end, we employ a range of empirical and theoretical measures. In the scope of this thesis, we were unfortunately unable to set up an extensive empirical campaign to validate the predictions of our ray tracer against measurements obtained by pairs of skywave radio transmitters and receivers. Instead, we compare our program against the predictions of Proplab Pro 3 [46], a commercial HF radio-prediction software. Comparison is made using two metrics: the difference in predicted ray paths, and the difference in predicted MUFs.

In section 4.9 it was proposed, that systematic errors in the ray tracing program could be identified by studying the reversibility of ray paths. In section 6.2, we find that the developed program indeed produces ray paths that can be reversed back to the transmitter, with the reversibility error decreasing when the integrator stepsize is made smaller. Moreover, the assertion was made that certain common model approximations can result in unphysical irreversibility. To this end, we investigate ray reversibility when assuming $\frac{\partial \mathbf{Y}}{\partial \mathbf{r}}$ to be zero in the ray equations, and find that such an assumption produces unphysical irreversibility that is not mended by a decrease in stepsize.

The developed program estimates absorption loss by integrating eqn. (4.73) in parallel with the ray equations. To better understand the validity of the procedurally generated ionosphere, we compare the predicted absorption loss against empirically observed measurements obtained from vertical ionosonde sounding. We find, that the developed program underestimates absorption loss by roughly a magnitude in a single sounding dataset, prompting further investigations into the used methods of evaluating signal loss.

Finally, we present how the proposed method of photon mapping could be used to estimate skywave coverage maps. We presents results for both transmitting ground stations and orbiting satellites, and investigate how photon mapping could be utilized in analyzing the measurements of the Suomi 100 MF/HF radio instruments.

6.1 Code validation using Proplab Pro 3

Proplab Pro 3 (hereafter referred to as simply Proplab) is a commercial HF radio-prediction software developed by the privately owned company Solar Terrestrial Dispatch [46]. Much like the program developed for this thesis (hereafter simply referred to as the ray tracer), Proplab uses ray tracing to compute the paths of transmitted radio waves through an IRI-generated ionosphere. Although our ray tracer uses the 2016 standard of IRI, the 3.1 version of Proplab uses IRI-2007 for ionosphere generation. Because of this, we may expect slightly different ionospheric profiles, and consequently, different ray paths between the two programs.

By virtue of being a commercial software tool, the specific implementations and design choices remain hidden from the user: If a difference is detected between the

predictions of Proplab and our ray tracer, we are undeniably limited in discerning the root cause. This being said, it is still interesting to see how our program fares against Proplab in the sought ways of code validation.

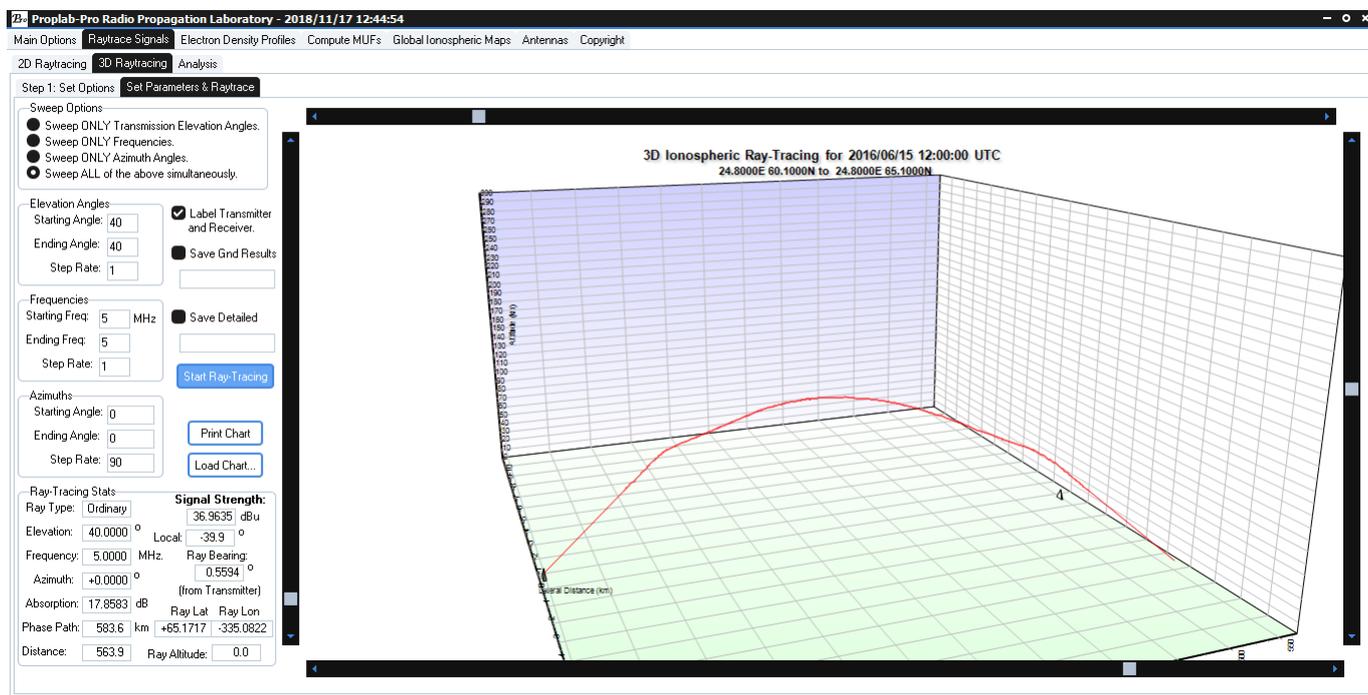


Figure 29: The graphic user interface of Proplab Pro 3, running in 3D ray tracing mode. A single 5 MHz ordinary ray is transmitted northward at a 40° elevation from $(60.1^\circ N 24.8^\circ E)$, and the ray eventually strikes the ground some 500 km from the transmitter.

6.1.1 Predicted ray paths

For the first test, we measure how the computed ray path position differs between the two programs. In these tests, an ordinary ray of frequency 5 MHz is transmitted northward (azimuth= 0°) at an elevation of 40° . In both programs, the transmitter is located near Otaniemi, at $(60.1^\circ N 24.8^\circ E)$, and tracing is performed in an ionosphere generated for the date and time 15.6.2016 12:00 UTC. The transmitted ray is traced for a single hop until it strikes the ground, and the distance between the strike-points predicted by both programs is evaluated. More precisely, we evaluate the great circle distance between the two points, which we then take as a metric to investigate the difference in ray paths.

In these tests, our ray tracer uses one-dimensional ionospheric profiles that have been generated directly above the transmitter. However, Proplab performs ray tracing in a three-dimensional ionosphere, arguably resulting in more precise ray paths further away from the transmitter. Moreover, the used profiles are not necessary identical between the two programs even directly above the transmitter, due to differing IRI options. The current version of our ray tracer always uses the standard

IRI options, while in Proplab the user is given some freedom in tweaking these options. In essence, such an experiment can be considered a blind experiment: How well do the programs agree with each other when running "as is", i.e. when using the standard settings of both programs without purposefully adjusting them with prior knowledge. In order to plot the output of both programs in the same graph, the ray data produced by Proplab is exported and converted to the ECEF coordinate system.

The results of such an experiment are presented in figure 30 with the corresponding (one-dimensional) electron density profiles. The electron density shown by Proplab has been evaluated at mid path, above circa ($62.1^\circ N$ $24.8^\circ E$). The ray paths now include the effects of both the geomagnetic dipole field and collisions. It can be seen, that the produced ray paths are initially nearly identical, until the reflection point at an altitude of roughly 140 km. From here, the produced ray paths diverge, and eventually strike the ground 76 kilometers apart. The outcome is virtually unchanged when collisions are not included, with the strike-points still being separated by roughly 76 kilometers. Curiously enough, the shown electron density profiles are nearly identical, save for a slightly thicker F1 region predicted by our ray tracer.

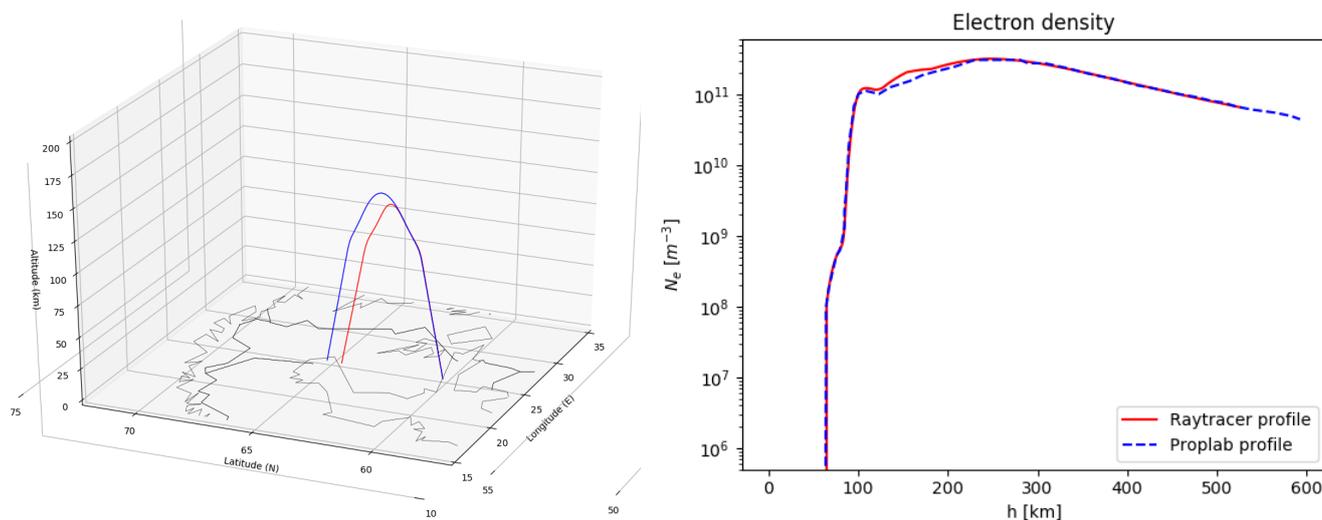


Figure 30: Left: The ray paths predicted by our ray tracer (red) and Proplab (blue) over a single hop. The predicted strike-points are separated by a great circle distance of roughly 76 kilometers. The ray computed by our ray tracer has traversed a ground path distance of 487 km from the transmitter. Right: Electron density profiles produced by both programs for this test. The profiles are seemingly identical, except for a slight difference in the F1 region.

To minimize the difference in the used ionospheric profiles, we import the Proplab electron density profile shown in figure 30 to our ray tracer. Furthermore, we do not include the effects of collisions in the ray equations. The same tests are repeated, both with and without the inclusion of the geomagnetic field. These results are shown in figure 31. With the geomagnetic field being included, the strike-points are

separated by 29 km. When we further exclude the geomagnetic field, the strike-points are separated by only 11 km.

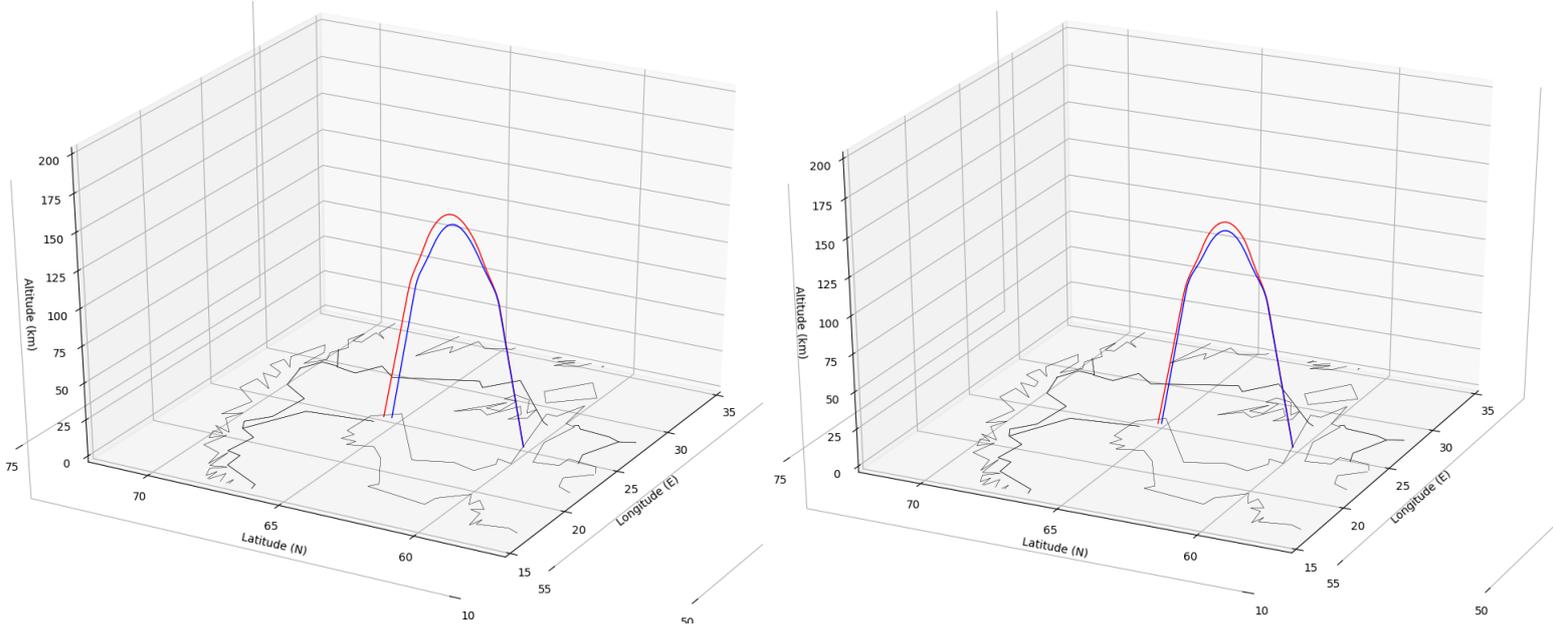


Figure 31: Ray paths predicted by our ray tracer (red) and Proplab (blue), using the Proplab electron density profile of figure 30 in our ray tracer. The leftmost figure depicts the case when the geomagnetic field is included, while in the rightmost figure the geomagnetic field is not included. The strike-points are separated respectively by 29 km and 11 km.

These results might suggest, that the three-dimensional ionosphere of Proplab is the primary culprit behind the observed difference in ray paths. In fact, in isotropic and collisionless plasmas it follows from the ray equations that $\dot{\mathbf{r}} \parallel \mathbf{k}$ and $\mathbf{k} \parallel -\frac{\partial N_{e0}}{\partial \mathbf{r}}$ (see eqns. (A.47)-(A.48)). Therefore, if the electron density only varies with altitude, it follows that \mathbf{k} is always normal to the Earth's spherical surface. Hence, the ray experiences no lateral deviation: a ray transmitted northward will continue to propagate northward or southward for all of its path. This is the case represented by the rightmost graph in figure 31, where the the path solved by our ray tracer produces a strike-point that lies on the initial longitude $24.8^\circ E$. This behavior is not observed in the path produced by Proplab. The strike-point has laterally deviated slightly from this initial longitude, suggesting the presence of a three-dimensional ionosphere that varies not only with altitude.

6.1.2 Predicted maximum usable frequencies

When designing skywave radio links, the maximum usable frequency (MUF) is a quantity of significant practical importance. Low frequency waves are attenuated much more strongly by absorption, favoring the use of high frequencies between two stations linked by a skywave channel. However, the transmission frequency cannot

be increased arbitrarily, as frequencies too high would penetrate the ionosphere and escape into outer space. At high frequencies, the skywaves transmitted at steep elevation angles are not reflected back to the ground, forming a skip zone where no skywaves are incident. An illustration of MUF and the skip zone is presented in figure 32 using Proplab.

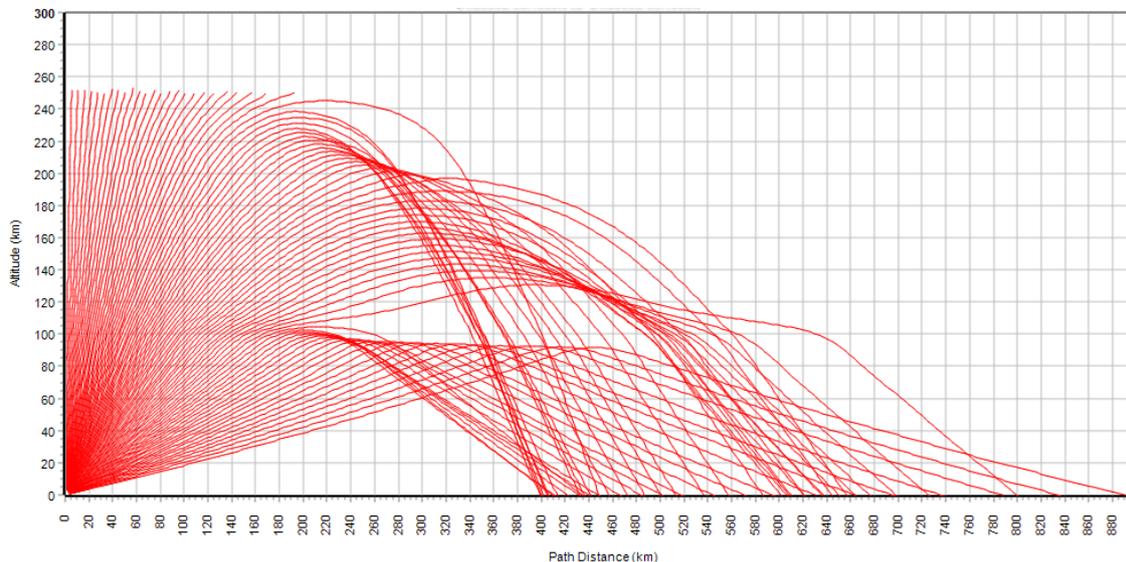


Figure 32: Illustration of MUF using Proplab. The target receiver is situated 400 km from the transmitter. At the particular MUF, the skywaves closest to the transmitter are just within reach of the receiver at the edge of the skip zone. When using any frequency higher than the MUF, the skip zone extends to a distance further than the receiver, rendering the frequency unusable for this pair of stations.

For our second validation test, it was decided to compare the MUF predictions between Proplab and our ray tracer. The transmitter is again situated at ($60.1^{\circ}N$ $24.8^{\circ}E$) on the 15th of June 2016 12:00 UTC. The target receiver is due north from the transmitter, situated between 200 km to 700 km from the transmitter as measured on the great circle path. The effects of the geomagnetic field as well as collisions were included. The ionosphere was generated by both programs using their respective standard settings, again resulting in the profiles presented in figure 30. For finding the MUF, it was decided to only study the the ordinary mode, as the extraordinary mode is attenuated more strongly over long distances. The results of these tests are presented in figure 33. To our satisfaction, it appears that the two programs achieve a satisfactory degree of agreement, evidenced by the similar predictions in MUF demonstrating at most a difference of 8% in these experiments. The largest difference is observed for receivers close to the transmitter.

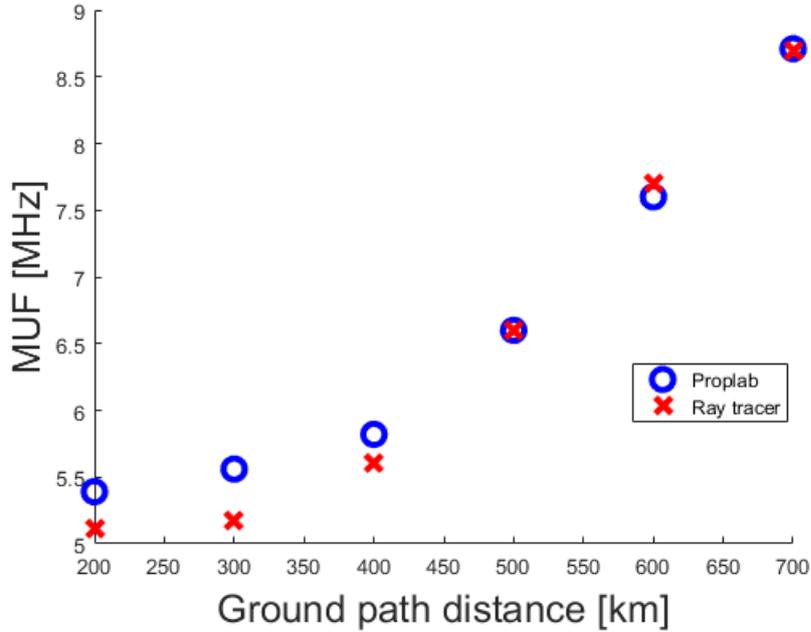


Figure 33: MUFs predicted by Proplab and our ray tracer, evaluated over a range of ground path distances. The two programs achieve satisfactory agreement, with the greatest measured difference being less than 8%.

6.2 Reversibility analysis

In section 4.9 it was argued that proven reversibility of the ray equations (4.20)-(4.21) could be used to validate the employed numerical integrator. Hence, it is necessary to investigate whether the transformation $\mathbf{k} \rightarrow -\mathbf{k}$ results in the ray being propagated back to the transmitter. To better understand how the stepsize Δt affects reversibility, the stepsize is held constant throughout integration for these tests. This way, we may shed some light on the convergence properties of the employed integrator. The test is laid out as follows: An ordinary ray of 5 MHz is transmitted at an elevation angle of 40° from the ground, and allowed to propagate until it strikes the ground some distance away. There, the sign of the wave vector is reversed, and the ray is transmitted in the direction it came from. When the ray strikes the ground again, the great circle distance to the original transmitter is evaluated. This distance is termed the reversibility error. We should expect the ray to strike fairly close to the transmitter, with the reversibility error decreasing with a smaller stepsize. However, should we observe a large reversibility error that does not decrease when Δt is made smaller, the integrator could be suspected of harboring a systematic error.

In all tests, we include both the geomagnetic field and collisions. It is important to note, that the used dipole field is not uniform, thus exhibiting spatial gradients that are derived in appendix B. If the field is non-uniform (like in these tests), the formal ray equations require the evaluation of $\frac{\partial \mathbf{B}_0}{\partial \mathbf{r}}$ along the ray path. Under these conditions, the reversibility error is evaluated as a function of (constant) stepsize

Δt . And indeed, in figure 34 we observe a small reversibility error that appears to decrease with smaller Δt , just as was expected. Moreover, the reversibility error appears to decrease linearly with Δt . When using $\Delta t = 10^{-9}$ s, the reversibility error was measured to be only 13 centimeters, a remarkably small error compared to the accumulated geometric distance of more than 1000 km.

What if we were to approximate $\frac{\partial \mathbf{B}_0}{\partial \mathbf{r}}$ as being zero in the ray equations, although $\mathbf{B}_0(\mathbf{r})$ is allowed to vary with position as in eqn. (2.12)? This is an assumption made in e.g. [33] in order to simplify the ray equations, with the argument that the spatial variations in $\mathbf{Y}(\mathbf{r})$ are small compared to the variations in $X(\mathbf{r})$. The results of this assumption are demonstrated in the rightmost graph of figure 34. First of all, the reversibility error is conspicuously greater than in the default case, respectively 4.9 km versus 1.4 km for the relatively large stepsize of $\Delta t = 10^{-5}$ s. Moreover, the reversibility error does not appear to respond to a decrease in Δt in a monotonous fashion, and no convergence towards a decreasing error is discerned from these results as the reversibility error still remains close to 5 km.

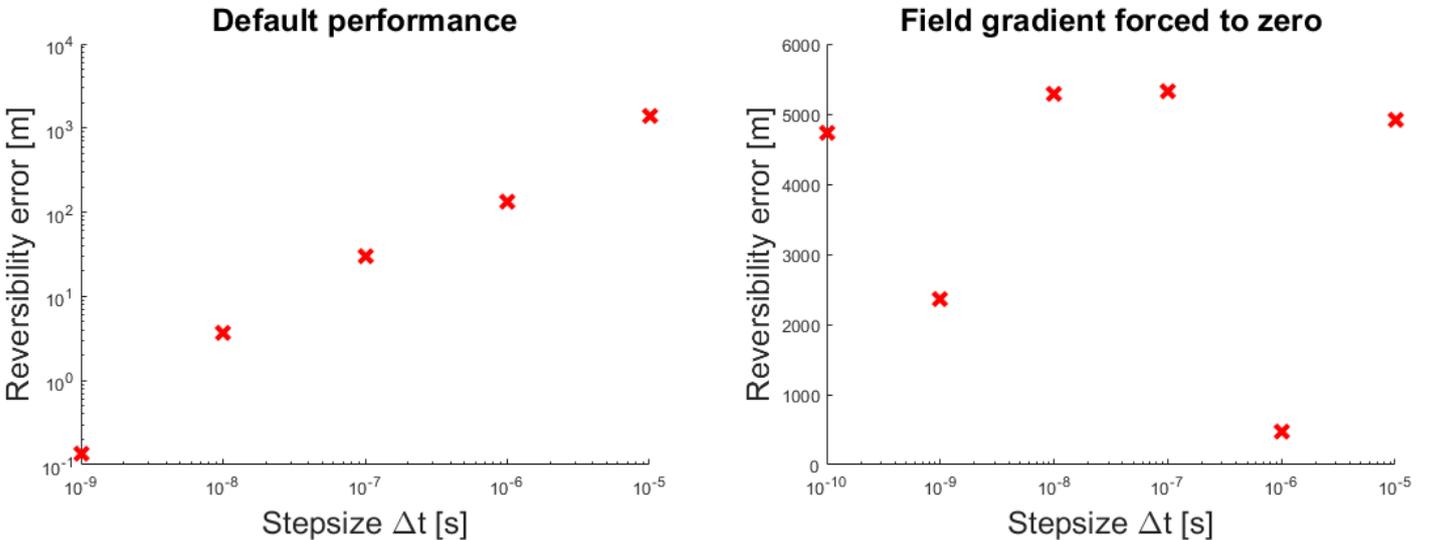


Figure 34: Reversibility error as a function of constant stepsize. When the ray equations are evaluated in the form given in appendix A, the reversibility error appears to decrease linearly with the stepsize (leftmost graph). When the spatial gradients of the geomagnetic field are approximated as being zero (rightmost graph), the reversibility error appears to persist even when the stepsize is decreased. Note, that the vertical axis in the leftmost graph is logarithmic, while the vertical axis is linear in the rightmost graph.

6.3 Attenuation of characteristic modes by absorption

The developed ray tracer accumulates for all rays an attenuation $A_{dB}(t)$ in decibels, consisting of four effects:

1. Gain resulting from the directional properties of the transmitting antenna,

2. magnetoionic splitting loss in the limiting region,
3. absorption loss evaluated by integrating eqn. (4.73), and
4. ground reflection loss.

A representative result for $A_{dB}(t)$ is demonstrated in figure 35, with the respective loss terms labeled in the graph. Transmitting from an isotropic antenna, the gain is equal to unity, resulting in $A_{dB}(0) = 0$ in figure 35. The figure traces an ordinary ray, shown by the discontinuous decrease in A_{dB} at magnetoionic splitting. Inside the ionosphere, the signal is attenuated most in the D region and least in the F region. When reflected from dry soil, the ray experiences a significant ground reflection loss. If the ray were reflected from e.g. saline sea water, the reflection loss would be close to zero. One should remember, that the computed A_{dB} does not include the effects of geometric attenuation, which is in turn estimated by photon mapping in this thesis.

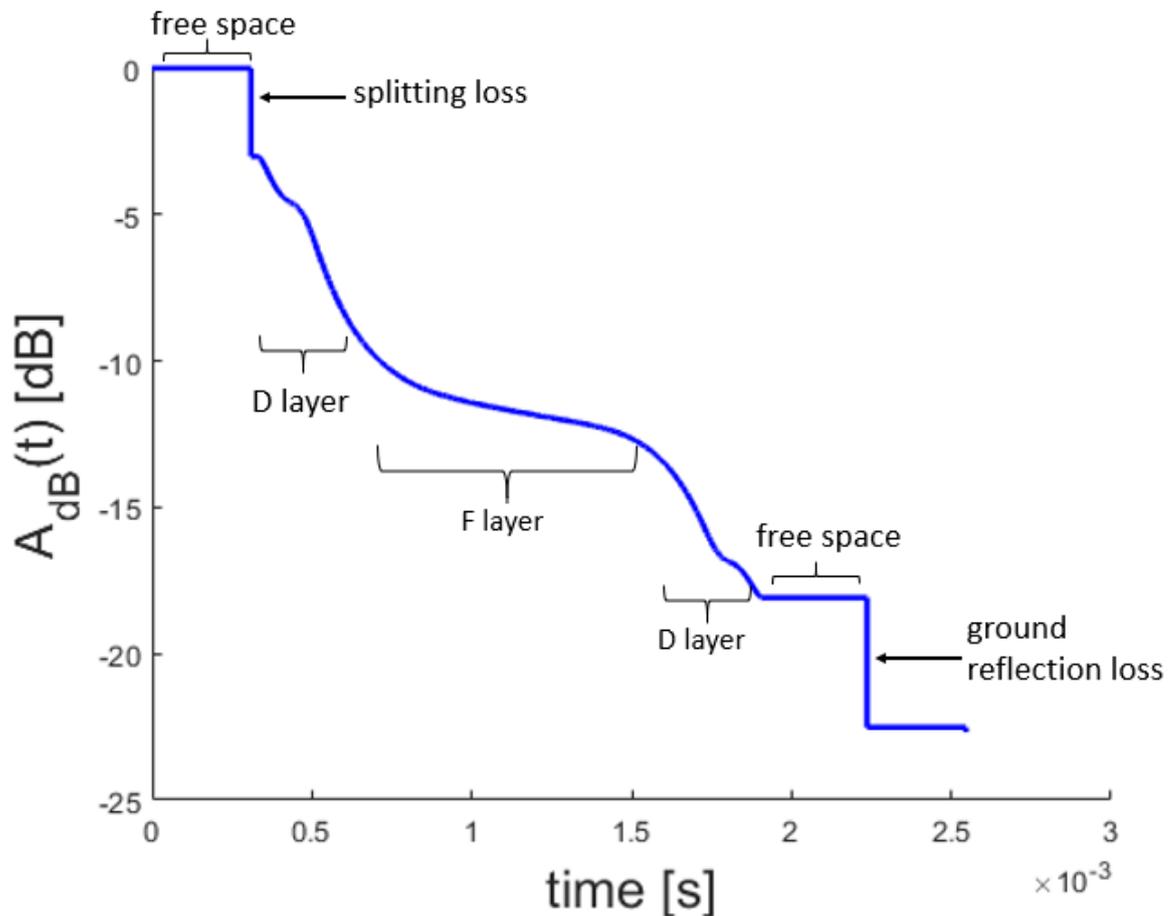


Figure 35: Breakdown of a typical result for the attenuation $A_{dB}(t)$ of an ordinary ray. The traced ray of frequency 5 MHz was launched at an elevation angle of 40° . The ray is attenuated most by absorption in the D region, with the absorption loss being greater the lower the ray frequency. The result would look mostly similar for an extraordinary ray, except that the absorption loss would be greater than for the ordinary ray.

In general, the absorption loss is greatest during the day and lowest during the night. Moreover, low frequencies experience a higher absorption loss in the D region, as evidenced in figure 36. It can be seen, that the extraordinary ray experiences significantly more absorption loss at low frequencies. In practice, this entails that the ordinary mode is sometimes the only viable option when communicating between two distant stations by skywave [30].

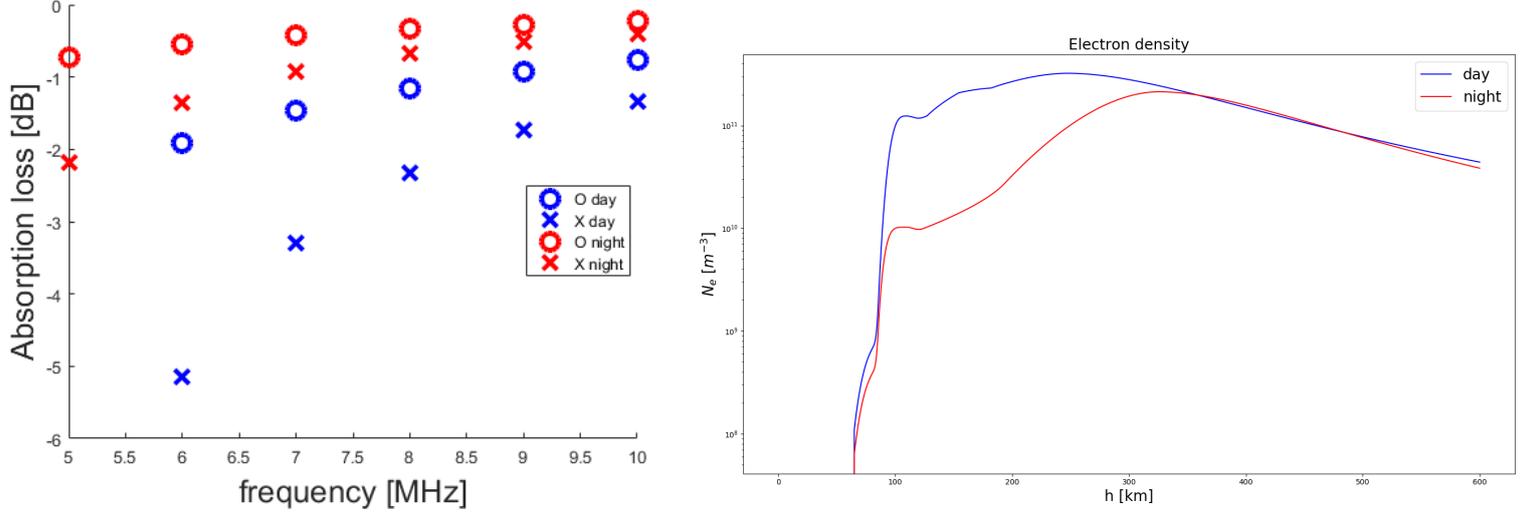


Figure 36: Left: Computed absorption loss of characteristic modes during the day (blue) and night (red). The rays were launched vertically (elevation 90°), and were traced through the ionosphere to an altitude of 600 km. Right: IRI electron density profiles used in computing the absorption loss.

As usual, we would certainly benefit from comparing the computed values of the absorption loss to empirical observations. At the Italian National Institution for Geophysics and Volcanology (INGV), a method of calibrating the AIS-INGV ionosonde for ionospheric attenuation has been devised [24], providing us with empirical data of the absorption loss. The results presented in [24] have been acquired in September 2016 after subtracting the geometric attenuation. However, the vertical sounding was performed using a relatively low frequency of 2.3 MHz: As a result, the vertically incident traced rays are not pseudoreal near reflection. In deriving eqn. (4.73) it has been explicitly assumed that $\chi \ll \mu$, a condition that no longer holds in the soundings presented in [24]. However, for the traced 2.3 MHz rays to pass the pseudoreal check (see section 5.4), the pseudoreal tolerance $\epsilon_{pr} = \max[|\chi/\mu|]$ has to be raised to an almost inexcusably high value of 1. The validity of ray tracing is clearly compromised in these tests, but fortunately, the ray is non-pseudoreal only for a brief moment near reflection, after which the ray continues to propagate without difficulty. Unfazed by these violations, we generate an ionosphere and trace vertically both O- and X-modes at the time and date of the soundings in [24]. When incident on the ground again, both rays have accumulated an attenuation $A_{dB}^{O/X}$ resulting from absorption and a single magnetoionic splitting event. The transmitting antenna was modelled as being isotropic to eliminate the gain factor in attenuation. The total ionospheric attenuation of a sounding is then estimated by the formula

$$A_{dB}^{tot} = 10 \log_{10} \left(10^{A_{dB}^O/10} + 10^{A_{dB}^X/10} \right). \quad (6.1)$$

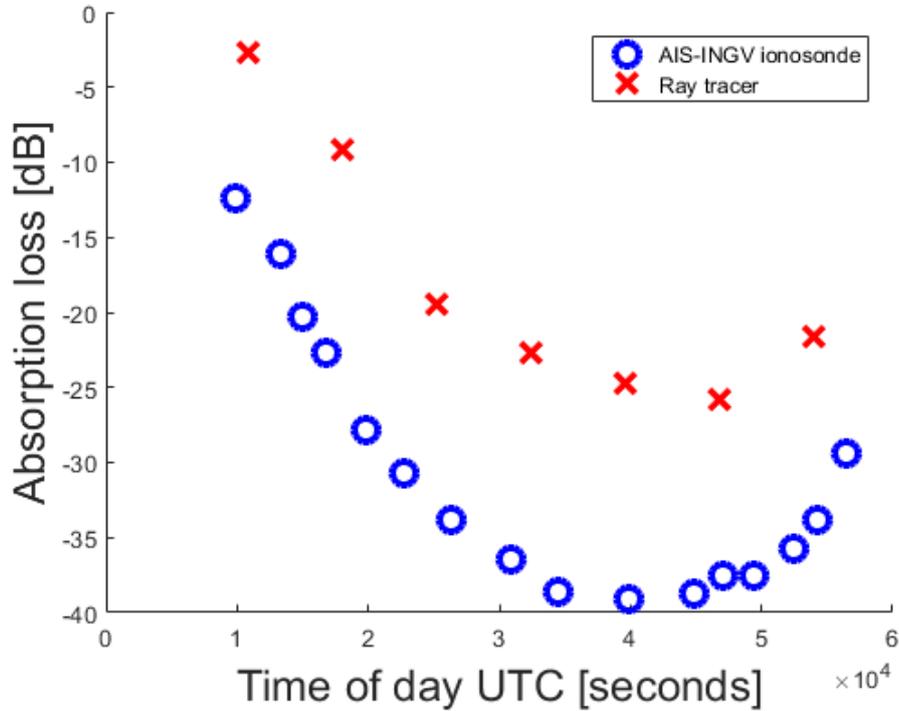


Figure 37: Measured ionospheric attenuation of 2.3 MHz by the AIS-INGV ionosonde [24] versus attenuation predicted by our ray tracer. The shown dataset was recorded 3.9.2016 from 00:00 to 17:00 UTC. The effects of geometric attenuation have been subtracted from the empirical dataset.

The computed ionospheric attenuation has been plotted in figure 37 with the empirical results in [24]. The sounding results, originally given in [24] figure 12, have been median filtered to acquire a smoother dataset. We see, that the ray tracer underestimates ionospheric attenuation by a decibel value of roughly 12, translating to a difference in signal intensity of more than a magnitude. Apart from the error arising from non-pseudoreal rays being traced, the generated electron density and collision frequency could differ from the actual ionosphere above AIS-INGV, partly explaining the discrepancy. However, there is also the likely explanation that the effects of collisions in the D region have not been accounted for correctly. A thorough treatment on calculating ionospheric HF absorption in [36] shows, that the Appleton-Hartree formula underestimates absorption in the D and E regions. The Sen-Wyller dispersion relation [3][36] works demonstrably better in the D and E regions in terms of absorption, but fails to estimate absorption in the F region. In the same paper, it is shown that the absorption per path length $dA_{dB}/d|\mathbf{r}|$ is higher by a factor of 5/2 when using the Sen-Wyller formulae. In light of these results, it would be interesting to evaluate the absorption (4.73) in the D region using the Sen-Wyller dispersion relation, while still using the ray equations derived from the Appleton-Hartree formula. Such an improvement in the program then might better account for the ionospheric attenuation of radio rays.

6.4 Coverage map prediction using photon mapping

This section is dedicated to showcasing some of the results obtained by photon mapping. As described in section 4.8.2, the primary use of photon mapping is to estimate geometric attenuation of a skywave transmission. When incorporating the other attenuation effects described in the previous section, photon mapping allows us to estimate the intensity of a skywave transmission on any target surface.

Photon mapping tests have been conducted using $N = 10^5$ photons and transmission power $P_{rad} = 1$ W, with the transmitter being once again located near Otaniemi at $(60.1^\circ N, 24.8^\circ E)$. The mesh has been constructed in a geographic region encompassing Central and Northern Europe, with the mesh resolution $(\Delta\lambda, \Delta\phi) = (0.6^\circ, 1.3^\circ)$. In the first part of the experiments, we showcase coverage maps predicted for a ground station transmitting at 7 MHz. The photons are allowed to undergo up to two ground reflections before being terminated. In all of the tests described in this section, both the geomagnetic field and collisions are being included in the ray equations. The results of photon mapping for a HF ground station in Otaniemi are shown in figure 38: the value on the colorbar is the decibel intensity I_{dB} , which is related to the real intensity I (in Wm^{-2}) by the formula

$$I = 10^{I_{dB}/10} \quad [\text{Wm}^{-2}]. \quad (6.2)$$

From the results in figure 38, a number of interesting observations can be made. The white areas indicate that no rays are incident on the corresponding mesh elements, resulting in a zero incident intensity. The figure then gives a two-dimensional illustration of the skip zone at this particular frequency. Moreover, the rightmost figure demonstrates the effects of a non-isotropic antenna: the intensity maxima are visibly focused in directions perpendicular to the dipole axes. Lastly, the produced photon map demonstrates local extrema where the incident intensity is visibly higher than at adjacent locations. These extrema suggest optimal receiver locations at this frequency and time of day. However, the extrema are partly due to multipath propagation, where rays incident on the mesh element have been reflected from different regions of the ionosphere (e.g. the E and F regions). The resulting multipath interference could decrease the signal quality at perceived points of maximal intensity in figure 38, rendering some of the extrema non-optimal receiver locations.

Lastly, we study the use of photon mapping for estimating coverage maps of ground stations as seen by a receiver in LEO. We may as well set the photon transmitter to coincide with the satellite position, and launch photons of the desired frequency to discern where on the ground the transmission can be heard. Relying on ray reversibility, we can then conclude with some degree of certainty that if the predicted incident intensity is zero on the ground location, the frequency cannot be used to communicate between the ground station and the satellite. In other words, we can use photon mapping to exclude certain ground stations from a list of plausible transmitters.

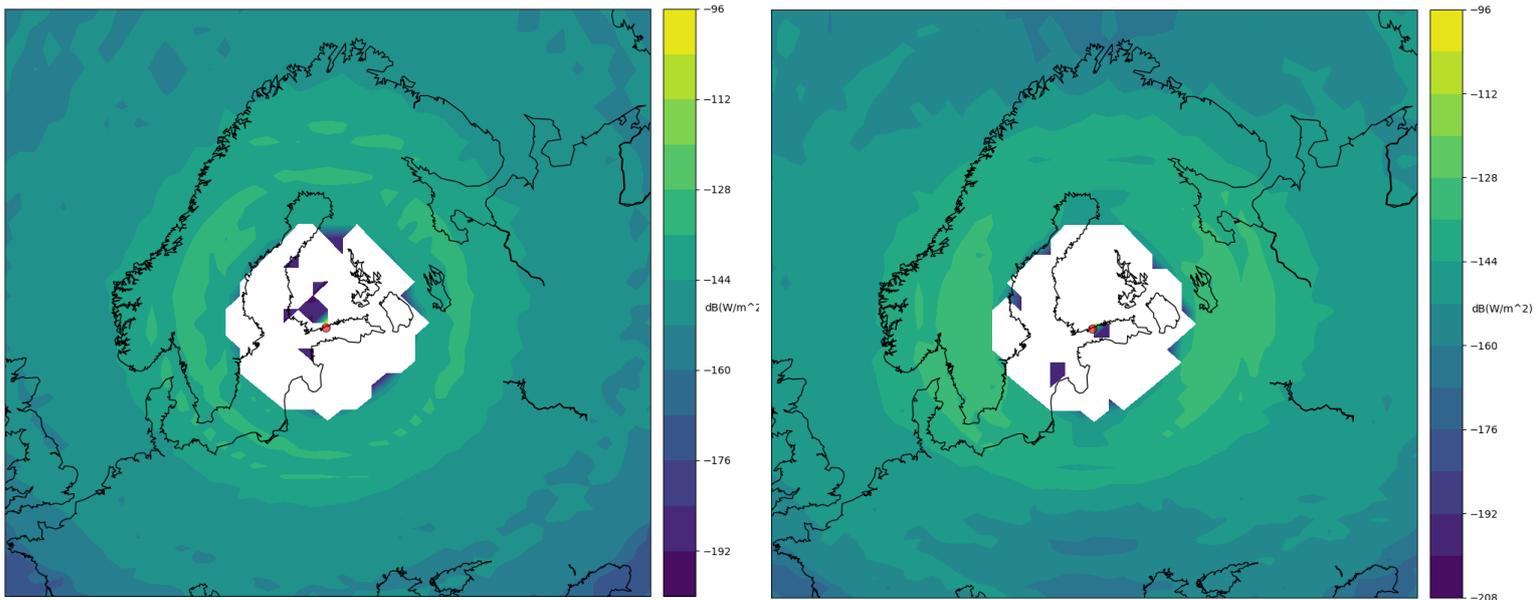


Figure 38: Photon mapping results for a 7 MHz ground station in Otaniemi, with the surface of the Earth being the target surface. The figure on the left represents the result when using an isotropic antenna, while the figure on the right is for a half-wave dipole directed northward (elevation $\alpha = 0^\circ$, bearing $\beta = 0^\circ$). The areas in white are mesh elements with no incident rays, indicating a zero incident intensity.

In figure 39, the satellite position is denoted by a red dot above Otaniemi. In addition to the satellite, the maps show the positions of two hypothetical ground stations as blue dots. At the relatively high frequency of 7 MHz, the rays penetrate the ionosphere more easily. As a result, the predicted coverage area spans a wide geographic region. However, if a frequency only slightly above the F2 critical frequency is used (e.g. 5.5 MHz), only near-vertical rays may penetrate the ionosphere and strike the ground. Consequently, the coverage area is decreased, spanning a geographic region just below the satellite. In the rightmost map of figure 39, the northernmost ground station is not within reach of the satellite at 5.5 MHz, while the ground station closer to Otaniemi may still reach the satellite. However, due to nonreciprocity of EM fields in the ionosphere, we cannot guarantee that the two antennas can communicate with each other, although the antennas are within the coverage regions of each other.

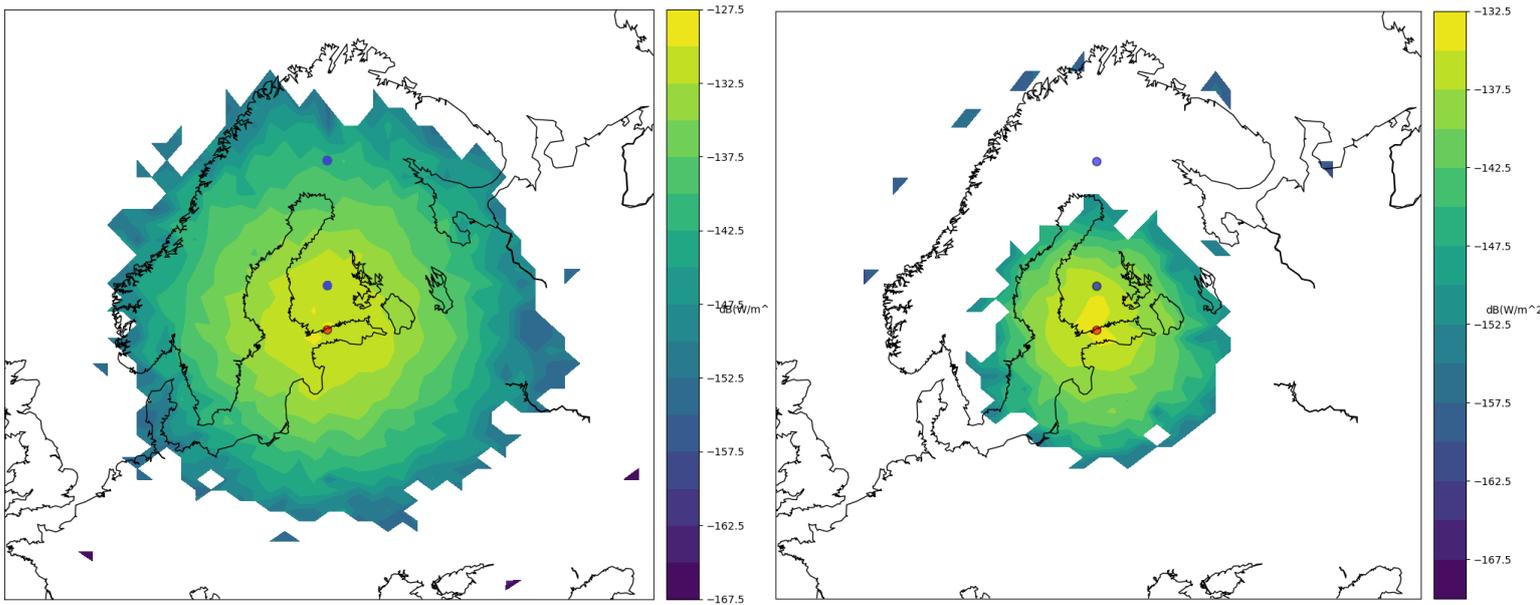


Figure 39: Photon mapping results for a satellite transmitting from an altitude of 575 km above Otaniemi, with the surface of the Earth being the target surface. Results are shown for a transmission frequency of 7 MHz (left) and 5.5 MHz (right). The transmitting antenna is assumed to be isotropic. The satellite ground track position is marked by a red dot. Two hypothetical ground stations have been marked by blue dots.

7 Summary

This thesis has studied the propagation of MF/HF radio waves in the terrestrial ionosphere by means of numerical ray tracing. Starting from the cold plasma approximation, we have derived the Appleton-Hartree dispersion relation and polarization equations of low-amplitude EM waves propagating in the anisotropic ionosphere. Assuming the absorption losses to be small, and further assuming that the dimension of spatial inhomogeneity in the ionosphere greatly exceeds the wavelength – approximations proven to be particularly useful in the case of HF radio waves – we may use geometrical optics to construct the Hamiltonian ray equations. Numerically integrating the ray equations using the adaptive Cash-Karp integrator of the fifth order, we obtain numerical solutions to the position and wave vector of rays transmitted at a given frequency. As the primary product of this thesis, we developed an ionospheric ray tracing program in C++ to study MF/HF radio propagation in practice.

Section 2 has been devoted to discussing the terrestrial ionosphere as a medium of MF/HF radio waves. We recognized the necessity of evaluating three ionospheric quantities, namely the electron number density $N_{e0}(\mathbf{r})$, the geomagnetic field $\mathbf{B}_0(\mathbf{r})$, and the electron collision frequency $\nu_{eff}(\mathbf{r})$. In the current version of the ray tracer, the electron density and collision frequency are procedurally generated using IRI-2016 and NRLMSISE-00 as one-dimensional altitude profiles, while the geomagnetic field is allowed to exhibit three-dimensional spatial variation in the form of a dipole magnet. The value of ν_{eff} has been taken to be the Schunk-Nagy electron-neutral collision frequency in the D region, the inclusion of which permits a mechanism for computing signal attenuation by absorption.

In section 3, we studied the properties of low-amplitude cold plasma waves in the ionosphere. Knowing that the MF/HF frequencies of interest greatly exceed the ionospheric ion gyrofrequency, we deemed it adequate to solely include electrons in the derivation of dispersion relations. The cold plasma approximation presented us with two suitable dispersion relations to be used as ray Hamiltonians; the Booker quartic equation, and the Appleton-Hartree dispersion relation. Moreover, anisotropy introduced by the geomagnetic field was shown to allow only two modes of propagation, commonly known as the ordinary and extraordinary modes. In order to decompose a transmitted signal into these modes by its initial polarization, we derived the formulae for computing the perpendicular and longitudinal polarizations of the magnetoionic wave modes. Moreover, we demonstrated how the wave may interchange its mode during propagation, and presented an algorithm for detecting wave mode interchange.

Section 4 discussed the essentials of the employed ray tracing formalism. We derived the Hamiltonian ray equations, which are numerically integrated in time to yield a solution to the ray path. The ray equations are generated by a ray Hamiltonian, an implicit relation signifying a physical condition to be satisfied along the ray path. It was shown, that a versatile ray tracer necessitates the use of at least two Hamiltonians: Since all the derivatives of the Booker quartic are zero in so-called degenerate plasma regions, the ray tracer needs to be supplemented with a Hamiltonian

whose derivatives are defined in degenerate regions as well, a requirement found to be satisfied by the Appleton-Hartree Hamiltonian. However, the Appleton-Hartree dispersion relation may be indeterminate at points where the plasma electron frequency equals the wave frequency. As a consequence, the developed ray tracer is instructed to switch between Hamiltonians in the middle of the path to circumvent problematic regions in parameter space.

Section 4 also details the methods of computing signal attenuation. The thesis documents four common antenna types, from which the initial ray polarization and transmission power can be evaluated. We present briefly the theory of limiting polarization, and how the attenuation by magnetoionic splitting can be computed in the limiting region. The developed ray tracer evaluates magnetoionic splitting loss by matching the transmitted polarization to the polarization ellipse of the selected magnetoionic mode (ordinary or extraordinary), as explained in section 4.7.2. However, we found that the formula used in evaluating the splitting loss (eqn. (4.60)) appears to violate the conservation of energy, as the sum signal amplitude of the ordinary and extraordinary modes was found in some cases to exceed the incident signal amplitude. This unfortunate discrepancy requires further theoretical investigation into applying the theory of limiting polarization (cf. section 4.7.1) to ray theory.

In addition, the developed ray tracer implements a method of estimating the signal loss and polarization shift at ground reflection. Having assumed the refractive index to be mostly real with only a small imaginary part, we derived an ODE to be integrated for estimating signal loss by absorption. Lastly, we have proposed a Monte Carlo method akin to photon mapping for estimating the geometric attenuation of a transmission. Launching a large number of photons in random directions, the photon paths are solved by ray tracing, simultaneously accumulating attenuation by absorption and other ionospheric effects. The strike-points on target surfaces are recorded, and the statistics of all strike-points are used to estimate the incident intensity on a discrete mesh.

We are urged to remember, that the ray tracing formalism employed in this thesis is subject to a substantial number of approximations and assumptions, some of which may be violated in the obtained numerical solutions unless special care is used. In particular, the presence of electron collisions introduces losses in the ionosphere that manifest as complex-valued dispersion relations. The geometrical optics formalism underlying ray tracing in real space requires, that the absorption per path is small enough to allow us to only use the real part of the Hamiltonian. This was referred to as the assumption of pseudoreal rays, and has proven to hold demonstrably well when tracing HF rays. However, the problems are accentuated when tracing rays of lower frequencies (such as MF waves), which often results in the pseudoreal assumption to be violated. Furthermore, geometrical optics fundamentally necessitates that the traced rays are of short wavelengths, a condition that is again more likely to be violated when propagating MF or lower frequency waves. The developed ray tracer has been equipped with diagnostics to detect rays that violate these conditions during propagation, warning the user and terminating tracing if such stopping conditions are encountered.

Section 5 covered the software engineering aspects of the developed ray tracing program. We discussed how mixed-language programming was utilized to incorporate an ionosphere model (IRI-2016) and a neutral atmosphere model (NRLMSISE-00) into the ray tracing program. We also elaborated on the possible inputs and outputs of the developed program, and how certain inputs (like low frequency or high elevation angle) can compromise the physical validity of the solved ray paths. Lastly, we discussed the methods of CPU parallelization, and presented benchmark performances for the program. The execution time of photon mapping was demonstrated to scale linearly with an increasing number of photons as expected. Moreover, the throughput of ray tracing and photon mapping was found to improve efficiently using multithread parallelization, although hyper-threading was not found to improve the performance on our system.

In section 6, we presented some results obtained with the developed ray tracer. We compared the solved ray paths to those of a commercial HF radio analysis program, Proplab Pro 3, and found that the final ray positions over a single hop could differ by several tens of kilometers. We hypothesized that the observed difference could be explained by three factors, namely (i) the programs are using a different standard of IRI to generate the ionosphere, (ii) the programs are using different geomagnetic field models, and (iii) the ionosphere used by Proplab varies in three dimensions, while the ionosphere used by our ray tracer only varies with altitude. The former assertion was supported by using the same mid-path ionospheric profile in both programs, which caused the difference in final ray positions to decrease by nearly 62%. The latter assertion was corroborated by observing, that the ray path solved by Proplab Pro 3 had undergone lateral deviation, a phenomenon that cannot occur in a vertically varying and isotropic ionosphere. Since the current version of the developed ray tracer uses vertically varying ionospheric profiles that are generated by default above the transmitter, any obtained ray paths inevitably suffer a hit in accuracy when traced far away from the transmitter. Arguably the most pressing need for improvement is in implementing a three-dimensional ionosphere for future versions of the ray tracer, in which N_{e0} and ν_{eff} are allowed to vary latitudinally and longitudinally as well. However, despite the fact that the developed ray tracer uses a one-dimensional ionosphere, predictions of the maximal usable frequency (MUF) were found to agree remarkably well with those of Proplab Pro 3, with the greatest difference in MUF being only 8% between the two programs.

Analysis of ionospheric attenuation in section 6.3 suggests, that the developed ray tracer underestimates absorption loss. We arrived to this conclusion by comparing the computed ray absorption to empirical ionosonde observations of 2.3 MHz sounding experiments. The difference in absorption loss was found to be roughly 12 decibels by average, corresponding to a discrepancy in signal intensity of more than a full order of magnitude. For a more reliable analysis, we are urged to repeat the comparison using waves of higher frequency, since the low ionosonde frequency of 2.3 MHz was found to clearly violate the pseudoreal assumption. However, there are several reasons to suspect that the current ray tracer would still underestimate absorption. First of all, the Appleton-Hartree refractive index used by our ray tracer has been demonstrated to underestimate absorption in the D region. One possible

solution is suggested in using the Sen-Wyller refractive index, which has been found to provide more accurate estimates for the absorption loss in the D region. Secondly, the used Schunk-Nagy electron-neutral collision frequency cannot be guaranteed to be valid beyond the D region, and higher up in the E and F regions, we would need to incorporate electron-ion collision processes as well. In the E and F regions, we are also well-advised to include more neutral particle species (such as oxygen atoms) in the collision mechanism. Fortunately, the NRLMSISE-00 atmosphere model used by the ray tracer is capable of also estimating the densities of oxygen and nitrogen atoms in the ionosphere, greatly facilitating such an improvement.

Finally, we presented results obtained by photon mapping when predicting radio coverage maps. The results were used to demonstrate a two-dimensional representation of the MUF and skip zone, as well as to highlight the effects of non-isotropic transmitters. It would be interesting to compare the coverage maps predicted by photon mapping to empirically-obtained link quality data in the field. Since the program is sought for data analysis of Suomi 100 radio measurements, further empirical validation of the developed ray tracer is certainly encouraged. Although the photon mapping method is fundamentally used to estimate geometric attenuation, the current version of the program inevitably loses some of the transmission power due to the selected strategy of magnetoionic splitting. After the first ground reflection, the program selects the strongest mode (i.e. the mode with lowest splitting loss) to be traced, discarding the weaker mode and thus irretrievably losing some small fraction of the power on each split. This strategy was selected to make the computation time significantly shorter; formally each splitting event following ground reflection generates two new rays, cumulatively generating $2^{(m+1)}$ rays at most to be traced after m ground reflections. By only tracing the strongest mode, this exponential increase in rays is averted at the cost of some small unphysical signal loss. We posited, that the splitting transmission coefficient is usually close to unity for one mode and zero for the other mode, given that the ray has already emerged once from the limiting region. Such a claim would require further investigations, on the basis of which it could be considered to always trace both rays after splitting.

Development of the ray tracing program was originally motivated by the Suomi 100 satellite and its AM radio instrument. It has been proposed, that the ray tracer could be used in analysis of the satellite in situ radio measurements. As an example, the photon mapping method analyzed in section 6.4 provides a method of estimating the coverage of a certain frequency, enabling us to possibly deduce the origins of certain received signals. Moreover, the ray tracer could be used in validation of HF radar measurements in ionospheric research; ray tracing allows the user to visualize the paths of HF sounding waves, as well as to estimate the attenuation of transmitted waves. On a more general level, the ray tracing program developed for this thesis could be used in conjunction with an HF prediction service, allowing users to estimate MUFs for a certain date and time.

References

- [1] C. Altman and K. Suchy, *Reciprocity, Spatial Mapping and Time Reversal in Electromagnetics*. Kluwer Academic Publishers 1991.
- [2] T.J.M. Boyd and J.J. Sanderson, *Plasma Dynamics*. Thomass Nelson and Sons Ltd. 1969.
- [3] K.G. Budden, *The Propagation of Radio Waves: The Theory of Radio Waves of Low Power in the Ionosphere and Magnetosphere*. Cambridge University Press 1985.
- [4] A.L. Fetter and J.D. Walecka, *Theoretical Mechanics of Particles and Continua*. Dover Publications 2015.
- [5] V.L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas*. Translated by J.B. Sykes and R.J. Tayler, Pergamon Press Ltd. 1964. Originally published as *Rasprostraneniye Elektromagnitnykh Voln v Plazme*, Fizmatgiz 1960.
- [6] D.F. Griffiths and D.J. Higham, *Numerical Methods for Ordinary Differential Equations*. Springer 2010.
- [7] J. Haselgrove, "*Ray theory and a new method for ray tracing*". The physics of the ionosphere, pp. 355-64. Physical Society, London, 1954.
- [8] R. Hunsucker and J. Hargreaves, *The High-Latitude Ionosphere and its Effects on Radio Propagation*. Cambridge University Press, 2003.
- [9] H.W. Jensen, *Realistic Image Synthesis Using Photon Mapping*. Peters 2001.
- [10] M. Kivelson and C. Russell, *Introduction to Space Physics*. Cambridge University Press, 1995.
- [11] H. Koskinen, *Johdatus plasmafysiikkaan ja sen avaruussovellutuksiin*. Limes ry. 2001.
- [12] L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields*. Translated by M. Hamermesh. Fourth revised English edition, Elsevier 2010. Originally published as *Teoriya Polya*, Teoreticheskaya Fizika 1941.
- [13] A. Lipson and S.G. Lipson, H. Lipson, *Optical Physics*. Cambridge University Press 2011, 4th edition.
- [14] W.H. Press and S.A. Teukolsky et al., *Numerical Recipes*. Second edition, Cambridge University Press 1992.
- [15] T. Rauber and G. Runger, *Parallel Programming for Multicore and Cluster Systems*. Springer 2010.

- [16] R. Schunk and A. Nagy, *Ionospheres*. Cambridge University Press, 2nd edition, 2009.
- [17] H.W. Silver, *The ARRL Handbook for Radio Communications 2013*. The American Radio Relay League, 2013.
- [18] T.H. Stix, *Waves in Plasmas*. American Institute of Physics, 1992.
- [19] W.L. Stutzman and G.A. Thiele, *Antenna Theory and Design*. John Wiley and Sons Inc., third edition, 2012.
- [20] A. Azzarone and C. Bianchi et al., "*IONORT: A Windows software tool to calculate the HF ray tracing in the ionosphere*". Computers & Geosciences 42, pp. 57-63, 2012.
- [21] P.A. Bradley, "*Ingesting a sporadic-E model to IRI*". Advances in Space Research, Vol. 31, Issue 3, pp. 577-588, 2003.
- [22] J.R. Cash and A.H. Karp, "*A variable order Runge-Kutta method for initial value problems with rapidly varying right-hand sides*". ACM Transactions on Mathematical Software 16, pp. 201-222, 1990.
- [23] L. Friedland and I.B. Bernstein, "*General Geometric Optics Formalism in Plasmas*". IEEE Transactions on Plasma Sciences, PS-8 90, 1980.
- [24] L. Gilli, U. Sciacca and E. Zuccheretti, "*Calibrating an Ionosonde for Ionospheric Attenuation Measurements*". Sensors, Vol. 18, Issue 5, 2018.
- [25] International Telegraph Union (ITU), "*Electrical Characteristics of the Surface of the Earth*". Recommendation ITU-R P.527-3, 1978-1982-1990-1992.
- [26] R.M. Jones and J.J. Stephenson, *A Versatile Three-Dimensional Ray Tracing Computer Program for Radio Waves in the Ionosphere*. U.S. Department of Commerce, Office of Telecommunication, OT Report 75-75, 1975.
- [27] P.J. Koskimaa, *Ferrite Rod Antenna in a Nanosatellite Medium and High Frequency Radio*. Master's thesis, Aalto University School of Electrical Engineering, 2016.
- [28] J.S.W. Lamb and J.A.G. Roberts, "*Time-reversal symmetry in dynamical systems: A survey*". Elsevier Physica D, Vol. 112, pp. 1-39, 1998.
- [29] J.D. Mathews, "*Sporadic E: current views and recent progress*". Journal of Atmospheric and Solar-Terrestrial Physics, Vol. 60, Issue 4, pp. 413-435, 1998.
- [30] G.J. Phillips and P. Knight, "*Effects of Polarization on a Medium-Frequency Sky-Wave Service, Including the Case of Multi-Hop Paths*". British Broadcasting Corporation (BBC) engineering division, research report No. G-083/2, 1964.

- [31] M.T. Rietveld et al., "*The Tromso dynasonde*". Polar Science, Volume 2, Issue 1, pp. 55-71, 2008.
- [32] M.S. Smith, "*Non-reversibility for radio rays in the ionosphere*". Journal of Atmospheric and Terrestrial Physics, Vol. 38, pp. 37-44, 1976.
- [33] D. Sonnenschein and D. Censor et al., "*Ray trajectories in an absorbing ionosphere*". Journal of Atmospheric and Terrestrial Physics, Vol. 59, No. 16, pp. 2101-2110, 1997.
- [34] A. Västberg and B. Lundborg, "*Signal intensity in the geometrical optics approximation for the magnetized ionosphere*". Radio Science, Vol. 31, No. 5, pp. 1579-1588, 1996.
- [35] K. Yee, "*Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media*". IEEE Transaction on Antennas and Propagation, Vol. 14, No. 3, pp. 302-307, 1966.
- [36] K.A. Zawdie et al., "*Calculating the absorption of HF radio waves in the ionosphere*". Radio Science, Vol. 52, Issue 6, 2017.
- [37] The CubeSat standard. Referenced 23.10.2018. Available at: <http://www.cubesat.org/>.
- [38] European Incoherent Scatter Scientific Association (EISCAT). Referenced 16.10.2018. Available at: <https://www.eiscat.se/>.
- [39] National Aeronautics and Space Administration (NASA): *NRLMSISE-00 Model 2001*. Referenced 16.10.2018. Available at: <https://ccmc.gsfc.nasa.gov/modelweb/atmos/nrlmsise00.html>.
- [40] National Aeronautics and Space Administration (NASA) : *International Reference Ionosphere*. Referenced 16.10.2018. Available at: <http://irimodel.org>.
- [41] National Oceanic and Atmospheric Administration (NOAA) : *The US/UK World Magnetic Model for 2015-2020*. Referenced 16.10.2018. Available at: <https://www.ngdc.noaa.gov/geomag/WMM/DoDWMM.shtml>.
- [42] NVIDIA, CUDA parallel computing API specification. Referenced 6.11.2018. Available at: <https://developer.nvidia.com/cuda-zone>.
- [43] Online C++11 standard library documentation. Referenced 6.11.2018. Available at: <https://en.cppreference.com/w/>.
- [44] OpenMP parallel computing API specification. Referenced 6.11.2018. Available at: <https://www.openmp.org/>.
- [45] Sodankylä Geophysical Observatory, the Sodankylä Ionosonde. Referenced 23.10.2018. Available at: <http://www.sgo.fi/Data/Ionosonde/ionosonde.php>.

- [46] Solar Terrestrial Dispatch, Proplab-Pro version 3. Referenced 23.10.2018. Available at: <http://www.spacew.com/www/proplab.html>.
- [47] E.W. Weisstein, "*Sphere Point Picking*". From MathWorld – A Wolfram Web Resource. Referenced 14.11.2018. Available at: <http://mathworld.wolfram.com/SpherePointPicking.html>.
- [48] Git repository of the developed ray tracing program at version.aalto.fi. Requires an IT account at Aalto University. Referenced 14.11.2018. Available at: <https://version.aalto.fi/gitlab/multi-plasma/raytracer>.

A Derivatives of the Hamiltonian

In this appendix, derivatives of the Hamiltonians are documented. Although many of the derivatives have been presented as well in [26], the aim of this thesis is to provide a stand-alone document for developing an ionospheric ray tracer, and hence, it is in our best interest to re-derive the results here.

The vector subscripts $i = \{1, 2, 3\}$ denote the basis vectors ($\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$) of the particular coordinate system used, which may be Cartesian or curvilinear. For example, the ray tracer developed for this thesis uses exclusively the familiar Cartesian basis vectors ($\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z$), while in e.g. [26] the local spherical basis vectors ($\hat{\mathbf{e}}_r, \hat{\mathbf{e}}_\theta, \hat{\mathbf{e}}_\phi$) are used in ray tracing. The ray tracer developed for this thesis uses Cartesian coordinates owing to their innate simplicity when transforming vectors to another coordinate system.

For ray tracing in anisotropic media, spatial derivatives of the geomagnetic field \mathbf{B}_0 are required. These are given in appendix B for a dipole field model.

Parameters and derivatives common to all Hamiltonians are presented below:

$$\omega_p = \sqrt{\frac{N_{e0}e^2}{m_e\varepsilon_0}} \quad (\text{A.1})$$

$$\Omega_c = \frac{e|\mathbf{B}_0|}{m_e} \quad (\text{A.2})$$

$$\nu_{N_2} = 2.33 \times 10^{-17} N_{N_2} (1 - 1.21 \times 10^{-4} T_e) T_e \text{ [Hz]} \quad (\text{A.3})$$

$$\nu_{O_2} = 1.82 \times 10^{-16} N_{O_2} (1 + 3.6 \times 10^{-2} \sqrt{T_e}) \sqrt{T_e} \text{ [Hz]} \quad (\text{A.4})$$

$$\nu_{eff} = \nu_{en} = \frac{N_{N_2}\nu_{N_2} + N_{O_2}\nu_{O_2}}{N_{N_2} + N_{O_2}} \quad (\text{A.5})$$

$$X = \frac{\omega_p^2}{\omega^2} \quad (\text{A.6})$$

$$Y = \frac{\Omega_c}{\omega} \quad (\text{A.7})$$

$$Z = \frac{\nu_{eff}}{\omega} \quad (\text{A.8})$$

$$U = 1 + iZ \quad (\text{A.9})$$

$$\mathbf{Y} = \frac{e\mathbf{B}_0}{m_e\omega} = Y_1\hat{\mathbf{e}}_1 + Y_2\hat{\mathbf{e}}_2 + Y_3\hat{\mathbf{e}}_3 \quad (\text{A.10})$$

$$\frac{\partial X}{\partial x_i} = \frac{e^2}{\varepsilon_0 m_e \omega^2} \frac{\partial N_{e0}}{\partial x_i} \quad (\text{A.11})$$

$$\frac{\partial Y_j}{\partial x_i} = \frac{e}{m_e \omega} \frac{\partial (\mathbf{B}_0)_j}{\partial x_i} \quad (\text{A.12})$$

The expression $(\mathbf{B}_0)_j$ denotes the component j of the geomagnetic field \mathbf{B}_0 .

$$\frac{\partial U}{\partial x_i} = i \frac{1}{\omega} \frac{\partial \nu_{eff}}{\partial x_i} \quad (\text{A.13})$$

$$\frac{\partial X}{\partial \omega} = - \frac{2X}{\omega} \quad (\text{A.14})$$

$$\frac{\partial Y_i}{\partial \omega} = - \frac{Y_i}{\omega} \quad (\text{A.15})$$

$$\frac{\partial U}{\partial \omega} = i \frac{\partial Z}{\partial \omega} = -i \frac{Z}{\omega} \quad (\text{A.16})$$

$$\frac{\partial (\mathbf{k} \cdot \mathbf{Y})^2}{\partial \omega} = -2 \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{\omega} \quad (\text{A.17})$$

A.1 Appleton-Hartree formula

From the Appleton-Hartree formula for the refractive index, the Appleton-Hartree Hamiltonian is defined as [26]

$$H = \text{Re} \left[\frac{c^2}{\omega^2} (k_1^2 + k_2^2 + k_3^2) - n^2 \right]. \quad (\text{A.18})$$

The components k_i of the wave vector are taken as real-valued quantities. Thus, the angle θ in the Appleton-Hartree formula is conveniently expressed by the dot product below:

$$\cos \theta = \frac{(\mathbf{k} \cdot \mathbf{Y})}{kY} \quad (\text{A.19})$$

We can rewrite the trigonometric functions in the Appleton-Hartree formula using the dot product above and the fundamental relation $\sin^2 \theta = 1 - \cos^2 \theta$:

$$n^2 = 1 - \frac{X(U - X)}{Q} \quad (\text{A.20})$$

$$Q = U(U - X) - \frac{1}{2} \left(Y^2 - \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \right) + S_R \quad (\text{A.21})$$

$$S_R = \pm \left[\frac{1}{4} \left(Y^2 - \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \right)^2 + (U - X)^2 \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \right]^{1/2} \quad (\text{A.22})$$

From here, the ray tracer branches to use one of four Appleton-Hartree Hamiltonians, depending on whether the geomagnetic field or collisions are included in the model. In the following sections, we present the complete set of equations for the most general case, i.e. the AHFWC-Hamiltonian, from which the rest can be obtained simply by dropping some terms.

A.1.1 With field and with collisions (AHFWC)

The most general of the Appleton-Hartree Hamiltonians is obtained by allowing X , \mathbf{Y} and U to vary. The derivatives used by the ray tracer are presented below.

$$\frac{\partial Q}{\partial X} = -U + S_R^{-1}(X - U) \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \quad (\text{A.23})$$

$$\begin{aligned} \frac{\partial Q}{\partial Y_i} = & -Y_i + \frac{(\mathbf{k} \cdot \mathbf{Y})}{k^2} k_i \quad (\text{A.24}) \\ & + \frac{1}{2} S_R^{-1} \left[\left(Y^2 - \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \right) \left(Y_i - \frac{(\mathbf{k} \cdot \mathbf{Y})}{k^2} k_i \right) + 2(U - X)^2 \frac{(\mathbf{k} \cdot \mathbf{Y})}{k^2} k_i \right] \end{aligned}$$

$$\frac{\partial Q}{\partial U} = 2U - X + S_R^{-1}(U - X) \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \quad (\text{A.25})$$

$$\frac{\partial n^2}{\partial X} = \frac{2X - U}{Q} + \frac{X(U - X)}{Q^2} \frac{\partial Q}{\partial X} \quad (\text{A.26})$$

$$\frac{\partial n^2}{\partial Y_i} = \frac{X(U - X)}{Q^2} \frac{\partial Q}{\partial Y_i} \quad (\text{A.27})$$

$$\frac{\partial n^2}{\partial U} = -\frac{X}{Q} + \frac{X(U-X)}{Q^2} \frac{\partial Q}{\partial U} \quad (\text{A.28})$$

$$\frac{\partial n^2}{\partial x_i} = \frac{\partial n^2}{\partial X} \frac{\partial X}{\partial x_i} + i \frac{\partial n^2}{\partial U} \frac{\partial Z}{\partial x_i} + \sum_{j=1}^3 \frac{\partial n^2}{\partial Y_j} \frac{\partial Y_j}{\partial x_i} \quad (\text{A.29})$$

$$\frac{\partial H}{\partial x_i} = \text{Re} \left[-\frac{\partial n^2}{\partial x_i} \right] \quad (\text{A.30})$$

$$\begin{aligned} f_{k_i} &= \frac{1}{2} \frac{\partial}{\partial k_i} \left(\frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \right) \\ &= \frac{(\mathbf{k} \cdot \mathbf{Y})}{k^2} Y_i - \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^4} k_i \end{aligned} \quad (\text{A.31})$$

$$\frac{\partial Q}{\partial k_i} = f_{k_i} \left(1 + S_R^{-1} \left[-\frac{1}{2} \left(Y^2 - \frac{(\mathbf{k} \cdot \mathbf{Y})^2}{k^2} \right) + (U - X)^2 \right] \right) \quad (\text{A.32})$$

$$\begin{aligned} \frac{\partial H}{\partial k_i} &= \text{Re} \left[\frac{2c^2}{\omega^2} k_i - \frac{\partial n^2}{\partial k_i} \right] \\ &= \text{Re} \left[\frac{2c^2}{\omega^2} k_i - \frac{X(U-X)}{Q^2} \frac{\partial Q}{\partial k_i} \right] \end{aligned} \quad (\text{A.33})$$

$$\begin{aligned} \frac{\partial n^2}{\partial \omega} &= \frac{\partial n^2}{\partial X} \frac{\partial X}{\partial \omega} + \frac{\partial n^2}{\partial U} \frac{\partial U}{\partial \omega} + \sum_{j=1}^3 \frac{\partial n^2}{\partial Y_j} \frac{\partial Y_j}{\partial \omega} \\ &= -2 \frac{X}{\omega} \frac{\partial n^2}{\partial X} - i \frac{Z}{\omega} \frac{\partial n^2}{\partial U} - \frac{1}{\omega} \sum_{j=1}^3 Y_j \frac{\partial n^2}{\partial Y_j} \end{aligned} \quad (\text{A.34})$$

$$\frac{\partial H}{\partial \omega} = \text{Re} \left[-\frac{2c^2}{\omega^3} k^2 - \frac{\partial n^2}{\partial \omega} \right] \quad (\text{A.35})$$

A.1.2 With field and no collisions (AHWFNC)

When collisions are not included, the equations are the same as for AHWFNC, except that

$$Z = 0, \quad U = 1, \quad \frac{\partial Z}{\partial x_i} = 0. \quad (\text{A.36})$$

A.1.3 No field and with collisions (AHNFWC)

In isotropic media, $Y = 0$, and consequently $S_R = 0$. The Appleton-Hartree refractive index is then greatly simplified to the form

$$n^2 = 1 - \frac{X}{U}. \quad (\text{A.37})$$

Consequently, the derivatives of the Hamiltonians are greatly simplified, and are given as follows:

$$\begin{aligned} \frac{\partial n^2}{\partial x_i} &= \frac{\partial n^2}{\partial X} \frac{\partial X}{\partial x_i} + \frac{\partial n^2}{\partial U} \frac{\partial U}{\partial x_i} \\ &= -\frac{1}{U} \frac{\partial X}{\partial x_i} + i \frac{X}{U^2} \frac{\partial Z}{\partial x_i} \end{aligned} \quad (\text{A.38})$$

$$\frac{\partial H}{\partial x_i} = \text{Re} \left[-\frac{\partial n^2}{\partial x_i} \right] \quad (\text{A.39})$$

$$\frac{\partial H}{\partial k_i} = \frac{2c^2}{\omega^2} k_i \quad (\text{A.40})$$

$$\begin{aligned} \frac{\partial n^2}{\partial \omega} &= \frac{\partial n^2}{\partial X} \frac{\partial X}{\partial \omega} + \frac{\partial n^2}{\partial U} \frac{\partial U}{\partial \omega} \\ &= \frac{1}{\omega} \left[\frac{2X}{U} - i \frac{XZ}{U^2} \right] \end{aligned} \quad (\text{A.41})$$

$$\frac{\partial H}{\partial \omega} = \text{Re} \left[-\frac{2c^2}{\omega^3} k^2 - \frac{\partial n^2}{\partial \omega} \right] \quad (\text{A.42})$$

A.1.4 No field and no collisions (AHNFNC)

Further, when neither collisions nor the magnetic field is included, the equations are the same as for AHNFWC except that

$$Z = 0, \quad U = 1, \quad \frac{\partial Z}{\partial x_i} = 0. \quad (\text{A.43})$$

Under these conditions, the dispersion relation can be simply expressed by $\omega^2 = \omega_p^2 + c^2 k^2$ [11]. Using the aforementioned dispersion relation and definition (A.1) of ω_p , the ray equations are simplified to the intuitively appealing form:

$$\frac{\partial H}{\partial \omega} = -\frac{2}{\omega} \quad (\text{A.44})$$

$$\frac{\partial H}{\partial k_i} = \frac{2c^2}{\omega^2} k_i \quad (\text{A.45})$$

$$\frac{\partial H}{\partial x_i} = \frac{e^2}{\varepsilon_0 m_e \omega^2} \frac{\partial N_{e0}}{\partial x_i} \quad (\text{A.46})$$

These relations then give the ray equations in isotropic and collisionless plasmas:

$$\dot{\mathbf{r}} = \frac{c^2}{\omega} \mathbf{k}, \quad (\text{A.47})$$

$$\dot{\mathbf{k}} = - \frac{e^2}{2\varepsilon_0 m_e \omega} \frac{\partial N_{e0}}{\partial \mathbf{r}}. \quad (\text{A.48})$$

A.2 Booker quartic

The Booker quartic Hamiltonian is defined as [26]

$$\begin{aligned} H = \text{Re} \left[& [U^2(U - X) - UY^2] c^4 k^4 + X(\mathbf{k} \cdot \mathbf{Y})^2 c^4 k^2 \right. \\ & + [Y^2(2U - X) - 2U(U - X)^2] c^2 k^2 \omega^2 - X(\mathbf{k} \cdot \mathbf{Y})^2 c^2 \omega^2 \\ & \left. + (U - X) [(U - X)^2 - Y^2] \omega^4 \right]. \end{aligned} \quad (\text{A.49})$$

This Hamiltonian works only in anisotropic plasmas, and hence, only two versions of H are provided: BQWFWC and BQWFNC. The more general of them, BQWFWC, is then simplified to BQWFNC by dropping the appropriate terms when collisions are not included in the model.

A.2.1 With field and with collisions (BQWFWC)

The most general case of the Booker quartic Hamiltonian is obtained when collisions are included. In this form, the equations are given as below [26]:

$$\begin{aligned} \frac{\partial H}{\partial X} = & -U^2 c^4 k^4 + (\mathbf{k} \cdot \mathbf{Y})^2 c^4 k^2 \\ & + [-Y^2 + 4U(U - X)] c^2 k^2 \omega^2 - (\mathbf{k} \cdot \mathbf{Y})^2 c^2 \omega^2 \\ & + [Y^2 - 3(U - X)^2] \omega^4 \end{aligned} \quad (\text{A.50})$$

$$\begin{aligned} \frac{\partial H}{\partial Y_i} = & 2[-U c^4 k^4 + (2U - X) c^2 k^2 \omega^2 - (U - X) \omega^4] Y_i \\ & + 2X(\mathbf{k} \cdot \mathbf{Y}) c^2 [c^2 k^2 - \omega^2] k_i \end{aligned} \quad (\text{A.51})$$

$$\begin{aligned} \frac{\partial H}{\partial U} &= [2U(U - X) + U^2 - Y^2] c^4 k^4 \\ &+ 2 [Y^2 - (U - X)^2 - 2U(U - X)] c^2 k^2 \omega^2 \\ &+ [3(U - X)^2 - Y^2] \omega^4 \end{aligned} \quad (\text{A.52})$$

$$\frac{\partial H}{\partial x_i} = \text{Re} \left[\frac{\partial H}{\partial X} \frac{\partial X}{\partial x_i} + i \frac{\partial H}{\partial U} \frac{\partial Z}{\partial x_i} + \sum_{j=1}^3 \frac{\partial H}{\partial Y_j} \frac{\partial Y_j}{\partial x_i} \right] \quad (\text{A.53})$$

$$\begin{aligned} \frac{\partial H}{\partial k_i} &= \text{Re} \left[\{4 [U^2(U - X) - UY^2] c^4 k^2 + 2X(\mathbf{k} \cdot \mathbf{Y})^2 c^4 \right. \\ &+ 2 [Y^2(2U - X) - 2U(U - X)^2] c^2 \omega^2 \} k_i \\ &+ 2X(\mathbf{k} \cdot \mathbf{Y}) c^2 [c^2 k^2 - \omega^2] Y_i \left. \right] \end{aligned} \quad (\text{A.54})$$

$$\begin{aligned} \frac{\partial H}{\partial \omega} &= \text{Re} \left[\frac{\partial H}{\partial X} \frac{\partial X}{\partial \omega} + \frac{\partial H}{\partial U} \frac{\partial U}{\partial \omega} + \sum_{j=1}^3 \frac{\partial H}{\partial Y_j} \frac{\partial Y_j}{\partial \omega} \right. \\ &+ 2 [Y^2(2U - X) - 2U(U - X)^2] c^2 k^2 \omega - 2X(\mathbf{k} \cdot \mathbf{Y})^2 c^2 \omega \\ &+ 4(U - X) [(U - X)^2 - Y^2] \omega^3 \left. \right] \\ &= \text{Re} \left[-2 \frac{X}{\omega} \frac{\partial H}{\partial X} - i \frac{Z}{\omega} \frac{\partial H}{\partial U} - \frac{1}{\omega} \sum_{j=1}^3 Y_j \frac{\partial H}{\partial Y_j} \right. \\ &+ 2 [Y^2(2U - X) - 2U(U - X)^2] c^2 k^2 \omega - 2X(\mathbf{k} \cdot \mathbf{Y})^2 c^2 \omega \\ &+ 4(U - X) [(U - X)^2 - Y^2] \omega^3 \left. \right] \end{aligned} \quad (\text{A.55})$$

A.2.2 With field and no collisions (BQWFNC)

When collisions are not included, the equations are the same as for BQWFNC except that

$$Z = 0, \quad U = 1, \quad \frac{\partial Z}{\partial x_i} = 0. \quad (\text{A.56})$$

B Derivatives of the geomagnetic dipole field

In appendix A, derivatives of the Hamiltonian were presented. When anisotropy is introduced in the form of an included geomagnetic field, the formal ray equations require the evaluation of the spatial derivatives of the geomagnetic field. The developed ray tracer allows the option between one of two magnetic field types: Constant, or dipolar. If the field is set as constant, all the derivatives of \mathbf{Y} are trivially zero. However, the dipolar field derivatives are generally nonzero, and are presented in this appendix. To reiterate, the dipolar field used by the ray tracer is of the form

$$\mathbf{B}_0 = \frac{B_m R_E^3}{r^5} (3(\mathbf{m} \cdot \mathbf{r})\mathbf{r} - r^2\mathbf{m}), \quad (\text{B.1})$$

where the dipole moment unit vector \mathbf{m} is placed at the center of the Earth ($r = 0$), and the position \mathbf{r} is measured from the origin at the center of the Earth. We are ultimately interested in the the component j of the field \mathbf{B}_0 , denoted as $(\mathbf{B}_0)_j$, and its spatial derivatives. The dipole field is, for clarity, decomposed into its Cartesian components:

$$\mathbf{B}_0 = B_1 \hat{\mathbf{e}}_1 + B_2 \hat{\mathbf{e}}_2 + B_3 \hat{\mathbf{e}}_3 \quad (\text{B.2})$$

$$\mathbf{r} = x_1 \hat{\mathbf{e}}_1 + x_2 \hat{\mathbf{e}}_2 + x_3 \hat{\mathbf{e}}_3, \quad |\mathbf{r}| = r \quad (\text{B.3})$$

$$\mathbf{m} = m_1 \hat{\mathbf{e}}_1 + m_2 \hat{\mathbf{e}}_2 + m_3 \hat{\mathbf{e}}_3, \quad |\mathbf{m}| = 1 \quad (\text{B.4})$$

The spatial derivatives of the field are then derived to be

$$\frac{\partial \mathbf{B}_0}{\partial x_i} = B_m R_E^3 \left[-\frac{5x_i}{r^7} (3(\mathbf{m} \cdot \mathbf{r})\mathbf{r} - r^2\mathbf{m}) + \frac{3(m_i \mathbf{r} + (\mathbf{m} \cdot \mathbf{r})\hat{\mathbf{e}}_i) - 2\mathbf{m}x_i}{r^5} \right], \quad (\text{B.5})$$

which can be easily recast into its Cartesian components, since $\frac{\partial \hat{\mathbf{e}}_j}{\partial x_i} = \mathbf{0}$ when using Cartesian basis vectors (such as the ECEF system presented in section 2). Consequently, the sought derivatives are simply

$$\frac{\partial (\mathbf{B}_0)_j}{\partial x_i} = \left(\frac{\partial \mathbf{B}_0}{\partial x_i} \right)_j. \quad (\text{B.6})$$

C Ray tracer configuration file options

This appendix presents a template configuration file below, containing all the necessary options for a sample ray tracing session. Moreover, this appendix tables all of the configuration file options supported by the current version of the ray tracing program. If the unit column has been left blank, it is understood that the option is either a dimensionless quantity (e.g. the relative permittivity ϵ_r) or to be selected from a list. The selection column then lists the inputs supported by this particular option.

DEFINE TRANSMITTER:

```
latitude=60N
longitude=25E
altitude=100
antenna_type=ideal_dipole
antenna_elevation=90
antenna_bearing=0
```

DEFINE RHOMBIC ANTENNA PARAMETERS (if antenna_type==rhombic):

```
rhombic_beta=6
rhombic_alpha=16
```

DEFINE GEOMAGNETIC SOUTH POLE:

```
pole_latitude=80.31S
pole_longitude=107.38E
B_EQ=30.4e-6
magfield=dipole
```

DEFINE EARTH SURFACE ELECTRICAL PROPERTIES:

```
ground_conductivity=5
ground_permittivity=70
```

DEFINE IRI/NRLMSISE-00 OPTIONS:

```
use_IRI=false
IRI_latitude=60.1N
IRI_longitude=24.8E
year=2016
month=6
day=15
time=UTC
hour=15
height_begin=65
height_end=550
height_step=1
```

```
DEFINE IONOSPHERE PROFILES (if use_IRI==false):  
density=dens_0615_2016_15UTC_otaniemi.txt  
collisions=colls_0615_2016_15UTC_otaniemi.txt
```

```
DEFINE INTEGRATOR:  
pos_abstol=1e-8  
pos_reltol=1e-8  
k_abstol=1e-8  
k_reltol=1e-8  
initial_step=1e-8  
max_step=1e-6  
min_height=60e3  
max_height=500e3  
max_hops=1  
max_number_steps=1e6  
include_collisions=true  
include_magfield=true  
pseudoreal_tolerance=0.1  
resonance_tolerance=1e4
```

```
DEFINE PROGRAM (raytrace/photonmap):  
program=raytrace  
frequency=5e6  
num_threads=16
```

```
DEFINE RAY TRACING OPTIONS (if program==raytrace):  
mode=0  
elevation=30:5:90  
azimuth=0:10:360
```

```
DEFINE PHOTON MAPPING OPTIONS (if program==photonmap):  
photons=100000  
elevation_range=-90:90
```

```
DEFINE OUTPUT:  
path_position=ON  
path_k=OFF  
path_time=ON  
path_attenuation=ON  
path_refractive_index=OFF  
path_groupvelocity=OFF  
strike_ground=ON  
strike_maxH=ON
```

option	explanation	unit	selection	example
latitude	transmitter latitude	degrees	N,S	60.1N
longitude	transmitter longitude	degrees	E,W	24.8E
altitude	altitude of transmitter above ground	meters		100
antenna_type	type of transmitting antenna		isotropic, ideal_dipole, halfwave_dipole, rhombic	ideal_dipole
antenna_elevation	elevation α of antenna, see figure 24	degrees		90
antenna_bearing	bearing β of antenna, see figure 24	degrees		0
rhombic_beta	L/λ of a rhombic antenna, see figure 15 and eqns. (4.51)-(4.55). Hence, $\gamma = \pi L/\lambda$.			6
rhombic_alpha	Rhombus angle of the rhombic antenna, see figure 15.	degrees		16
pole_latitude	latitude of geomagnetic south pole	degrees	N,S	80.31S
pole_longitude	longitude of geomagnetic south pole	degrees	E,W	107.38E
B_EQ	The value B_E in eqn. (2.12), defined as the field strength on the magnetic equator. If the field is set to be constant, $ \mathbf{B}_0 = B_EQ$	tesla		30.4e-6
magfield	type of geomagnetic field model		dipole, constant	dipole
ground_conductivity	electrical conductivity σ of the Earth's surface	siemens/m		5
ground_permittivity	electrical relative permittivity ϵ_r of the Earth's surface			70
use_IRI	if 'true', IRI-2016 and NRLMSISE-00 are used to generate the ionosphere. If 'false', the user provides the profiles		true, false	true
IRI_latitude	latitude where the altitude profile is generated		N,S	60.1N
IRI_longitude	longitude where the altitude profile is generated		E,W	24.8E
year	year for profile generation			2018
month	month for profile generation		1-12	6
day	day of month for profile generation		1-31	15
time	time standard for profile generation		UTC, local	UTC
hour	hour of day for profile generation	hours	0-24	12
height_begin	starting height for profile generation	km	60-1500	60
height_end	final height for profile generation	km	60-1500	600
height_step	height distance between two consecutive generated values in the profile	km		1
density	path to user-defined electron density altitude profile, relative to input/userdefined			dens.txt
collisions	path to user-defined collision frequency altitude profile, relative to input/userdefined			colls.txt

pos_abstol	absolute tolerance ϵ_{atol}^r for integrating position. See eqn. (4.33)	m		1e-8
pos_reltol	relative tolerance ϵ_{rtol}^r for integrating position. See eqn. (4.33)			1e-8
k_abstol	absolute tolerance ϵ_{atol}^k for integrating wave vector. See eqn. (4.33)	1/m		1e-8
k_reltol	relative tolerance ϵ_{rtol}^k for integrating wave vector. See eqn. (4.33)			1e-8
initial_step	stepsize Δt of first step	s		1e-7
max_step	maximum stepsize	s		1e-6
min_height	altitude of limiting region h_{min}	m		60e3
max_height	maximum altitude of integration h_{max}	m		600e3
max_hops	maximum number of ground reflections			1
max_number_steps	maximum number of time steps			1e6
include_collisions	if 'true', collisions are included in ray equations. If 'false', $Z = 0$		true, false	true
include_magfield	if 'true', anisotropy included in ray equations. If 'false', $Y = 0$		true, false	true
pseudoreal_tolerance	maximum value ϵ_{pr} of $ \text{Im}[n]/\text{Re}[n] $. See section 5.4			0.2
resonance_tolerance	maximum value ϵ_{res} of $\text{Re}[n]$. See section 5.4			1e4
program	program mode		raytrace, photonmap	raytrace
frequency	frequency $f = \omega/2\pi$ of traced rays	Hz		5e6
num_threads	maximum number of threads used by OpenMP			16
mode	mode of ray after first magneto-ionic splitting event, if in ray tracing mode. Has no effect in photon mapping mode		O, X	O
elevation	vector of launch elevation angles to be swept in ray tracing mode. Format is start:step:end. See figure 24	degrees		40:1:90
azimuth	vector of launch azimuth angles to be swept in ray tracing mode. Format is start:step:end. See figure 24	degrees		0:10:360
photons	Number of randomly launched photons N in photon mapping mode. The initial powers U_k are normalized using eqn. (4.77) for these N photons			1e5
elevation_range	Elevation limits of launched photons. Format is minimum:maximum. Note, that the power is normalized using photons launched in all directions irrespective of this option	degrees		-90:90

path_position	if 'ON', value of $\mathbf{r}(t)$ is written to output for every time step	ON, OFF	ON
path_k	if 'ON', value of $\mathbf{k}(t)$ is written to output for every time step	ON, OFF	ON
path_time	if 'ON', the value of t is written to output for every time step	ON, OFF	ON
path_attenuation	if 'ON', the value of $A_{dB}(t)$ is written to output for every time step	ON, OFF	ON
path_refractive_index	if 'ON', the value of $n(t) = \mu + i\chi$ is written to output for every time step	ON, OFF	ON
path_groupvelocity	if 'ON', the value of $\dot{\mathbf{r}}(t)$ is written to output for every time step	ON, OFF	ON
strike_ground	if 'ON', the values of \mathbf{r} , $\dot{\mathbf{r}}$ and A_{dB} immediately before ground intersection are written to output	ON, OFF	ON
strike_maxH	if 'ON', the values of \mathbf{r} , $\dot{\mathbf{r}}$ and A_{dB} immediately before intersection with h_{max} (defined by max_height) are written to output	ON, OFF	ON