SANDIA REPORT

SAND2016-9089
Unlimited Release
Printed September 2016

Simulation of Optical Phenomena in the Upper Atmosphere

Mark C. Woods, William C. Sailor

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2016-9089
Unlimited Release
Printed September 2016

Simulation of Optical Phenomena in the Upper Atmosphere

Mark C. Woods
Phenomenology & Sensor Science
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0406
mcwood@sandia.gov

William C. Sailor
Phenomenology & Sensor Science
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0406
wcsailo@sandia.gov

Abstract

This SAND report investigates the electron transport equation in the upper atmosphere and how it relates to auroral light emissions. The electron transport problem is a very stiff boundary value problem, so standard numerical methods such as symmetric collocation and shooting methods will not succeed unless if the boundary conditions are altered with unrealistic assumptions. We show this to be unnecessary and demonstrate a method in which the fast and slow modes of the boundary value problem are essentially decoupled. This allows for an upwind finite difference method to be applied to each mode as is appropriate. This greatly reduces the number of points needed in the mesh, and we demonstrate how this eliminates the need to define new boundary conditions. This method can be verified by showing that under certain restrictive assumptions, the electron transport equation has an exact solution that can be written as an integral. The connection between electron transport and the aurora is made explicit and a kinetic model for calculating auroral light emissions is given.

Contents

Appendix

Figures

Tables

1 Introduction

The aurora is a phenomenon that has fascinated people fortunate enough to see one for millennia. The mostly green and sometimes red emissions that emanate from the sky are some of the most spectacular displays of the natural world. They are most commonly seen at high and low latitudes. At high latitudes it is called the aurora borealis and at low latitudes the aurora australis. Occasionally, aurorae can be seen in the middle latitudes. A famous example occurred in 37 AD when Emperor Tiberius mistakenly dispatched the Roman army to put out a fire that he thought was due to a battle in a northern city (see Bryant [14]). Evidently, he saw a red aurora on the horizon.

In more modern times, scientists have found that the aurora is caused by the interaction of charged particles from the sun (the solar wind) and the constituents of the earth's atmosphere. Due to an unexplained mechanism, the earth's magnetic field accelerates these charged particles and guides them to the poles of the planet. Once they are close enough to the earth, they scatter and impart energy to the atmospheric particles, which in turn release this energy in the form of light. This is called fluorescence. The scattering and energy deposition of the charged particles has been modeled by many researchers over the past fifty years.

The methods used to model this phenomena have largely focused on electron transport. Although protons contribute to the aurora, their contribution is not nearly as important. The models were generally split into methods valid for higher energy electrons ($\geqslant$ 1000 eV) and methods valid for lower energy electrons ($\leqslant$ 500 eV). Monte Carlo methods are the exception to this rule, but they are limited by computing power. In a series of papers, Cicerone et al. [19], Swartz [100], and Swartz and Stamnes [101] showed that Monte Carlo methods do not compare favorably to the deterministic methods. Nevertheless, they are still used. The most recent pure Monte Carlo study for auroral research is due to Solomon [84]. He later came up with a hybrid method consisting of his earlier Monte Carlo work and a deterministic method given by Banks et al. [8]. This hybrid method is described in Solomon [85].

Some of the first to consider a transport model valid for the high, low, and intermediate energy regimes were Banks and Nagy [7] and Nagy and Banks [73]. Their model is very similar to the one that will be presented in this thesis, but they assume that all scattering is isotropic. That is, they assume an electron will scatter without any preference to pitch angle (pitch angle will be defined in the next section). In actuality, electrons have a tendency for forward scattering (i.e. little or no change in pitch angle) with the tendency being very strong for high energy electrons. A slightly more sophisticated model was given in Mantas [58, 59], but the isotropic assumption was still used. Regardless, the model was applied to various problems in Mantas and Bowhill [61], Mantas and Walker [62], and Mantas et al. [63].

The assumption of isotropic scattering limited the applicability of the above approaches to lower energy electrons, but Strickland et al. [96] showed that relaxing this assumption yields

a model valid for all energy regimes. They demonstrated that anisotropic scattering could be accounted for as long as enough points were used in the pitch angle discretization. The main shortcoming was the numerical method used to solve the governing equation. They end up with a boundary value problem, for which they devised a modified form of shooting. They turned the problem into a coupled problem of upward and downward moving electrons. They solved the downward moving stream assuming no upward moving electrons. That solution was then used as an input to the upward moving stream. This process was repeated until the solution converged. Although they claim that this process always converges, it can be demonstrated that this is not true for general boundary value problems. Interestingly enough, this modified shooting was also applied by Link [52]. The difference was that the upward and downward streams were first added and subtracted to give a new set of coupled equations.

The most commonly used method for solving the electron transport problem had its origin with Stolarski [95], who showed that the problem was similar to the radiative transfer problem. At the time, there existed much research on radiative transfer, so the idea was to apply the methods of radiative transfer to electron transport. This idea was seized upon and expanded in Stamnes [86, 87, 88, 89]. However, the shortcoming is still in the numerical method. A change of variables is applied and the domain is split into subintervals. The boundary value problem coefficients are assumed to be constant within each subinterval, and a matrix exponential is calculated, giving the homogeneous solution. An exponential is then fit to the source term, and a particular solution is found for each subinterval. The subinterval solutions are pieced together using the boundary conditions and enforcing solution continuity.

In a series of papers (Stamnes and Swanson [93], Stamnes and Dale [91], Stamnes [90], and Stamnes et al. [94]), the matrix exponential method was written into a general purpose code named DISORT (DIScrete Ordinates Radiative Transfer). Although the numerical method did not change, the authors of DISORT meant it to be used for radiative transfer problems, not electron transport. Nevertheless, DISORT was slightly modified by Lummerzheim [54] and applied to auroral electron transport. It was also used in Lummerzheim et al. [56], Lummerzheim and Lilensten [55], Min et al. [69], and most recently in Lanchester and Gustavsson [51]. DISORT has become the standard in electron transport studies. Other than the occasional Monte Carlo study, the only exception is in Porter et al. [77] where the boundary value problem is turned into an integral equation and solved through Neumann iteration.

What has been missed in the various numerical methods (other than orthodoxy) is an accurate accounting of the lower atmosphere. To explain, at the upper boundary a downward (toward the earth) electron distribution is specified. Similarly, at some low altitude the upward electron distribution is set to zero. The top of the upper atmosphere is simply chosen to be an altitude where the density is relatively small and scattering effects are negligible. The bottom of the atmosphere is more troublesome. Theoretically, the ground (an altitude of zero) could be chosen because there are no free electrons at ground level. However, the electron transport equation is very stiff at low altitudes and becomes exponentially more stiff as the altitude decreases, so it is computationally more efficient to choose a higher

altitude. If this altitude is chosen too high, not only might this selection not conform to reality, but the numerical solution may be negative and oscillate in certain places, which is not physically meaningful. Further, if the boundary is chosen too low, standard numerical methods such as collocation at Gaussian or Lobatto points also may produce negative and oscillating solutions due to the stiffness.

For the most part, all of the above studies of the electron transport problem deemphasize the boundary conditions. Although they may be briefly mentioned, there is little discussion or explanation of the numerical complications they introduce. Some do not mention the boundary conditions at all (see Stamnes [88] for example) and some do not give enough detail to understand what exactly they used as a boundary condition (see Porter et al. [77]). In particular, almost all articles neglect to mention the consequences of how to handle the lower boundary. Only Mantas [59] discusses the difficulty of choosing the lower boundary and that negative intensities lead to instabilities and meaningless solutions. In a later article however, this same author uses an arbitrary reflecting lower boundary condition (see Mantas and Bowhill [61]). He forces 60% of all electrons reaching an altitude of 120 km back upward. The claim is made that although this is not realistic, the condition does not adversely affect the solution.

The approach developed here is based on the eigenvalue decomposition methods developed for stiff problems by Kreiss et al. [43] and Brown and Lorenz [13], which makes it possible to avoid any unrealistic assumptions at the lower boundary and solve the problem without using too fine of a mesh. In this SAND report, we discuss the physics of the electron transport problem in detail and derive the governing equation. We then show how the problem can be discretized and solved using the upwind method of Kreiss et al. [43]. This method is designed for very stiff boundary value problems, which is appropriate for the electron transport problem. We then discuss a simplified problem that can be solved exactly with the boundary element method. If the method of Kreiss et al. [43] agrees with this solution, then this helps verify that the upwind method is working properly. Finally, this report shows how the output of the electron transport equation can be used to compute auroral light emissions.

2 The Electron Transport Equation

2.1 Physics of the Aurora

For our purposes, the upper atmosphere will be defined as any altitude between 50 and 1000 km. This range includes the thermosphere and parts of the mesosphere and exosphere. The thermosphere is where the aurora occurs. The major neutral species or constituents are O, N_2, and O_2, and the minor species are H, He, N, Ar, and NO (see Rees [78]). O is the dominant species between about 200 and 600 km, whereas N_2 and O_2 dominate below 200 km. Above 600 km, the lighter atoms H and He are the most abundant, but the gas is not dense enough to include them as "major species". The number densities of the various species are shown in Figure 1. This data comes from the MSIS (Mass Spectrometer and Incoherent Scatter) atmospheric model of Hedin [30, 31] for all species except for NO, which come from Sharma et al. [82]. It should be noted that atmospheric density is difficult to model, and Figure 1 is nominal. Strickland et al. [97], Meier et al. [67], and Hecht et al. [29] have shown that model atmospheres can give densities that are up to a factor of two different from the actual densities. However, it will be used as our model neutral atmosphere.

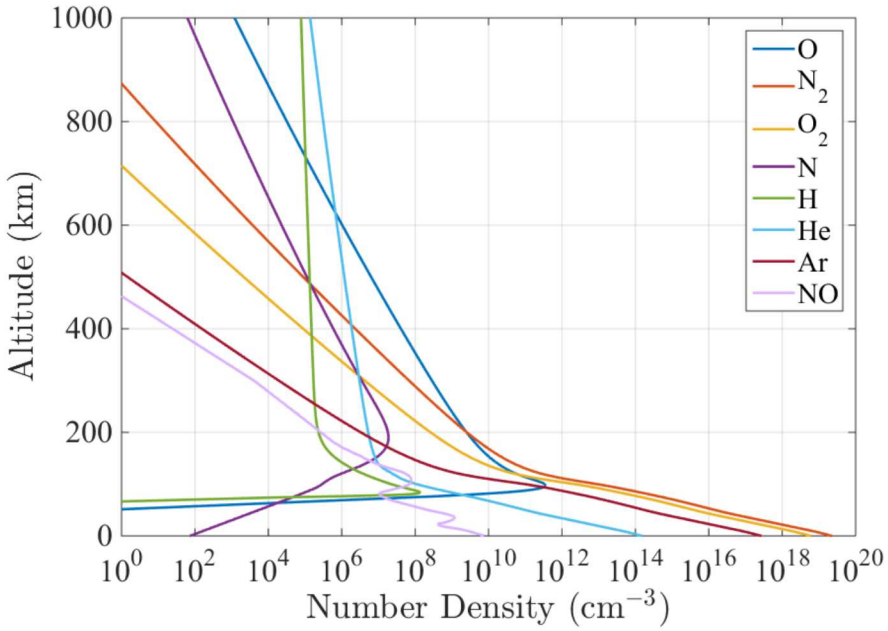

Figure 1. MSIS densities (Hedin [30, 31]) are from March 27, 1985, 1:30 UTC at 75° N, 90° W. NO densities are from Sharma et al. [82].

In addition to the neutral species, there are many different charged particles present in the upper atmosphere. The most abundant charged particle is the electron, and the most abundant ion is O^+. At the higher altitudes, H^+ (or protons p^+) are the most abundant,

and NO^+ is the most abundant at lower altitudes. The neutrals are far more abundant than the charged particles. The charged particle number densities are shown in Figure 2. This data comes from the IRI (International Reference Ionosphere) model of Bilitza and Reinisch [12]. Just as with the neutral atmospheric model, this data is nominal – there is a large variation in charged particle density depending on the local time and where the sun is in its 11 year solar cycle. Nevertheless, Figure 2 will be used as our model charged particle atmosphere. We also point out that this report is concerned with electron transport, but not the transport of the ambient electrons. From now on, any reference to electrons will mean the auroral electrons and not the ambient electrons. The word "ambient" will be used if ambient electrons are meant.

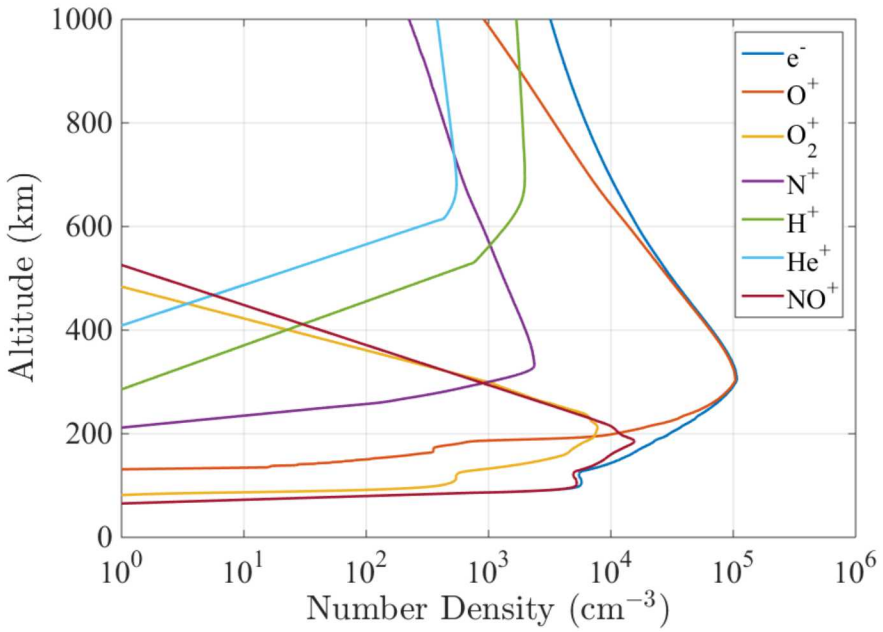

Figure 2. IRI densities (Bilitza and Reinisch [12]) are from March 27, 1985, 1:30 UTC at 75° N, 90° W.

The temperature profile of the upper atmosphere is highly variable. The thermosphere in particular is characterized by large increases in temperature as altitude increases. The neutrals, ambient electrons, and ions have separate profiles. Nominal profiles are shown in Figure 3. The temperature information comes from MSIS (Hedin [30, 31]) for the neutral temperatures and from IRI (Bilitza and Reinisch [12]) for the ambient ion and electron temperatures.

The aurora is caused by the interaction of the solar wind and the constituents of the upper atmosphere. The solar wind consists mostly of energetic electrons and protons. It also contains relatively small amounts of heavier ions such as He^{2+} and O^+, but their contributions to auroral light emissions is negligible (see Rees [78]). Most of the auroral light is a consequence of the streaming electrons, and for this reason we will be concerned with the

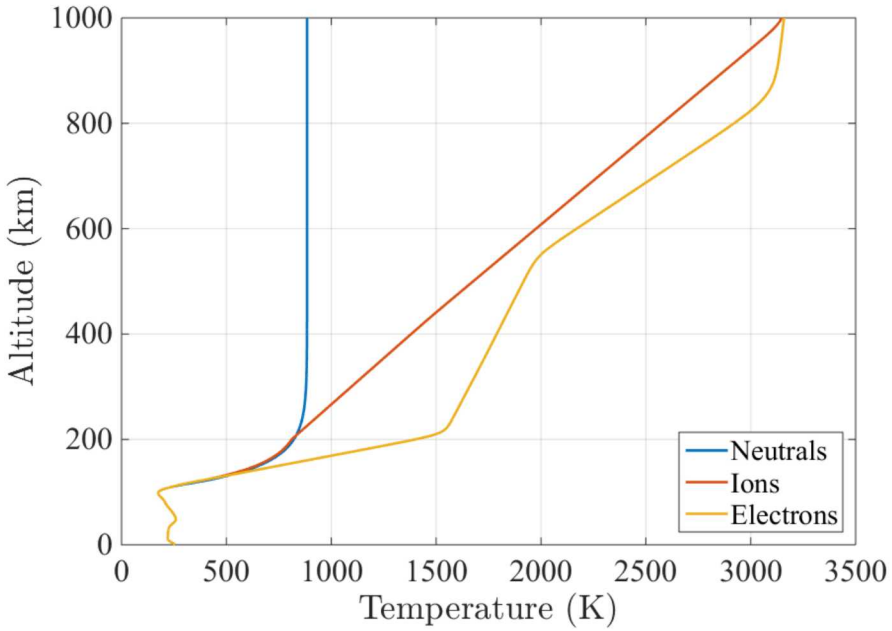

Figure 3. MSIS (Hedin [30, 31]) and IRI (Bilitza and Reinisch [12]) temperatures are from March 27, 1985, 1:30 UTC at 75° N, 90° W.

transport of electrons. Streaming protons contribute less directly to light emissions. Protons tend to steal electrons from the neutral gas, thus turning into hydrogen atoms. This process is called electron capture and is defined by the reaction

$$\mathrm{p}^+ + \mathrm{M} \rightarrow \mathrm{H} + \mathrm{M}^+ \tag{1}$$

where M is any neutral species. The streaming hydrogen atoms in turn tend to lose their electron in subsequent collisions. This process is called ionization stripping and is defined by the reaction

$$\mathrm{H} + \mathrm{M} \rightarrow \mathrm{p}^+ + \mathrm{e}^- + \mathrm{M}. \tag{2}$$

Thus, proton and hydrogen transport is a coupled process that is a source of streaming electrons. These electrons are much lower in abundance and energy than the solar wind electrons, so they do not contribute as much to light emissions. While proton and hydrogen transport could be incorporated into the electron transport equation, it is outside the scope of this report.

When the solar wind comes into the vicinity of the earth, the earth's magnetic field guides the charged particles to the poles of the planet. This is why aurorae are typically only seen at high and low latitudes (the aurora borealis and aurora australis, respectively). In addition, the magnetic field constrains the motion of the charged particles. In the altitude range of interest the magnetic field is almost uniform (see Chulliat et al. [18]), which causes the charged particles to move in helical paths. One such helical path for an electron is shown

in Figure 4. The kinetic energy E and pitch angle θ (the angle between the velocity vector and the magnetic field lines) of a charged particle is constant until it encounters an atom or molecule, where there is a probability of scattering. When scattering occurs, the particle's kinetic energy and/or pitch angle changes.

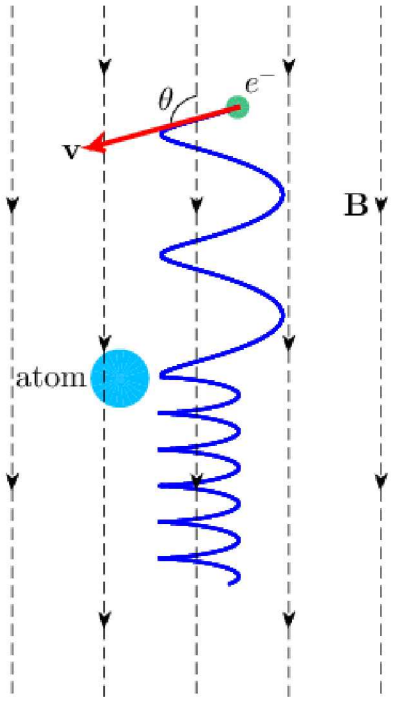

Figure 4. An example helical path along a uniform magnetic field **B**. The energy E and pitch angle θ change upon scattering.

Scattering is generally categorized as either elastic or inelastic. Elastic scattering is defined as a collision in which the kinetic energy of the streaming particle and atmospheric particle is conserved. For electrons, this can be modeled simply as a change in pitch angle. The energy exchange is very small due to the small mass of an electron. Conversely, inelastic scattering is defined as a collision in which kinetic energy is not conserved – part of it is converted to internal energy. Inelastic scattering can further be split into excitation, ionization, and dissociation. All of these processes require some energy threshold T in order to take place, and the streaming particle loses at least T in kinetic energy. This means that it must have kinetic energy greater than T to cause the inelastic reaction. An excitation is defined as an event where the atmospheric particle leaves its ground state and enters an excited state. Each excited state has its own energy threshold and probability of occurring. Ionization is defined as an event where the atmospheric particle loses at least one of its electrons. Just as there are many excited states for a neutral atom or molecule, there are many excited states for ions. A dissociation is defined as an event where a molecule breaks apart into two or more smaller molecules or atoms. It is also possible for a combination

of these to occur. For instance, an example of excitation, ionization, and dissociation in a single scattering event could be

$$N_2 + e^- \rightarrow N^+ + N^* + 2e^- \tag{3}$$

where N* is an excited state of nitrogen.

As these collisions occur, the atmospheric particles that are in excited states will decay back to their ground state. There are two ways that this takes place. The first is through a process called quenching. This is where one excited particle and another particle (not necessarily excited) collide, and the excited particle's internal energy is converted to kinetic energy. This is summarized as

$$M^* + X \rightarrow M + X \tag{4}$$

where M is any atmospheric particle, M* is the same particle in an excited state, and X is any ground state particle (M and X can also be ions). When quenching occurs, there is no light emission. The excess energy is converted back to kinetic energy.

The other way excited particles decay to their ground state is through fluorescence or spontaneous emission. This is when the excited particle randomly drops to a lower energy level without any external influence. This is summarized as

$$M^* \rightarrow M + h\nu \tag{5}$$

where $h\nu$ denotes light (h is Planck's constant and ν is the photon frequency – their product gives the photon energy). Each excited state has some lifetime associated with it. The lifetime of a state is defined as the amount of time it takes for the number of excited states to decay to e^{-1} of its original value. If a lifetime is short (on the order of nano and microseconds), then spontaneous emission dominates. For states with longer lifetimes, quenching dominates as long as the surrounding air is dense enough.

For instance, if the lifetime of some state is 2 minutes and if the surrounding air is dense enough, then it is likely that many collisions can take place before spontaneous emission occurs. Hence, quenching would be the most probable way for that atom or molecule to return to the ground state. If the surrounding air is less dense, then the probability of quenching decreases. For this reason, states with longer lifetimes tend to decay by spontaneous emission at higher altitudes where the air is less dense. If instead the lifetime was 10 nanoseconds, then it is unlikely any collisions will take place before spontaneous emission occurs, even at the most dense altitudes. Therefore, excited states with longer lifetimes have a variable branching between spontaneous emission and collisional quenching at a function of altitude, but excited states with short lifetimes do not.

The previous three equations are a bit misleading. They imply that an excited state always decays directly to the ground state. This is not the case. An excited particle may decay directly to the ground state, but it may instead decay to some lower excited state. In any case, the wavelength of the released photon in a vacuum is given by the Planck-Einstein relation

$$\lambda = \frac{hc}{\Delta E} \tag{6}$$

where c is the speed of light and ΔE is the change in energy between the two states. If there are n transitions with energy difference ΔE_i for the ith transition from the original excited state with threshold T back to the ground state, then conservation of energy gives $\sum_{i=1}^{n} \Delta E_i = T$.

A cross section is a theoretical area that quantifies the likelihood of some scattering event per unit electron flux per unit volume. It varies for each excited state and is a function of the incident electron energy. Cross sections can be determined theoretically through quantum mechanics or through experiment. In order to conduct any atmospheric study with scattering, a compilation of cross sections must be available. The cross sections used in this study are given in Appendix A.

2.2 Derivation of the Electron Transport Equation

Incomplete (and sometimes incorrect) derivations of the electron transport equation can be found in a number of sources. For instance, see Lummerzheim [54], Mantas [59, 60], Min et al. [69], and Stamnes and Rees [92]. The purpose of this section is to give a complete derivation of the electron transport equation, specifying the necessary assumptions as they are needed.

In order to quantify the transport of electrons through the upper atmosphere, we define the non-relativistic electron distribution function $f(\mathbf{r}, \mathbf{v}, t)$ to be the number of electrons per unit phase space (position space $\mathbf{r} = (x, y, z)$ and velocity space $\mathbf{v} = (v_x, v_y, v_z)$) at time t. The units of the distribution function are $\text{cm}^{-6}\ \text{s}^3$. Thus, the total number of electrons $N(t)$ at time t is some volume of phase space $\Omega_{\mathbf{r}} \times \Omega_{\mathbf{v}}$ is given by

$$N(t) = \iiint_{\Omega_{\mathbf{v}}} \iiint_{\Omega_{\mathbf{r}}} f(\mathbf{r}, \mathbf{v}, t)\, dx\, dy\, dz\, dv_x\, dv_y\, dv_z. \tag{7}$$

To find the electron distribution function, we use the continuity equation (also called the Boltzmann equation in this context) which describes the statistical behavior of the transporting electrons. It is given by

$$\frac{\partial f}{\partial t} + \nabla_{\mathbf{r}} \cdot (\mathbf{v} f) + \nabla_{\mathbf{v}} \cdot (\mathbf{a} f) = Q(\mathbf{r}, \mathbf{v}, t, f) \tag{8}$$

where $\mathbf{a} = (a_x, a_y, a_z)$ is acceleration and the source term $Q(\mathbf{r}, \mathbf{v}, t, f)$ is yet to be specified. Since $\mathbf{v}$ is an independent variable, we can write

$$\nabla_{\mathbf{r}} \cdot (\mathbf{v} f) = \mathbf{v} \cdot \nabla_{\mathbf{r}} f. \tag{9}$$

Further, in the absence of external forces, $\mathbf{a} = \mathbf{F}/m = \mathbf{0}$ where $\mathbf{F}$ is the sum of the external forces and m is the electron rest mass. This gives

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f = Q(\mathbf{r}, \mathbf{v}, t, f). \tag{10}$$

The electron distribution function is a common concept in physics and kinetic theory, but instruments on rockets for auroral studies do not normally measure it. Rather, they measure what is called electron intensity (also called differential number flux or just flux). The electron intensity is related to the distribution function by

$$I = \frac{v^2}{m} f \tag{11}$$

where $v = \|\mathbf{v}\|$ is speed. The units of intensity are $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}\ \mathrm{eV}^{-1}\ \mathrm{sr}^{-1}$. If we substitute (11) into (10) and multiply by m/v, we obtain

$$\frac{1}{v}\frac{\partial I}{\partial t} + \frac{\mathbf{v}}{v} \cdot \nabla_{\mathbf{r}} I = Q(\mathbf{r}, \mathbf{v}, t, I) \tag{12}$$

where the definition of the source term has changed.

It is convenient to represent $\mathbf{r}$ in the usual Cartesian coordinates where z is the local vertical direction and x and y are orthogonal local horizontal directions. Further, we represent $\mathbf{v}$ in spherical coordinates (v, θ, ϕ) where θ is the polar angle and ϕ is the azimuthal angle. We let θ be the same as the pitch angle so that it is the angle between the magnetic field lines and the electron velocity. This makes ϕ the angle of rotation about the magnetic field lines (see Figure 4). The z axis and magnetic field lines do not in general coincide. It is also convenient to let one of the independent variables be kinetic energy E instead of speed v. For non-relativistic energies, the relation between the two is $E = mv^2/2$.

At this point, some further simplifications can be made. First, the time it takes for even the slowest electrons to fully penetrate the atmosphere is much less than the time it takes for the atmosphere to respond to the streaming electrons or for the incident electron flux to vary. For this reason, steady state conditions will be assumed. That is,

$$\frac{\partial I}{\partial t} = 0. \tag{13}$$

Further, in the altitudes where the aurora is present the geomagnetic field can be accurately modeled as uniform. This implies that if the magnetic field is along some direction $\hat{\mathbf{e}}_s = \mathbf{B}/\|\mathbf{B}\|$, then the electron intensity only varies along $\hat{\mathbf{e}}_s$ due to how they are confined to helical paths about the field lines. That is, $\|\nabla_{\mathbf{r}} I\| = \partial I/\partial s$ and

$$\frac{\mathbf{v}}{v} \cdot \nabla_{\mathbf{r}} I = \cos\theta \|\nabla_{\mathbf{r}} I\| = \cos\theta \frac{\partial I}{\partial s}. \tag{14}$$

If we let $\mu = \cos\theta$ and substitute (13) and (14) into (12), we get

$$\mu \frac{\partial I}{\partial s} = Q(\mathbf{r}, E, \mu, \phi, I). \tag{15}$$

Atmospheric density (which will appear shortly in the source term) varies as a function of position $\mathbf{r}$. However, the variation in the xy plane is weak compared to the variation in the z direction. Since none of the other quantities (except electron intensity) depend on

position, we will only be concerned with the altitude direction $\hat{\mathbf{e}}_z$. Let α be the magnetic inclination or dip angle (the angle between $\hat{\mathbf{e}}_s$ and the xy plane). Then

$$\frac{\partial I}{\partial s} = \frac{dz}{ds}\frac{\partial I}{\partial z} = \sin\alpha \frac{\partial I}{\partial z}. \tag{16}$$

It can be shown that the gyroperiod of the electrons is much less than the average time between collisions (see Stamnes [87]). This implies that electron intensity is invariant to rotations about the magnetic field lines, and we can take our equation to be independent of ϕ. With these assumptions and (16), (15) becomes

$$\mu \sin\alpha \frac{\partial I(z, E, \mu)}{\partial z} = Q(z, E, \mu, I). \tag{17}$$

We now turn our attention to the source term $Q(z, E, \mu, I)$. Electron sources can come from a number of places. The most important for the aurora is the electron scattering source. All other sources will be specified later. We write this as

$$Q(z, E, \mu, I) = Q_{\text{sc}}(z, E, \mu, I) + Q_{\text{other}}(z, E, \mu). \tag{18}$$

In a scattering event, the electron changes its energy and/or pitch angle. That is, the phase space distribution changes. Mathematically, this can be modeled as the incident electron being destroyed and a new electron being created at the new energy and pitch angle (out scattering and in scattering, respectively). Physically they are the same electron, but this will appear as two different terms in $Q_{\text{sc}}(z, E, \mu, I)$. The two terms are given by

$$Q_{\text{sc}}(z, E, \mu, I) = Q_{\text{in}}(z, E, \mu, I) - Q_{\text{out}}(z, E, \mu, I). \tag{19}$$

We will consider elastic, excitation, dissociation, and ionization scattering. For every scattering source, it is assumed that all collisions are binary and the atmospheric particles are at rest. Compared to the speeds of the streaming electrons, this is a reasonable assumption. This also implies that streaming electrons cannot gain energy in a collision. Rather, the cascade of energy is from the streaming electrons to the atmosphere.

For elastic scattering, the rate of incoming electrons is given by

$$Q_{\text{in}}^{\text{el}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} n_\xi(z) \int_0^\infty \int_{-1}^1 \sigma_\xi^{\text{el}}(E') K_\xi^{\text{el}}(E, E', \mu, \mu') I(z, E', \mu')\, d\mu'\, dE' \tag{20}$$

where $n_\xi(z)$ is the number density of species ξ, $\sigma_\xi^{\text{el}}(E')$ is the elastic scattering cross section, and $K_\xi^{\text{el}}(E, E', \mu, \mu')$ is the elastic scattering kernel. Here, primed variables denote that quantity before scattering and unprimed variables denote that quantity after scattering. As discussed in the previous section, the elastic cross section $\sigma_\xi^{\text{el}}(E')$ governs the probability of an elastic collision between an electron with incident energy E' and species ξ. The elastic scattering kernel describes how electrons are redistributed in energy and pitch angle resulting from an elastic collision.

This can be simplified by assuming there is no exchange of kinetic energy from an elastic collision due to the extremely small mass of the electron. For example, mass ratio of a proton to an electron is $m_\mathrm{p}/m_\mathrm{e} \approx 1836.2$ (see Rees [78]). Hence, the elastic scattering kernel can be written as

$$K_\xi^{\mathrm{el}}(E, E', \mu, \mu') = \delta(E' - E) P_\xi(E, \mu, \mu') \tag{21}$$

where $\delta(\cdot)$ is the Dirac delta function and $P_\xi(E, \mu, \mu')$ is called the phase function. $P_\xi(E, \mu, \mu')$ is a probability density function describing how electrons at energy E are redistributed in pitch angle due to an elastic collision with species ξ. With this, (20) becomes

$$Q_{\mathrm{in}}^{\mathrm{el}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} n_\xi(z) \sigma_\xi^{\mathrm{el}}(E) \int_{-1}^{1} P_\xi(E', \mu, \mu') I(z, E, \mu')\, d\mu'. \tag{22}$$

The phase function can be determined from quantum mechanics. Using the Born approximation and the screened Coulomb or Yukawa potential (see Griffiths [26], Rees [78], and Sakurai and Napolitano [80]), we find that

$$P_\xi(E, \Theta) = \frac{\varepsilon_\xi(1 + \varepsilon_\xi)}{\pi(1 + 2\varepsilon_\xi - \cos\Theta)^2} \tag{23}$$

where $\cos\Theta = \mu\mu' + \sqrt{(1 - \mu^2)(1 - \mu'^2)}\cos(\phi - \phi')$ and ε_ξ is an energy dependent parameter which will be discussed further in the next section. Since our problem is azimuth independent, we can integrate out ϕ' to find

$$\begin{aligned} P_\xi(E, \mu, \mu') &= \frac{\varepsilon_\xi(1 + \varepsilon_\xi)}{\pi} \int_0^{2\pi} \frac{1}{\left[1 + 2\varepsilon_\xi - \mu\mu' - \sqrt{(1 - \mu^2)(1 - \mu'^2)}\cos(\phi - \phi')\right]^2}\, d\phi' \\ &= \frac{2\varepsilon_\xi(1 + \varepsilon_\xi)(1 + 2\varepsilon_\xi - \mu\mu')}{\left[(1 + 2\varepsilon_\xi - \mu\mu')^2 - (1 - \mu^2)(1 - \mu'^2)\right]^{3/2}}. \end{aligned} \tag{24}$$

The integration can be carried out by converting the cosine term to exponentials and making the substitution $\zeta = e^{-i(\phi - \phi')}$. It is a simple (but somewhat tedious) matter to carry out the integration using residues (see Ablowitz and Fokas [1]). The rate of outgoing electrons is a simpler expression than that of incoming electrons. For elastic scattering, it is given by

$$Q_{\mathrm{out}}^{\mathrm{el}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} n_\xi(z) \sigma_\xi^{\mathrm{el}}(E) I(z, E, \mu). \tag{25}$$

Using (22) and (25), the elastic scattering rate is

$$Q_{\mathrm{sc}}^{\mathrm{el}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} n_\xi(z) \sigma_\xi^{\mathrm{el}}(E) \left(\int_{-1}^{1} P_\xi(E', \mu, \mu') I(z, E, \mu')\, d\mu' - I(z, E, \mu) \right). \tag{26}$$

For excitation scattering, the rate of incoming electrons is given by

$$Q_{\mathrm{in}}^{\mathrm{ex}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} n_\xi(z) \sum_{\substack{\text{channels} \\ \eta}} \int_0^\infty \int_{-1}^{1} \sigma_\xi^\eta(E') K_\xi^\eta(E, E', \mu, \mu') I(z, E', \mu')\, d\mu'\, dE' \tag{27}$$

where channels denotes the various excitation states. In other words, each channel excites the atom or molecule to a different state. Again, the excitation kernel can be simplified. We will make the assumption that electrons do not change pitch angle in an excitation collision. This assumption is validated from experiments that show only a very small change in pitch angle (see Lummerzheim and Lilensten [55]). Further, Strickland et al. [96] tested a variety of excitation phase functions, but the effect on the solution was negligible. Hence, only the elastic scattering phase function has an appreciable affect on the solution of the electron transport problem. As discussed in the previous section, there is some energy threshold T_ξ^η that is required for excitation, i.e. the electron loses T_ξ^η in kinetic energy. These assumptions and simplifications give

$$K_\xi^\eta(E, E', \mu, \mu') = \delta(E' - (E + T_\xi^\eta))\delta(\mu' - \mu). \tag{28}$$

Substituting (28) into (27) gives

$$Q_{\text{in}}^{\text{ex}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} \sum_{\substack{\text{channels} \\ \eta}} n_\xi(z)\sigma_\xi^\eta(E + T_\xi^\eta)I(z, E + T_\xi^\eta, \mu). \tag{29}$$

The rate of outgoing electrons for excitation scattering is given by

$$Q_{\text{out}}^{\text{ex}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} \sum_{\substack{\text{channels} \\ \eta}} n_\xi(z)\sigma_\xi^\eta(E)I(z, E, \mu). \tag{30}$$

From (29) and (30), the excitation rate is given by

$$Q_{\text{sc}}^{\text{ex}}(z, E, \mu, I) = \sum_{\substack{\text{species} \\ \xi}} \sum_{\substack{\text{channels} \\ \eta}} n_\xi(z)\big[\sigma_\xi^\eta(E + T_\xi^\eta)I(z, E + T_\xi^\eta, \mu) - \sigma_\xi^\eta(E)I(z, E, \mu)\big]. \tag{31}$$

For dissociation, we will model it in the same way as excitation. That is, the dissociation cross sections will be included in the summations in (31). Technically, this should also result in a decrease in the number density of the molecule that was dissociated and an increase in the number density of the resulting atoms (since we are only considering diatomic molecules). However, the number of dissociations is very small compared to the overall number of molecules. This is especially true at low altitudes (less than 300 km) where most of the dissociations take place.

Lastly, we consider ionization scattering. Before giving the scattering kernel, we note that there are many ionized states for each species just as there are many neutral states for each species. Further, ionization means that one or more electrons are dislodged from the atom or molecule. However, most ionizations are single ionizations to the ion ground state. This is the only ionization channel we will include in our model. Including other ionization states or double and triple ionization is a matter of obtaining accurate cross sections for those channels. Their inclusion in the model is straightforward. With this, the rate of incoming

electrons for ionization scattering is given by

$$Q_{\text{in}}^{\text{ion}}(z,E,\mu,I) = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z)\bigg(\int_0^\infty \int_{-1}^1 \sigma_\xi^{\text{ion}}(E')K_\xi^{\text{pri}}(E,E',\mu,\mu')I(z,E',\mu')\,d\mu'\,dE' + \int_0^\infty \int_{-1}^1 \sigma_\xi^{\text{ion}}(E')K_\xi^{\text{sec}}(E,E',\mu,\mu')I(z,E',\mu')\,d\mu'\,dE'\bigg) \tag{32}$$

where "pri" denotes the primary electron and "sec" denotes the secondary electron. In the context of ionization, the primary electron is the incident electron after scattering occurs, and the secondary electron is the electron dislodged from the atom or molecule.

For ionization scattering, there is an energy threshold T_ξ^{ion} required to ionize the atom or molecule just like with excitation scattering. Unlike excitation scattering though, the loss of kinetic energy for the incident electron is greater than T_ξ^{ion} due to the presence of the secondary electron. The secondary electron's kinetic energy is taken from the primary electron. By conservation of energy, we have

$$E' = E_{\text{pri}} + E_{\text{sec}} + T_\xi^{\text{ion}}. \tag{33}$$

The energy of the secondary electron is governed by an energy-dependent distribution. The probability density function for an incident electron with energy E' to become a primary electron with energy E is given by the empirical function (see Rees [78])

$$R_\xi^{\text{pri}}(E,E') = \frac{1}{A(E')}\frac{1}{E'-E}\exp\bigg[-\frac{E'-E}{31.5} - 339\exp\bigg(-\frac{E'-E}{2.49}\bigg)\bigg] \times \log\bigg(\frac{\sqrt{E'}+\sqrt{E}}{\sqrt{E'}-\sqrt{E}}\bigg) \tag{34}$$

where the normalization $A(E')$ is determined from

$$\int_0^{E'-T_\xi^{\text{ion}}} R_\xi^{\text{pri}}(E,E')\,dE = 1. \tag{35}$$

Similarly, the probability density function for an incident electron with energy E' to dislodge a secondary electron with energy E is given by

$$R_\xi^{\text{sec}}(E,E') = R_\xi^{\text{pri}}(E'-E-T_\xi^{\text{ion}},E'). \tag{36}$$

For the primary electron, it is assumed that there is no change in pitch angle just as it was assumed for excitations and dissociations. For the secondary electron, however, we assume that it is dislodged isotropically. These assumptions along with (34) and (36) give

$$K_\xi^{\text{pri}}(E,E',\mu,\mu') = R_\xi^{\text{pri}}(E,E')\delta(\mu'-\mu), \tag{37}$$

$$K_\xi^{\text{sec}}(E,E',\mu,\mu') = \frac{1}{2}R_\xi^{\text{sec}}(E,E'). \tag{38}$$

Substituting (37) and (38) into (32) yields

$$Q_{\text{in}}^{\text{ion}}(z,E,\mu,I) = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z)\bigg(\int_{E+T_\xi^{\text{ion}}}^{\infty} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{pri}}(E,E') I(z,E',\mu)\, dE' + \frac{1}{2}\int_{E+T_\xi^{\text{ion}}}^{\infty}\int_{-1}^{1} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{sec}}(E,E') I(z,E',\mu')\, d\mu'\, dE' \bigg) \tag{39}$$

where we have used the fact that the cross sections vanish for $E' < E+T_\xi^{\text{ion}}$ (this is true for the excitation cross sections also replacing T_ξ^{ion} with the T_ξ^η, but it didn't make any difference since the argument of the Dirac delta in (28) automatically satisfied this requirement). Similar to before, the ionization rate out is given by

$$Q_{\text{out}}^{\text{ion}}(z,E,\mu,I) = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z)\sigma_\xi^{\text{ion}}(E) I(z,E,\mu), \tag{40}$$

and the ionization rate is

$$Q_{\text{sc}}^{\text{ion}}(z,E,\mu,I) = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z)\bigg(\int_{E+T_\xi^{\text{ion}}}^{\infty} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{pri}}(E,E') I(z,E',\mu)\, dE' + \frac{1}{2}\int_{E+T_\xi^{\text{ion}}}^{\infty}\int_{-1}^{1} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{sec}}(E,E') I(z,E',\mu')\, d\mu'\, dE' - \sigma_\xi^{\text{ion}}(E) I(z,E,\mu)\bigg). \tag{41}$$

The total scattering source is given by the sum of (26), (31), and (41). If we define the total cross section for species ξ by

$$\sigma_\xi^{\text{tot}}(E) = \sigma_\xi^{\text{el}}(E) + \sum_{\substack{\text{channels}\\ \eta}} \sigma_\xi^\eta(E) + \sigma_\xi^{\text{ion}}, \tag{42}$$

then we can write the electron transport equation as

$$\begin{aligned}\mu\,\sin\alpha \frac{\partial I(z,E,\mu)}{\partial z} = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z)\bigg(&- \sigma_\xi^{\text{tot}}(E) I(z,E,\mu) \\ &+ \sigma_\xi^{\text{el}}(E)\int_{-1}^{1} P_\xi(E',\mu,\mu') I(z,E,\mu')\, d\mu' + \sum_{\substack{\text{channels}\\ \eta}} \sigma_\xi^\eta(E+T_\xi^\eta) I(z,E+T_\xi^\eta,\mu) \\ &+ \int_{E+T_\xi^{\text{ion}}}^{\infty} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{pri}}(E,E') I(z,E',\mu)\, dE' \\ &+ \frac{1}{2}\int_{E+T_\xi^{\text{ion}}}^{\infty}\int_{-1}^{1} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{sec}}(E,E') I(z,E',\mu')\, d\mu'\, dE' \bigg) + Q_{\text{other}}(z,E,\mu).\end{aligned} \tag{43}$$

Let us now turn our attention to the "other" source term $Q_{\text{other}}(z, E, \mu, I)$. In the previous section, it was already pointed out that ionization stripping (2) resulting from streaming hydrogen atoms is a source of electrons. This requires that we first solve the coupled transport problem of streaming protons and hydrogen atoms. The output of this problem could then be used to define the ionization stripping source for the electron transport problem. However, this report is concerned with the numerical solution of the electron transport equation, and an additional source term does not add any new substance to our study of the electron transport equation. Addition of this source term may be required for analysis of experimental data using the results of the current work, however. For a further understanding of proton and hydrogen atom transport, see Basu et al. [10] and Strickland et al. [98].

Another source term comes from the interaction of streaming electrons with ambient electrons. An assumption can be made that as electrons stream through the atmosphere, they are continuously slowed down by ambient electrons (see Lummerzheim [54] or Mantas [59] for instance), as if by a frictional force. This approximation is said to be fairly accurate for energies greater than about 500 eV. In reality, however, the streaming electrons elastically scatter off the ambient electrons. A frictional force does not exist. The energy loss to ambient electrons is a discrete process just as it is for all other species. For high energy electrons, viewing it as a continuous process is reasonable since the electron loses a very small fraction of its total energy. For electrons with energies less than about 100 eV, a continuous approximation no longer holds because the streaming electron may lose a significant fraction of its total energy in an elastic collision with an ambient electron. Further, the hypothetical frictional force given in the continuous slowing down approximation is very small at the energies where it is valid. In other words, the force is negligible at energies where it is valid and not valid at energies where it is non-negligible.

Another source of contention with the continuous slowing down approximation is in how it treats energy transfer in collisions. In an elastic scattering event with a neutral species, we have ignored any energy transfer due to the large mass ratio. For a collision with an ambient electron, this ratio is unity. Therefore, the transfer of kinetic energy can no longer be ignored. Further, it is possible for the streaming electron to gain energy in such a collision. Authors who use the continuous slowing down approximation always assume the streaming electron does not gain energy in a collision. It would be better if these collisions could be modeled in the same way as the electron-neutral collisions are in (43). However, there are plasma effects to consider, and the equations have not yet been derived. Such a derivation, followed by a complete solution, may be necessary for the correct analysis of certain types of experimental measurements. Here, however, streaming electron-ambient electron collisions will not be considered. To see how the continuous slowing down approximation could be incorporated into the electron transport equation, see Stamnes and Rees [92].

There are more sources of electrons that could be included, but these are negligible. One such source is photoionization. This occurs when solar photons ionize the various atmospheric species. In general, since we are concerned with the aurora and the visible light emissions, we shall only consider nighttime conditions and assume no photoionization. To see how photoionization could be included, see Link [52].

Another source is from electron-ion collisions. As shown in Figure 2, various ions are present in the atmosphere, but Figure 1 shows that they are much less abundant than the neutral species. The inclusion of electron-ion collisions would be handled the same way as what was done above for electron-neutral collisions. One last source that could be included is electron-ion recombination. This is defined by the reaction

$$\mathrm{M}^+ + \mathrm{e}^- \rightarrow \mathrm{M}^* \rightarrow \mathrm{M} + h\nu. \tag{44}$$

However, this is a slow reaction, occurs at lower altitudes, and involves mostly low energy electrons.

The amount of physics included is limited by the data that is available. The cross sections for the various above reactions all need to be known in order to include them. The calculation of a cross section in all but the most simple cases is so complicated that they are usually found through experiments. For this reason, scientists have found the most important cross sections (i.e. those that contribute the most to atmospheric studies). These are the terms that we have included in (43).

Finally, the boundary conditions are important and some discussion is necessary. Physically, the electrons from the sun travel toward the ground. Thus, one boundary condition is that at the "top of the atmosphere" some downward electron intensity is specified, but this is not enough. There is some altitude that the electrons are unable to penetrate because the atmosphere is too thick for them to do so. Then another boundary condition is that at the "bottom of the atmosphere" the upward electron intensity vanishes. These can be written as

$$I(z_{\mathrm{top}}, E, \mu) = I_{\mathrm{top}}(E, \mu), \quad \mu < 0, \tag{45}$$

$$I(z_{\mathrm{bot}}, E, \mu) = 0, \quad \mu > 0, \tag{46}$$

Although these look unassuming, they cause many difficulties. One has to define what is meant by the "top and bottom of the atmosphere". The "top of the atmosphere" is simple. We say that this is an altitude where scattering is negligible so that the electron intensity has not been appreciably attenuated. Scattering becomes negligible when the density is very low. An altitude of 1000 km meets the above requirement.

The "bottom of the atmosphere" is more difficult. This is the altitude where the electron intensity is zero. We know this altitude exists because there are no free electrons at the ground. After all, another name for an energetic electron is a beta particle. Since radiation poisoning is not an epidemic, the auroral electrons must stop at some altitude above the ground. The problem is that this altitude is not easy to define. One main reason is that it varies as a function of electron energy. Electrons with greater kinetic energy tend to penetrate further into the atmosphere.

One way to circumvent this problem is to simply let the ground ($z = 0$ km) be the "bottom of the atmosphere". However, as we will see later, this is not wise because (43) is very stiff at low altitudes. Hence, this choice will result in much more computation than is

necessary. Another consequence of the extreme stiffness is that regardless of what altitude we let the "bottom of the atmosphere" be, the numerical solver may yield negative intensities, which are physically meaningless. Worse still, if the electron intensity is negative in one part of the domain, it causes other parts of the numerical solution to blow up. In other words, a negative electron intensity creates instability in the numerical solution. These issues will be addressed shortly.

2.3 Reduction of the Phase Function for Elastic Scattering

In order to discretize (43), the integrals should be approximated by appropriate quadrature sums. Many quadrature techniques approximate the integrand by a low-order polynomial or a piecewise polynomial. The problem here is that the phase function for elastic scattering $P_\xi(E, \mu, \mu')$ is not well approximated by a low-order polynomial due to the fact that it contains sharp peaks at $\mu = \mu'$. The formula for the phase function is given by (24) and is repeated here

$$P_\xi(E, \mu, \mu') = \frac{2\varepsilon_\xi(1+\varepsilon_\xi)(1+2\varepsilon_\xi - \mu\mu')}{\left[(1+2\varepsilon_\xi - \mu\mu')^2 - (1-\mu^2)(1-\mu'^2)\right]^{3/2}} \qquad \text{(24 revisited)}$$

where ε_ξ is an energy-dependent parameter that approaches zero as E becomes large. The result is that for large energies, $P_\xi(E, \mu, \mu')$ is sharply peaked at $\mu = \mu'$ and an inordinate number of points is required to approximate it with polynomials. There is an idea called the δ-M method, originally formulated by Wiscombe [104] for radiative transfer, that overcomes this difficulty. The formulas used here are slightly different than the original formulation, but the concepts are the same.

We begin by expanding the phase function as

$$P_\xi(E, \mu, \mu') = \sum_{m=0}^{\infty} \frac{2m+1}{2} \chi_{\xi,m}(E) P_m(\mu) P_m(\mu') \tag{47}$$

where $P_m(\cdot)$ is the mth degree Legendre polynomial and the $\chi_{\xi,m}$ are the phase function moments. The moments can be obtained using the well-known orthogonality relation of the Legendre polynomials (see Koornwinder et al. [39]). Multiplying (47) by $P_n(\mu)$, integrating over μ from -1 to 1, and solving for $\chi_{\xi,m}(E)$ yields

$$\chi_{\xi,m}(E) = \frac{1}{P_m(\mu')} \int_{-1}^{1} P_m(\mu) P_\xi(E, \mu, \mu')\, d\mu. \tag{48}$$

Since the moments do not depend on μ', we simply set $\mu' = 1$ because $P_m(1) = 1$ for all m. With this, we find

$$\chi_{\xi,m}(E) = \int_{-1}^{1} P_m(\mu) P_\xi(E, \mu, 1)\, d\mu. \tag{49}$$

It is convenient calculate the phase function moments using recursion since the Legendre polynomials have a well-known recursion relation (see Koornwinder et al. [39]). Calculating

the first two moments directly, we find that

$$\chi_0 = 1, \tag{50}$$

$$\chi_1 = 1 + 2\varepsilon + 2\varepsilon(1+\varepsilon)\log\left(\frac{\varepsilon}{1+\varepsilon}\right) \tag{51}$$

where we have dropped the E and ξ dependency. Then from (49),

$$\begin{aligned}
\chi_m &= \int_{-1}^{1} P_m(\mu)P(E,\mu,1)\,d\mu \\
&= 2\varepsilon(1+\varepsilon)\int_{-1}^{1}\frac{P_m(\mu)}{(1+2\varepsilon-\mu)^2}\,d\mu \\
&= 2\varepsilon(1+\varepsilon)\int_{-1}^{1}\left(\frac{2m-1}{m}\mu P_{m-1}(\mu) - \frac{m-1}{m}P_{m-2}(\mu)\right)\frac{1}{(1+2\varepsilon-\mu)^2}\,d\mu \\
&= 2\varepsilon(1+\varepsilon)\frac{2m-1}{m}\int_{-1}^{1}\frac{\mu P_{m-1}(\mu)}{(1+2\varepsilon-\mu)^2}\,d\mu - \frac{m-1}{m}\chi_{m-2} \\
&= 2\varepsilon(1+\varepsilon)\frac{2m-1}{m}\int_{-1}^{1}\left(\frac{1+2\varepsilon}{(1+2\varepsilon-\mu)^2} - \frac{1}{1+2\varepsilon-\mu}\right)P_{m-1}(\mu)\,d\mu \\
&\quad - \frac{m-1}{m}\chi_{m-2} \\
&= \frac{(2m-1)(1+2\varepsilon)}{m}\chi_{m-1} - 2\varepsilon(1+\varepsilon)\frac{2m-1}{m}\int_{-1}^{1}\frac{P_{m-1}(\mu)}{1+2\varepsilon-\mu}\,d\mu \\
&\quad - \frac{m-1}{m}\chi_{m-2} \\
&= \frac{(2m-1)(1+2\varepsilon)}{m}\chi_{m-1} - 2\varepsilon(1+\varepsilon)\frac{2m-1}{m}\left(\frac{P_m(\mu)-P_{m-2}(\mu)}{(2m-1)(1+2\varepsilon-\mu)}\right|_{-1}^{1} \\
&\quad - \frac{1}{2m-1}\int_{-1}^{1}\frac{P_m(\mu)-P_{m-2}(\mu)}{(1+2\varepsilon-\mu)^2}\,d\mu\Bigg) - \frac{m-1}{m}\chi_{m-2} \\
&= \frac{(2m-1)(1+2\varepsilon)}{m}\chi_{m-1} - \frac{m-1}{m}\chi_{m-2} + \frac{\chi_m - \chi_{m-2}}{m}
\end{aligned} \tag{52}$$

where we have used the Legendre polynomial recursion formula in going from the second equality to the third and integration by parts with a lesser-known property of the Legendre polynomials going from the sixth equality to the seventh. This property is (see Koornwinder et al. [39])

$$(2m-1)P_{m-1}(\mu) = \frac{d}{d\mu}\big[P_m(\mu) - P_{m-2}(\mu)\big]. \tag{53}$$

Now the recursion formula for the phase function moments can be written as

$$\chi_m = \frac{1}{m-1}\big[(2m-1)(1+2\varepsilon)\chi_{m-1} - m\chi_{m-2}\big], \qquad m = 2,3,\ldots \tag{54}$$

Although this formula has been reported in the literature (see Link [52] and Lummerzheim [54]), it appears it has never been derived analytically. In fact, Lummerzheim [54] states it

has not been verified and only becomes convinced of its accuracy after numerically testing it up to $m = 16$. The above calculation serves to establish its validity.

As mentioned earlier, the phase function is not well approximated by a low-order polynomial due to the sharp peak at $\mu = \mu'$. The idea behind the δ-M method is to approximate $P_\xi(E, \mu, \mu')$ by a Dirac delta function and M Legendre polynomials. This way, the delta function can capture the sharp peak and the polynomials can capture the rest. Let our approximation be

$$P^*_{\xi,M}(E, \mu, \mu') = f_\xi(E)\delta(\mu - \mu') + \big[1 - f_\xi(E)\big] \sum_{m=0}^{M-1} \frac{2m+1}{2} \chi^*_{\xi,m}(E) P_m(\mu) P_m(\mu') \tag{55}$$

where $f_\xi(E)$ is the "fraction" that represents the sharp peak and $\chi^*_{\xi,m}(E)$ are the modified moments. Substitute (55) into (49) in place of $P_\xi(E, \mu, \mu')$ and use orthogonality to obtain

$$\begin{aligned} \chi_{\xi,m}(E) &= \int_{-1}^{1} P^*_{\xi,M}(E, \mu, 1) P_m(\mu)\, d\mu \\ &= f_\xi(E) + \begin{cases} \big[1 - f_\xi(E)\big]\chi^*_{\xi,m}(E), & m \leqslant M-1 \\ 0, & m \geqslant M \end{cases}. \end{aligned} \tag{56}$$

From this, we find

$$\chi^*_{\xi,m}(E) = \frac{\chi_{\xi,m}(E) - f_\xi(E)}{1 - f_\xi(E)}, \qquad m = 0, 1, \ldots, M-1, \tag{57}$$

$$f_\xi(E) = \chi_{\xi,M}(E). \tag{58}$$

Using the definitions above and yet another property of the Legendre polynomials

$$\delta(\mu - \mu') = \sum_{m=0}^{\infty} \frac{2m+1}{2} P_m(\mu) P_m(\mu'), \tag{59}$$

we find that

$$\begin{aligned} P_\xi(E, \mu, \mu') - P^*_{\xi,M}(E, \mu, \mu') &= \sum_{m=0}^{\infty} \frac{2m+1}{2} \chi_m P_m(\mu) P_m(\mu') - f\delta(\mu - \mu') \\ &\quad - (1 - f) \sum_{m=0}^{M-1} \frac{2m+1}{2} \chi^*_m(E) P_m(\mu) P_m(\mu') \\ &= \sum_{m=0}^{\infty} \frac{2m+1}{2} \chi_m P_m(\mu) P_m(\mu') - \sum_{m=0}^{\infty} \frac{2m+1}{2} \chi_M P_m(\mu) P_m(\mu') \\ &\quad - \sum_{m=0}^{M-1} \frac{2m+1}{2} (\chi_m - \chi_M) P_m(\mu) P_m(\mu') \\ &= \sum_{m=M+1}^{\infty} \frac{2m+1}{2} (\chi_m - \chi_M) P_m(\mu) P_m(\mu'). \end{aligned} \tag{60}$$

This shows that $P_\xi(E,\mu,\mu') - P^*_{\xi,M}(E,\mu,\mu') \to 0$ as $M \to \infty$. Hence, the error in using the δ-M method approaches zero as M is increased to infinity.

Putting everything together, the electron transport equation is given by

$$\begin{aligned}
\mu \sin\alpha \frac{\partial I(z,E,\mu)}{\partial z} = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z) \Bigg(& - \big[\sigma^{\text{tot}}_\xi(E) - f_\xi(E)\sigma^{\text{el}}_\xi(E)\big] I(z,E,\mu) \\
& + \big[1 - f_\xi(E)\big]\sigma^{\text{el}}_\xi(E) \sum_{m=0}^{M-1} \frac{2m+1}{2} \chi^*_{\xi,m}(E) P_m(\mu) \int_{-1}^{1} P_m(\mu') I(z,E,\mu')\, d\mu' \\
& + \sum_{\substack{\text{channels}\\ \eta}} \sigma^\eta_\xi(E + T^\eta_\xi) I(z, E + T^\eta_\xi, \mu) + \int_{E+T^{\text{ion}}_\xi}^{\infty} \sigma^{\text{ion}}_\xi(E') R^{\text{pri}}_\xi(E,E') I(z,E',\mu)\, dE' \\
& + \frac{1}{2} \int_{E+T^{\text{ion}}_\xi}^{\infty} \int_{-1}^{1} \sigma^{\text{ion}}_\xi(E') R^{\text{sec}}_\xi(E,E') I(z,E',\mu')\, d\mu'\, dE' \Bigg) + Q(z,E,\mu)
\end{aligned} \tag{61}$$

where the subscript "other" has been dropped from the final term. The power of the δ-M method is that M does not need to be large to accurately approximate the elastic scattering integral. This means that a small number of points can be used, which greatly reduces the amount of computation.

3 Numerical Solution of the Electron Transport Equation

3.1 Discretization

The first step in many numerical procedures is to define an appropriate mesh and discretization. From (61), we see that electron intensity depends on intensities at higher energies. For this reason, energy is discretized starting with the largest energies. That is,

$$E_0 > E_1 > E_2 > \cdots > E_{K-1} > E_K > 0. \tag{62}$$

We want to pick E_0 large enough so that electron intensities at larger energies do not contribute very much to the right-hand-side of (61). This way, terms like $I(z, E + T_\xi^\eta, \mu)$ can be estimated by an asymptotic expression derived from the boundary condition (45) for $E > E_0$. For example, we can set

$$I(z, E, \mu) = \left(\frac{E_0}{E}\right)^\beta I(z, E_0, \mu), \qquad E > E_0 \tag{63}$$

if the incident electron intensity $I_{\text{top}}(E, \mu) \sim E^{-\beta} I_{\text{top}}(\mu)$ as $E \to \infty$. For auroral studies, $E_0 = 10^5$ eV is usually sufficient. We point out that since we are only concerned with light emissions, E_K does not need to be 0. The human eye can only see wavelengths between about 3800 and 7500 Å, and most auroral experiments are conducted in the visible range. By the Planck-Einstein relation (6), this means that energies less that 1.65 eV need not be considered. Due to the large range of energies considered, it is advisable to choose a logarithmic spacing in energy.

Evaluating (61) at $E = E_i$, we get

$$\begin{aligned}
\mu \sin\alpha \frac{\partial I(z, E_i, \mu)}{\partial z} = \sum_{\substack{\text{species} \\ \xi}} n_\xi(z) \Bigg(& - \big[\sigma_\xi^{\text{tot}}(E_i) - f_\xi(E_i)\sigma_\xi^{\text{el}}(E_i)\big] I(z, E_i, \mu) \\
& + \big[1 - f_\xi(E_i)\big]\sigma_\xi^{\text{el}}(E_i) \sum_{m=0}^{M-1} \frac{2m+1}{2} \chi_{\xi,m}^*(E_i) P_m(\mu) \int_{-1}^{1} P_m(\mu') I(z, E_i, \mu')\, d\mu' \\
& + \sum_{\substack{\text{channels} \\ \eta}} \sigma_\xi^\eta(E_i + T_\xi^\eta) I(z, E_i + T_\xi^\eta, \mu) + \int_{E_i + T_\xi^{\text{ion}}}^{\infty} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{pri}}(E_i, E') I(z, E', \mu)\, dE' \\
& + \frac{1}{2} \int_{E_i + T_\xi^{\text{ion}}}^{\infty} \int_{-1}^{1} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{sec}}(E_i, E') I(z, E', \mu')\, d\mu'\, dE' \Bigg) + Q(z, E_i, \mu)
\end{aligned} \tag{64}$$

for $i = 0, 1, 2, \ldots, K$. Now, we can approximate $I(z, E_i + T_\xi^\eta, \mu)$ for $i \neq 0$ with a linear interpolation. That is, if j is the index such that $E_j < E < E_{j-1}$ for some E not on the energy mesh, then

$$I(z, E, \mu) \approx H(z, E, \mu) = \frac{E - E_j}{E_{j-1} - E_j} I(z, E_{j-1}, \mu) + \frac{E - E_{j-1}}{E_j - E_{j-1}} I(z, E_j, \mu). \tag{65}$$

The integrals over energy can thus be approximated with the trapezoidal rule. The spacing used for this approximation is different from the energy mesh. This is because the integrands can vary on scales much smaller than the energy mesh, especially at large energies. The integrands can be calculated provided that $I(z, E, \mu) \approx H(z, E, \mu)$ is used for points not on the energy mesh. For our purposes, we will write this approximation as

$$\begin{aligned}\int_{E_i+T_\xi^{\text{ion}}}^{\infty} \sigma_\xi^{\text{ion}}(E') R_\xi^{\text{pri}}(E_i, E') I(z, E', \mu)\, dE' \\ \approx \sum_k u_k \sigma_\xi^{\text{ion}}(E_k) R_\xi^{\text{pri}}(E_i, E_k) H(z, E_k, \mu)\end{aligned} \tag{66}$$

where u_k and E_k are the trapezoidal rule weights and nodes, respectively. An adaptive quadrature scheme could be used, but this level of sophistication is not required. We simply use about 100 points between $E_i + T_\xi^{\text{ion}}$ and $E_i + T_\xi^{\text{ion}} + 200$ with more points toward the beginning of the interval. A expression similar to (66) is used for the secondary ionization integral, but the weights u_k and nodes E_k may be different. We use about 130 points between $E_i + T_\xi^{\text{ion}}$ and $E_i + T_\xi^{\text{ion}} + 800$ with more points toward the beginning of the interval.

As demonstrated above, the electron intensities are decoupled from each other in energy. That is, if we are trying to solve for the intensity at energy E_i, then the third, fourth, and fifth terms on the right-hand-side of (64) can be approximated as outlined above. The intensities are not, however, decoupled in pitch angle. This is due to the second term on the right-hand-side of (64). The integral is over the entire range of μ and the electron intensity under the integrand is unknown. The number of points used in the pitch angle discretization will determine the size of the boundary value problem system, so it is best to choose a quadrature rule that uses the fewest points possible. Since the limits on the pitch angle integrals do not change, it is convenient to use Gaussian quadrature. Further, the number of points should be even because an odd Gaussian quadrature would give $\mu = 0$ as one of the nodes. It can be shown that this would result in a differential-algebraic system of equations, which are often more difficult to solve numerically than their differential counterparts (see Ascher and Petzold [5]).

Using Gaussian quadrature with $2L$ points gives

$$\int_{-1}^{1} P_m(\mu') I(z, E_i, \mu')\, d\mu' \approx \sum_{\substack{\ell=-L \\ \ell\neq 0}}^{L} w_\ell P_m(\mu_\ell) I(z, E_i, \mu_\ell) \tag{67}$$

where w_ℓ and μ_ℓ are the Gaussian quadrature weights and nodes, respectively. In electron transport studies, it is common to use a "double-Gauss" rule. In this rule, μ_ℓ and w_ℓ for $\ell = 1, 2, \ldots, L$ are the nodes and weights for a Gaussian rule on the interval $(0, 1)$. That is

$$\int_0^1 f(\mu)\, d\mu \approx \sum_{\ell=1}^{L} w_\ell f(\mu_\ell). \tag{68}$$

Then the nodes and weights for the interval $(-1, 0)$ are given by $\mu_{-\ell} = -\mu_\ell$ and $w_{-\ell} = w_\ell$ for $\ell = 1, 2, \ldots, L$. This rule is important for several reasons. First, the electron intensity

often is discontinuous across $\mu = 0$ due to boundary conditions (45) and (46). Second, many electron transport studies are concerned with calculating the upward and downward flux. These are given by

$$\Phi_+(z,E) = +2\pi \int_0^1 \mu I(z,E,\mu)\, d\mu \tag{69}$$

$$\Phi_-(z,E) = -2\pi \int_{-1}^0 \mu I(z,E,\mu)\, d\mu. \tag{70}$$

Hence, flux is a simple matter to calculate if the double-Gauss rule is used. Regardless of whether ordinary Gaussian quadrature or the double-Gauss rule is used, the approximation can be written in the form (67).

We are now ready to fully discretize (64). We find

$$\begin{aligned}
\mu_l \sin\alpha \frac{dJ_{i,l}(z)}{dz} = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z) \Bigg(& -\big[\sigma_\xi^{\text{tot}}(E_i) - f_\xi(E_i)\sigma_\xi^{\text{el}}(E_i)\big] J_{i,l}(z) \\
& + \big[1 - f_\xi(E_i)\big]\sigma_\xi^{\text{el}}(E_i) \sum_{m=0}^{M-1} \frac{2m+1}{2} \chi^*_{\xi,m}(E_i) P_m(\mu_l) \sum_{\substack{\ell=-L\\ \ell\neq 0}}^{L} w_\ell P_m(\mu_\ell) J_{i,\ell}(z) \\
& + \sum_{\substack{\text{channels}\\ \eta}} \sigma_\xi^\eta(E_i + T_\xi^\eta) H(z, E_i + T_\xi^\eta, \mu_l) + \sum_{k_1} u_{k_1} \sigma_\xi^{\text{ion}}(E_{k_1}) R_\xi^{\text{pri}}(E_i, E_{k_1}) H(z, E_{k_1}, \mu_l) \\
& + \frac{1}{2} \sum_{k_2} u_{k_2} \sigma_\xi^{\text{ion}}(E_{k_2}) R_\xi^{\text{sec}}(E_i, E_{k_2}) \sum_{\substack{\ell=-L\\ \ell\neq 0}}^{L} w_\ell H(z, E_{k_2}, \mu_\ell) \Bigg) + Q(z, E_i, \mu_l)
\end{aligned} \tag{71}$$

for $i = 0, 1, \ldots, K$ and $l = \pm 1, \pm 2, \ldots, \pm L$ where $J_{i,l}(z) \approx I(z, E_i, \mu_l)$. This equation can be written in matrix form. We can define

$$\begin{aligned}
\mathbf{J}_i(z) &= \begin{bmatrix} J_{i,-L}(z) & \cdots & J_{i,-1}(z) & J_{i,1}(z) & \cdots & J_{i,L}(z) \end{bmatrix}^T \\
\mathbf{H}(z,E) &= \begin{bmatrix} H(z,E,\mu_{-L}) & \cdots & H(z,E,\mu_{-1}) & H(z,E,\mu_1) & \cdots & H(z,E,\mu_L) \end{bmatrix}^T \\
\mathbf{P}_m &= \begin{bmatrix} P_m(\mu_{-L}) & \cdots & P_m(\mu_{-1}) & P_m(\mu_1) & \cdots & P_m(\mu_L) \end{bmatrix}^T \\
\mathbf{Q}(z,E) &= \begin{bmatrix} Q(z,E,\mu_{-L}) & \cdots & Q(z,E,\mu_{-1}) & Q(z,E,\mu_1) & \cdots & Q(z,E,\mu_L) \end{bmatrix}^T \\
\mathbf{P} &= \begin{bmatrix} | & | & | & & | \\ \mathbf{P}_0 & \mathbf{P}_1 & \mathbf{P}_2 & \cdots & \mathbf{P}_{M-1} \\ | & | & | & & | \end{bmatrix} \\
\mathbf{D}_\xi(E) &= \frac{1}{2} \operatorname{diag}\Big(\chi^*_{\xi,0}(E),\, 3\chi^*_{\xi,1}(E),\, 5\chi^*_{\xi,2}(E),\, \ldots,\, (2M-1)\chi^*_{\xi,M-1}(E)\Big) \\
\mathbf{W} &= \operatorname{diag}\big(w_{-L},\, \ldots,\, w_{-1},\, w_1,\, \ldots,\, w_L\big) \\
\boldsymbol{\mu} &= \operatorname{diag}\big(\mu_{-L},\, \ldots,\, \mu_{-1},\, \mu_1,\, \ldots,\, \mu_L\big) \sin\alpha.
\end{aligned}$$

With these definitions, our discretized equation becomes

$$\begin{aligned}\boldsymbol{\mu}\frac{d\mathbf{J}_i(z)}{dz} = \sum_{\substack{\text{species}\\ \xi}} n_\xi(z)\Bigg(&- \big[\sigma_\xi^{\text{tot}}(E_i) - f_\xi(E_i)\sigma_\xi^{\text{el}}(E_i)\big]\mathbf{J}_i(z) \\ &+ \big[1 - f_\xi(E_i)\big]\sigma_\xi^{\text{el}}(E_i)\mathbf{P}\mathbf{D}_\xi(E)\mathbf{P}^T\mathbf{W}\mathbf{J}_i(z) \\ &+ \sum_{\substack{\text{channels}\\ \eta}} \sigma_\xi^\eta(E_i + T_\xi^\eta)\mathbf{H}(z, E_i + T_\xi^\eta) \\ &+ \sum_{k_1} u_{k_1}\sigma_\xi^{\text{ion}}(E_{k_1})R_\xi^{\text{pri}}(E_i, E_{k_1})\mathbf{H}(z, E_{k_1}) \\ &+ \frac{1}{2}\mathbf{1}\mathbf{W}\sum_{k_2} u_{k_2}\sigma_\xi^{\text{ion}}(E_{k_2})R_\xi^{\text{sec}}(E_i, E_{k_2})\mathbf{H}(z, E_{k_2})\Bigg) + \mathbf{Q}(z, E_i)\end{aligned} \tag{72}$$

where $\mathbf{1} \in \mathbb{R}^{2L\times 2L}$ is a matrix of ones. It is now a simple matter to convert this to

$$\frac{d\mathbf{J}_i(z)}{dz} = \mathbf{A}(z, E_i)\mathbf{J}_i(z) + \mathbf{q}(z, E_i) \tag{73}$$

where $\mathbf{A}(z, E_i) \in \mathbb{R}^{2L\times 2L}$ and $\mathbf{q}(z, E_i) \in \mathbb{R}^{2L}$. The only parts that may not be straightforward are the terms involving $\mathbf{H}(z, E)$. From (65) and depending on the energy mesh, one of the interpolating points may be the current energy. For example, we might find $E_{i-1} < E_i + T_\xi^\eta < E_i$ for the third term on the right-hand-side of (72). In this case, one of the terms in (65) would contribute to $\mathbf{A}(z, E_i)$ and the other term would contribute to $\mathbf{q}(z, E_i)$. If instead $E_{j-1} < E_i + T_\xi^\eta < E_j$ where $j < i$, then both terms in (65) would contribute to $\mathbf{q}(z, E_i)$.

We can also rewrite the boundary conditions (45) and (46) in matrix form. This is simply

$$\mathbf{B}_{\text{top}}\mathbf{J}_i(z_{\text{top}}) + \mathbf{B}_{\text{bot}}\mathbf{J}_i(z_{\text{bot}}) = \mathbf{I}_{\text{BC}}(E_i) \tag{74}$$

where $\mathbf{B}_{\text{top}}, \mathbf{B}_{\text{bot}} \in \mathbb{R}^{2L\times 2L}$ and $\mathbf{I}_{\text{BC}}(E_i) \in \mathbb{R}^{2L}$. Here, $\mathbf{B}_{\text{top}}$ contains the $L \times L$ identity matrix in its upper-left corner and zeros elsewhere. Similarly, $\mathbf{B}_{\text{bot}}$ contains the identity matrix in its lower-right corner and zeros elsewhere. The vector $\mathbf{I}_{\text{BC}}(E_i)$ is constructed from the right-hand-sides of (45) and (46) in a straightforward manner.

3.2 Stiffness of the Electron Transport Equation

Equations (73) and (74) together form a linear boundary value problem (BVP). The difficulty with this particular problem is that it is very stiff. We can see from (72) that the eigenvalues of the system are proportional to the number densities $n_\xi(z)$. We also see from Figure 1 that at high altitudes the number densities are relatively small, whereas at low altitudes the number densities are relatively large. In fact, at low altitudes the largest number densities increase exponentially as altitude decreases. This means that the same trend holds true for the eigenvalues of the system.

The largest eigenvalue for a representative problem at two altitudes and energies are shown in Table 1. Using a low order BVP solver, it is not unreasonable to require that the product of the local step size and largest local eigenvalue be less than 2. It is well known that keeping this product small helps keep the error small (see Ascher and Petzold [5]). Table 1 shows that this is not a problem for the high altitudes. Step sizes on the order of kilometers are possible. However, at the low altitudes this requires step sizes on the order of centimeters at 100000 eV and microns at 10 eV. Clearly, if we wish to accurately solve the problem without using hundreds of thousands of points, something else must be done.

Table 1. The largest eigenvalue in cm^{-1} for $z = 50, 1000$ km and $E = 10, 100000$ eV.

Altitude \ Energy	10 eV	100000 eV
50 km	1.639×10^{3}	1.513×10^{-1}
1000 km	7.960×10^{-9}	1.194×10^{-13}

The problem under consideration qualifies as being a very stiff BVP. We make a distinction between a stiff and a very stiff BVP. For a stiff BVP, standard methods such as collocation at Gaussian and Lobatto points (to be discussed shortly) can still be used as long as enough points are placed in the stiff regions of the domain. For a very stiff BVP, this requires far too many points and is not computationally practical. Standard methods will be unacceptable in this case because they will be inaccurate due to a failure to damp out the rapidly increasing and/or decreasing modes. Even worse, for the electron transport problem they can yield negative intensities, which are meaningless. This is likely the reason why all previous numerical attempts at this problem have made spurious assumptions such as 60% of all electrons reaching 120 km are reflected back upward. Numerical methods for very stiff BVPs are not widely used or well-known. The purpose of this chapter is to describe a numerical method that avoids these spurious assumptions.

3.3 Numerical Methods for Stiff Boundary Value Problems

Two of the more elementary ways of solving BVPs are by the shooting method or by replacing the derivative with a divided difference and solving the resulting linear system. The flaws of the shooting method are numerous and outlined in Ascher et al. [6] and Ascher and Petzold [5]. For this reason, some researchers resort to the multiple shooting method, but this method still has its drawbacks. The divided difference method's main drawback is that it is difficult to extend it beyond second order accuracy. That is, if a higher order method is desired, then the resulting system will be underdetermined.

The most popular methods for solving BVPs are based on collocation at either Gauss or Lobatto points. In these methods, a set of collocation points $0 \leqslant \rho_1 < \rho_2 < \cdots < \rho_P \leqslant 1$ is chosen along with a set of basis functions (usually polynomials). A mesh is created, and for

every subinterval of the domain the numerical solution is given by the linear combination of the basis functions that satisfy the differential equation at the collocation points. For our problem, this would mean that a polynomial will satisfy (73) at $z_{n-1} + \rho_j \Delta z_n$ for every subinterval $[z_{n-1}, z_n]$ where $\Delta z_n = z_n - z_{n-1}$. The difference between different collocation methods is the choice of collocation points. Gauss and Lobatto points are popular because they are symmetric, give the highest order accuracy possible, and they are A-stable (see Ascher and Petzold [5]). Hence, they are an easy choice for non-stiff and moderately stiff problems, and software is freely available.

As the BVP becomes more stiff, however, symmetric methods begin to lose their appeal. It has been demonstrated by Ascher and Weiss [3, 4] that under certain assumptions, Gauss and Lobatto collocation can still be used as long as more mesh points are placed in the stiff regions of the domain. The assumptions are that the system matrix ($\mathbf{A}(z, E_i)$ in our case) can be immediately separated into a stiff part and a non-stiff part and no turning points are present. A turning point can mean different things in different contexts, but in this context it is a point where one of the system eigenvalues changes by an order of magnitude. These assumptions can be very restrictive depending on the problem (see Kreiss [42]).

For this reason, other methods were devised. One such method is given in Dieci et al. [21, 22]. A transformation is found such that the BVP is replaced by a nonlinear matrix initial value problem (IVP), a linear vector IVP, and a linear vector terminal value problem (TVP). The BVP solution can be constructed from the solutions to these three problems. The theoretical underpinnings and numerical difficulties are discussed at length. The biggest advantage of this method is that stiff IVP solvers are readily available. The main drawback is that the method is largely untested on large systems, so its robustness is unknown. For this reason, symmetric collocation methods still dominate (see Ascher et al. [6]).

All of the above methods have difficulty with turning points. For symmetric collocation, this is clearly demonstrated in Kreiss et al. [43]. They give an example where even when a large number of points are placed in the vicinity of the turning point, the numerical solution is unsatisfactory for both Gauss and Lobatto collocation. In fact, Kreiss [42] argues that the only way to know if turning points exist is to find the system eigenvalues throughout the domain. However, if the eigenvalues are found, then it is possible to decouple the system. That is, the system can be block diagonalized so that the eigenvalues are separated according to their magnitude and sign. Then an appropriate difference method can be applied to each block. Appropriate one-sided or upwind difference methods are applied to the blocks that contains large eigenvalues, and symmetric difference methods are applied to the block that contains the small eigenvalues. These ideas will be made more clear in this chapter.

The idea of upwinding also has its drawbacks. As we will see, it requires much more computation than symmetric collocation methods. This is not unexpected since the underlying BVP is much more difficult numerically. One of the difficulties is in how to select a mesh. Stiff problems have a tendency to have large gradients in the solution. These regions are called boundary and interior layers (see Holmes [32]). The difficulty is that in order for the mesh to accurately resolve the solution, the location of these layers is required because more points need to be placed in their vicinity. One way to handle this is to use adaptive mesh

refinement. This is used in conjunction with upwinding in Kreiss and Kreiss [41]. Another way to handle this is to use the BVP coefficients ($\mathbf{A}(z, E_i)$ and $\mathbf{q}(z, E_i)$ in our case) to determine where the layers will be and create an a priori mesh. The prescription given in Kreiss et al. [43] for a priori mesh construction guarantees that the solution will be resolved on that mesh. Further, it naturally leads to upwinding. This will be demonstrated later on in this chapter. The method of Kreiss et al. [43] is a low order method, but it has been shown by Brown and Lorenz [13] and Ringhofer [79] that the method can be extended to higher orders. The problem is reframed in a collocation setting, and various sets of collocation points are compared.

3.4 Overview of the Upwinding Method

To explain what is done in the upwind method, consider the problem

$$\frac{d\mathbf{y}(x)}{dx} = \mathbf{A}(x)\mathbf{y}(x) + \mathbf{f}(x), \qquad a \leqslant x \leqslant b \tag{75}$$

with boundary conditions

$$\mathbf{B}_a\mathbf{y}(a) + \mathbf{B}_b\mathbf{y}(b) = \mathbf{c} \tag{76}$$

where $\mathbf{y}(x), \mathbf{f}(x), \mathbf{c} \in \mathbb{R}^M$ and $\mathbf{A}(x), \mathbf{B}_a, \mathbf{B}_b \in \mathbb{R}^{M\times M}$. It turns out that the eigenvalues for the electron transport problem are real, so we will assume real eigenvalues throughout this section. The method we will use is derived from and introduced by Kreiss et al. [43] and Brown and Lorenz [13]. These authors were principally focused on proving error estimates and did not adequately explain its implementation. Consequently, we will be focused on implementation.

Throughout this section, we will assume some mesh $a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b$ with $\Delta x_n = x_n - x_{n-1}$ for $n = 1, 2, \ldots, N$ and

$$\Delta x = \max_{1\leqslant n\leqslant N} \Delta x_n. \tag{77}$$

In addition, all norms used denote the infinity norm. The notation

$$\|\mathbf{z}(x)\|_{(\alpha,\beta)} = \sup_{\alpha<x<\beta} \|\mathbf{z}(x)\|_\infty = \sup_{\alpha<x<\beta} \|\mathbf{z}(x)\| \tag{78}$$

will be used throughout for the norm of a vector function $\mathbf{z}(x)$ on an interval (α, β).

As stated in Kreiss et al. [43], a function $\mathbf{y}(x)$ is resolved on an interval (α, β) if

$$\left\| \frac{d^\nu \mathbf{y}(x)}{dx^\nu} \right\|_{(\alpha,\beta)} \leqslant K\big(\|\mathbf{y}(x)\|_{(\alpha,\beta)} + 1\big) \tag{79}$$

for $\nu = 0, 1, \ldots, p$ where $K\Delta x \ll 1$. The degree of smoothness p can vary with the BVP. The importance of using a mesh that resolves the solution follows from the error analysis for any finite difference method. For example, the local truncation error for the trapezoidal rule

is bounded by $\|\boldsymbol{\tau}_n\| \leqslant \frac{1}{12}\Delta x^3 \|\mathbf{y}'''(x)\|_{(x_{n-1},x_n)}$. For non-stiff BVPs, Δx can be made small enough so that this error is small. However, for a stiff BVP the derivatives of the solution can be very large so that the error is large unless if Δx is made prohibitively small. For this reason, a solution is only resolved if a number p of its derivatives are bounded by a constant K that is not too large (i.e. $K\Delta x \ll 1$). Further, if $\mathbf{y}(x)$ is resolved on (α, β) and (β, γ), then it is resolved on (α, γ). This means that we only need to worry about resolving $\mathbf{y}(x)$ in the neighborhood of every point $x \in (a, b)$.

The obvious problem with (79) is that in order to know if the mesh resolves the solution, it appears we need to already have the solution. It turns out that we can find a mesh that resolves the solution using only information about the BVP coefficients $\mathbf{A}(x)$ and $\mathbf{f}(x)$. This means that an adequate mesh can be found before obtaining the solution.

Definition 1. Suppose a matrix function $\mathbf{D}(x) \in \mathbb{R}^{M\times M}$ can be partitioned into the form

$$\mathbf{D}(x) = \begin{bmatrix} \mathbf{D}^{11}(x) & \mathbf{D}^{12}(x) & \mathbf{D}^{13}(x) \\ \mathbf{D}^{21}(x) & \mathbf{D}^{22}(x) & \mathbf{D}^{23}(x) \\ \mathbf{D}^{31}(x) & \mathbf{D}^{32}(x) & \mathbf{D}^{33}(x) \end{bmatrix} \tag{80}$$

where $\mathbf{D}^{ij}(x) \in \mathbb{R}^{m_i\times m_j}$ for $i, j = 1, 2, 3$ and $m_1+m_2+m_3 = M$. $\mathbf{D}(x)$ is *essentially diagonally dominant* on (α, β) if $\mathbf{D}^{11}(x)$ and $\mathbf{D}^{33}(x)$ are strictly diagonally dominant,

$$\left\| \left[\widetilde{\mathbf{D}}^{ii}(x)\right]^{-1} \mathbf{D}^{ij}(x) \right\|_{(\alpha,\beta)} \leqslant K_0 \tag{81}$$

for $i = 1, 3$ and $j = 1, 2, 3$, and

$$\|\mathbf{D}^{2j}(x)\|_{(\alpha,\beta)} \leqslant K_0 \tag{82}$$

for $j = 1, 2, 3$. Here $\widetilde{\mathbf{D}}^{ii}(x)$ is the diagonal matrix containing the diagonal elements of $\mathbf{D}^{ii}(x)$ and $K_0\Delta x \ll 1$.

Note that in this definition, there is no requirement that $\|\mathbf{D}^{ij}(x)\|_{(\alpha,\beta)}$ for $i = 1, 3$ and $j = 1, 2, 3$ is small. As we will shortly see, the first and last row blocks will correspond to the stiff portion of the BVP. Also note that there is nothing special about submatrices $\mathbf{D}^{11}(x)$ and $\mathbf{D}^{33}(x)$. The definition is written with those blocks being strictly diagonally dominant because the algorithm we will use to solve the BVP will put the system in this form. Now we are in a position to determine if a given mesh resolves the solution to our BVP.

Theorem 1 (Kreiss et al. [43])**.** *Consider the BVP*

$$\frac{d\mathbf{w}(x)}{dx} = \mathbf{D}(x)\mathbf{w}(x) + \mathbf{g}(x), \qquad a \leqslant x \leqslant b \tag{83}$$

where $\mathbf{D}(x)$ is partitioned as in (80). Let $\mathbf{w}(x)$ and $\mathbf{g}(x)$ be similarly partitioned so that $\mathbf{w}^i(x), \mathbf{g}^i(x) \in \mathbb{R}^{m_i}$ for $i = 1, 2, 3$. If $\mathbf{D}(x)$ is essentially diagonally dominant on (α, β) and there are constants K_1 and K_2 such that

$$\left\| \left[\widetilde{\mathbf{D}}^{ii}(x)\right]^{-1} \frac{d^\nu \mathbf{D}^{ij}(x)}{dx^\nu} \right\|_{(\alpha,\beta)} \leqslant K_1 \quad \textit{and} \quad \left\| \left[\widetilde{\mathbf{D}}^{ii}(x)\right]^{-1} \frac{d^\nu \mathbf{g}^{i}(x)}{dx^\nu} \right\|_{(\alpha,\beta)} \leqslant K_1 \tag{84}$$

for $i = 1, 3$, $j = 1, 2, 3$, *and* $\nu = 0, 1, \ldots, p$;

$$\left\| \frac{d^\nu \mathbf{D}^{2j}(x)}{dx^\nu} \right\|_{(\alpha,\beta)} \leqslant K_1 \quad \textit{and} \quad \left\| \frac{d^\nu \mathbf{g}^2(x)}{dx^\nu} \right\|_{(\alpha,\beta)} \leqslant K_1 \tag{85}$$

for $j = 1, 2, 3$ *and* $\nu = 0, 1, \ldots, p$; *and*

$$\left\| \widetilde{\mathbf{D}}^{11}(a) \right\|_{(\alpha,\beta)} \leqslant K_2 \quad \textit{and} \quad \left\| \widetilde{\mathbf{D}}^{33}(b) \right\|_{(\alpha,\beta)} \leqslant K_2, \tag{86}$$

then $\mathbf{w}(x)$ *is resolved on* (α, β). *Here again,* $\widetilde{\mathbf{D}}^{ii}(x)$ *is the diagonal matrix containing the diagonal elements of* $\mathbf{D}^{ii}(x)$, $K_1 \Delta x \ll 1$, *and* $K_2 \Delta x \ll 1$.

Theorem 1 requires the matrix of the BVP to be essentially diagonally dominant. This will not in general be the case. Thus, we must find a way to transform the BVP (75) to this form. Suppose we are able to find an invertible matrix function $\mathbf{V}(x) \in \mathbb{R}^{M \times M}$ such that

$$\mathbf{V}^{-1}(x)\mathbf{A}(x)\mathbf{V}(x) = \mathbf{\Lambda}(x) = \begin{bmatrix} \mathbf{\Lambda}^{11}(x) & & \\ & \mathbf{\Lambda}^{22}(x) & \\ & & \mathbf{\Lambda}^{33}(x) \end{bmatrix} \tag{87}$$

where $\mathbf{\Lambda}^{ii}(x) \in \mathbb{R}^{m_i \times m_i}$ and $m_1 + m_2 + m_3 = M$. Here, the eigenvalues of $\mathbf{\Lambda}^{11}(x)$ are large and negative, the eigenvalues of $\mathbf{\Lambda}^{33}(x)$ are large and positive, and the eigenvalues (both positive and negative) of $\mathbf{\Lambda}^{22}(x)$ are small. Using (87), the BVP (75) becomes

$$\frac{d\mathbf{y}(x)}{dx} = \mathbf{V}(x)\mathbf{\Lambda}(x)\mathbf{V}^{-1}(x)\mathbf{y}(x) + \mathbf{f}(x). \tag{88}$$

Let $\mathbf{w}(x) = \mathbf{V}^{-1}(x)\mathbf{y}(x)$. Then

$$\frac{d\mathbf{w}(x)}{dx} = \mathbf{D}(x)\mathbf{w}(x) + \mathbf{g}(x) \tag{89}$$

where $\mathbf{D}(x) = \mathbf{\Lambda}(x) - \mathbf{V}^{-1}(x)\mathbf{V}'(x)$ (the prime denotes a derivative with respect to x) and $\mathbf{g}(x) = \mathbf{V}^{-1}(x)\mathbf{f}(x)$. With Theorem 1 in mind, suppose that $\mathbf{\Lambda}^{11}(x)$ and $\mathbf{\Lambda}^{33}(x)$ are strictly diagonally dominant. Now if $\left\| \mathbf{V}^{-1}(x)\mathbf{V}'(x) \right\|_{(\alpha,\beta)}$ is small, then $\mathbf{D}(x)$ is essentially diagonally dominant on (α, β). Hence, the application of Theorem 1 relies on our ability to find an appropriate matrix function $\mathbf{V}(x)$.

Before we do this however, (87) and (89) are suggestive of a finite difference method. If an IVP or a TVP have large eigenvalues, it is known that methods such as backward Euler work very well. If either an IVP or TVP has only small eigenvalues, then the trapezoidal rule is more accurate. With this in mind, let us use backward Euler for the $\mathbf{\Lambda}^{11}(x)$ block integrating from a to b, the trapezoidal rule for the $\mathbf{\Lambda}^{22}(x)$ block (integration direction does not matter since it is a symmetric method), and backward Euler for the $\mathbf{\Lambda}^{33}(x)$ block integrating from b to a. This finite difference method gives us

$$\begin{aligned} &- \begin{bmatrix} \mathbf{I}_{m_1} & & \\ \frac{\Delta x_n}{2}\mathbf{D}_{n-1}^{21} & \mathbf{I}_{m_2} + \frac{\Delta x_n}{2}\mathbf{D}_{n-1}^{22} & \frac{\Delta x_n}{2}\mathbf{D}_{n-1}^{23} \\ \Delta x_n \mathbf{D}_{n-1}^{31} & \Delta x_n \mathbf{D}_{n-1}^{32} & \mathbf{I}_{m_3} + \Delta x_n \mathbf{D}_{n-1}^{33} \end{bmatrix} \mathbf{V}_{n-1}^{-1}\mathbf{u}_{n-1} \\ &+ \begin{bmatrix} \mathbf{I}_{m_1} - \Delta x_n \mathbf{D}_n^{11} & -\Delta x_n \mathbf{D}_n^{12} & -\Delta x_n \mathbf{D}_n^{13} \\ -\frac{\Delta x_n}{2}\mathbf{D}_n^{21} & \mathbf{I}_{m_2} - \frac{\Delta x_n}{2}\mathbf{D}_n^{22} & -\frac{\Delta x_n}{2}\mathbf{D}_n^{23} \\ & & \mathbf{I}_{m_3} \end{bmatrix} \mathbf{V}_n^{-1}\mathbf{u}_n = \Delta x_n \begin{bmatrix} \mathbf{g}_n^1 \\ \frac{1}{2}(\mathbf{g}_{n-1}^2 + \mathbf{g}_n^2) \\ \mathbf{g}_{n-1}^3 \end{bmatrix} \end{aligned} \tag{90}$$

for $n = 1, 2, \ldots, N$ where $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{u}_n \approx \mathbf{y}(x_n)$. Also, we use the shorthand $\mathbf{V}_n^{-1} = \mathbf{V}^{-1}(x_n)$ and similarly for $\mathbf{D}^{ij}(x_n)$ and $\mathbf{g}^i(x_n)$. This notation will be used throughout this section.

To numerically solve the BVP (75), the finite difference equations (90) and the boundary conditions (76) are assembled into a system of equations of size $M(N+1)$. This turns out to be a very sparse system. To solve this, we could use an incomplete LU factorization as a preconditioner and a stabilized biconjugate gradient method. In MATLAB and GNU Octave, the functions `ilu` and `bicgstab` are available for this.

3.5 The Schur Method

We now turn our attention to finding an appropriate $\mathbf{V}(x)$. From (87), we see that we need a similarity transform. One way to find $\mathbf{V}(x)$ is to find a series of similarity transforms and construct $\mathbf{V}(x)$ from those. At any mesh point x_n we can find the Schur decomposition of $\mathbf{A}(x_n)$. This is given by

$$\widetilde{\mathbf{V}}_n^{-1}\mathbf{A}_n\widetilde{\mathbf{V}}_n = \widetilde{\mathbf{\Lambda}}_n = \begin{bmatrix} \widetilde{\mathbf{\Lambda}}^{11} & \widetilde{\mathbf{\Lambda}}^{12} & \widetilde{\mathbf{\Lambda}}^{13} \\ & \widetilde{\mathbf{\Lambda}}^{22} & \widetilde{\mathbf{\Lambda}}^{23} \\ & & \widetilde{\mathbf{\Lambda}}^{33} \end{bmatrix} \tag{91}$$

where $\widetilde{\mathbf{V}}_n$ is an orthogonal matrix and $\widetilde{\mathbf{\Lambda}}_n$ is an upper triangular matrix. This implies that $\widetilde{\mathbf{\Lambda}}^{11}$, $\widetilde{\mathbf{\Lambda}}^{22}$, and $\widetilde{\mathbf{\Lambda}}^{33}$ are upper triangular, so the eigenvalues of $\mathbf{A}_n$ are on the diagonal of $\widetilde{\mathbf{\Lambda}}_n$. Further, the Schur decomposition can be done in such a way that the large negative eigenvalues are on the diagonal of $\widetilde{\mathbf{\Lambda}}^{11}$, the large positive eigenvalues are on the diagonal of $\widetilde{\mathbf{\Lambda}}^{33}$, and the small eigenvalues (both positive and negative) are on the diagonal of $\widetilde{\mathbf{\Lambda}}^{22}$. In MATLAB and GNU Octave, the functions `schur` and `ordschur` accomplish this task.

Next, we zero out the remaining off-diagonal blocks. Let

$$\widehat{\mathbf{V}}_n = \begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{S}_1 & \mathbf{S}_3 \\ & \mathbf{I}_{m_2} & \mathbf{S}_2 \\ & & \mathbf{I}_{m_3} \end{bmatrix} \iff \widehat{\mathbf{V}}_n^{-1} = \begin{bmatrix} \mathbf{I}_{m_1} & -\mathbf{S}_1 & \mathbf{S}_1\mathbf{S}_2 - \mathbf{S}_3 \\ & \mathbf{I}_{m_2} & -\mathbf{S}_2 \\ & & \mathbf{I}_{m_3} \end{bmatrix} \tag{92}$$

so that we obtain

$$\widehat{\mathbf{V}}_n^{-1}\widetilde{\mathbf{\Lambda}}_n\widehat{\mathbf{V}}_n = \widehat{\mathbf{\Lambda}}_n = \begin{bmatrix} \widehat{\mathbf{\Lambda}}_n^{11} & & \\ & \widehat{\mathbf{\Lambda}}_n^{22} & \\ & & \widehat{\mathbf{\Lambda}}_n^{33} \end{bmatrix}. \tag{93}$$

This will occur if $\mathbf{S}_1$, $\mathbf{S}_2$, and $\mathbf{S}_3$ solve the Sylvester equations

$$\widetilde{\mathbf{\Lambda}}^{11}\mathbf{S}_1 - \mathbf{S}_1\widetilde{\mathbf{\Lambda}}^{22} = -\widetilde{\mathbf{\Lambda}}^{12}, \tag{94}$$

$$\widetilde{\mathbf{\Lambda}}^{22}\mathbf{S}_2 - \mathbf{S}_2\widetilde{\mathbf{\Lambda}}^{33} = -\widetilde{\mathbf{\Lambda}}^{23}, \tag{95}$$

$$\widetilde{\mathbf{\Lambda}}^{11}\mathbf{S}_3 - \mathbf{S}_3\widetilde{\mathbf{\Lambda}}^{33} = -\widetilde{\mathbf{\Lambda}}^{13} - \widetilde{\mathbf{\Lambda}}^{12}\mathbf{S}_2. \tag{96}$$

Two algorithms for solving the Sylvester equation are the Bartels-Stewart algorithm and the Hessenberg-Schur algorithm, both of which are given in Golub et al. [25]. The Sylvester equation $\mathbf{AX} - \mathbf{XB} = \mathbf{C}$ has a unique solution if and only if $\mathbf{A}$ and $\mathbf{B}$ do not share any eigenvalues (see Golub et al. [25]), which is guaranteed due to how we have defined the blocks of $\widetilde{\mathbf{\Lambda}}_n$. Routines for solving the Sylvester equation are implemented in MATLAB and GNU Octave with the function `lyap`.

As stated above, the diagonal blocks of (93) are upper triangular, but $\widehat{\mathbf{\Lambda}}^{11}$ and $\widehat{\mathbf{\Lambda}}^{33}$ are not necessarily strictly diagonally dominant as required in Theorem 1. Fortunately, it is simple to find a similarity transform to make an upper triangular matrix strictly diagonally dominant. Let $\mathbf{Q} \in \mathbb{R}^{m \times m}$ be a diagonal matrix and $\mathbf{\Lambda} \in \mathbb{R}^{m \times m}$ be an upper triangular matrix. Then the product

$$\mathbf{Q}^{-1}\mathbf{\Lambda}\mathbf{Q} = \begin{bmatrix} \lambda_{11} & q_1^{-1}q_2\lambda_{12} & \cdots & q_1^{-1}q_m\lambda_{1m} \\ & \lambda_{22} & \cdots & q_2^{-1}q_m\lambda_{2m} \\ & & \ddots & \vdots \\ & & & \lambda_{mm} \end{bmatrix} \tag{97}$$

is strictly diagonally dominant if we choose $q_m = 1$ and

$$q_i = \frac{\sum_{j=i+1}^{m} q_j|\lambda_{ij}|}{\gamma|\lambda_{ii}|} \tag{98}$$

for $i = m-1, m-2, \ldots, 1$ and $0 < \gamma < 1$. With this in mind, let

$$\bar{\mathbf{V}}_n = \begin{bmatrix} \mathbf{Q}^{11} & & \\ & \mathbf{I}_{m_2} & \\ & & \mathbf{Q}^{33} \end{bmatrix} \iff \bar{\mathbf{V}}_n^{-1} = \begin{bmatrix} (\mathbf{Q}^{11})^{-1} & & \\ & \mathbf{I}_{m_2} & \\ & & (\mathbf{Q}^{33})^{-1} \end{bmatrix} \tag{99}$$

where $\mathbf{Q}^{11}$ and $\mathbf{Q}^{33}$ are diagonal matrices with elements set according to (97) and (98) so that

$$\bar{\mathbf{V}}_n^{-1}\widehat{\mathbf{\Lambda}}_n\bar{\mathbf{V}}_n = \mathbf{\Lambda}_n = \begin{bmatrix} \mathbf{\Lambda}^{11} & & \\ & \mathbf{\Lambda}^{22} & \\ & & \mathbf{\Lambda}^{33} \end{bmatrix}. \tag{100}$$

We now have the desired block diagonal matrix $\mathbf{\Lambda}_n$ with strictly diagonally dominant matrices $\mathbf{\Lambda}^{11}$ and $\mathbf{\Lambda}^{33}$. However, we still need $\left\|\mathbf{V}_n^{-1}\mathbf{V}_n'\right\|$ to be small. From (91), (93), and (100), let

$$\widetilde{\mathbf{V}}_n\widehat{\mathbf{V}}_n\bar{\mathbf{V}}_n = \mathbf{U}_n = \begin{bmatrix} \mathbf{U}^{11} & \mathbf{U}^{12} & \mathbf{U}^{13} \\ \mathbf{U}^{21} & \mathbf{U}^{22} & \mathbf{U}^{23} \\ \mathbf{U}^{31} & \mathbf{U}^{32} & \mathbf{U}^{33} \end{bmatrix}. \tag{101}$$

Also define

$$\frac{1}{d_i} = \left\| \begin{bmatrix} \mathbf{U}^{1i} \\ \mathbf{U}^{2i} \\ \mathbf{U}^{3i} \end{bmatrix} \right\| \tag{102}$$

for $i = 1, 2, 3$. Now let

$$\breve{\mathbf{V}}_n = \begin{bmatrix} d_1\mathbf{I}_{m_1} & & \\ & d_2\mathbf{I}_{m_2} & \\ & & d_3\mathbf{I}_{m_3} \end{bmatrix} \Longleftrightarrow \breve{\mathbf{V}}_n^{-1} = \begin{bmatrix} d_1^{-1}\mathbf{I}_{m_1} & & \\ & d_2^{-1}\mathbf{I}_{m_2} & \\ & & d_3^{-1}\mathbf{I}_{m_3} \end{bmatrix} \tag{103}$$

so that

$$\breve{\mathbf{V}}_n^{-1}\mathbf{\Lambda}_n\breve{\mathbf{V}}_n = \mathbf{\Lambda}_n = \begin{bmatrix} \mathbf{\Lambda}^{11} & & \\ & \mathbf{\Lambda}^{22} & \\ & & \mathbf{\Lambda}^{33} \end{bmatrix}. \tag{104}$$

Note that this last similarity transform does not change $\mathbf{\Lambda}_n$. It only has the effect of scaling $\mathbf{V}_n$ so that $\left\|\mathbf{V}_n^{-1}\mathbf{V}_n'\right\|$ is small. This scaling is effective because right-multiplying $\mathbf{U}_n$ by $\breve{\mathbf{V}}_n$ normalizes the "column blocks" of $\mathbf{U}_n$. This makes all elements of $\mathbf{U}_n$ of moderate size. Similarly, left-multiplying $\mathbf{U}_n^{-1}$ by $\breve{\mathbf{V}}_n^{-1}$ normalizes the "row blocks" of $\mathbf{U}_n^{-1}$. Certainly, other choices of $\breve{\mathbf{V}}_n$ may be better in this regard, but (104) has been found to be adequate for the electron transport problem.

Finally, we find using (91), (93), (100), and (104) that if we let

$$\mathbf{V}_n = \widetilde{\mathbf{V}}_n\widehat{\mathbf{V}}_n\bar{\mathbf{V}}_n\breve{\mathbf{V}}_n, \tag{105}$$

then we obtain the necessary similarity transformation for (87).

3.6 The Riccati Method

The Schur method would be expensive if calculated for every mesh point and does not guarantee a smooth $\mathbf{V}(x)$. However, if we already have $\mathbf{V}_{n-1}$, then

$$\mathbf{V}_{n-1}^{-1}\mathbf{A}_n\mathbf{V}_{n-1} = \widetilde{\mathbf{\Lambda}}_n = \begin{bmatrix} \widetilde{\mathbf{\Lambda}}^{11} & \widetilde{\mathbf{\Lambda}}^{12} & \widetilde{\mathbf{\Lambda}}^{13} \\ \widetilde{\mathbf{\Lambda}}^{21} & \widetilde{\mathbf{\Lambda}}^{22} & \widetilde{\mathbf{\Lambda}}^{23} \\ \widetilde{\mathbf{\Lambda}}^{31} & \widetilde{\mathbf{\Lambda}}^{32} & \widetilde{\mathbf{\Lambda}}^{33} \end{bmatrix} \tag{106}$$

can be viewed as a perturbation to $\mathbf{\Lambda}_{n-1}$ so long as Δx_n is sufficiently small. That is, the off-diagonal blocks should be small. If we can eliminate the off-diagonal blocks through similarity transforms, then we can avoid calculating the Schur decomposition of $\mathbf{A}_n$.

Let us partition $\widetilde{\mathbf{\Lambda}}_n$ so that

$$\widetilde{\mathbf{\Lambda}}_n = \left[\begin{array}{cc|c} \widetilde{\mathbf{\Lambda}}^{11} & \widetilde{\mathbf{\Lambda}}^{12} & \widetilde{\mathbf{\Lambda}}^{13} \\ \widetilde{\mathbf{\Lambda}}^{21} & \widetilde{\mathbf{\Lambda}}^{22} & \widetilde{\mathbf{\Lambda}}^{23} \\ \hline \widetilde{\mathbf{\Lambda}}^{31} & \widetilde{\mathbf{\Lambda}}^{32} & \widetilde{\mathbf{\Lambda}}^{33} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{B}^{11} & \mathbf{B}^{12} \\ \hline \mathbf{B}^{21} & \mathbf{B}^{22} \end{array}\right] = \mathbf{B}_n. \tag{107}$$

Now let

$$\widetilde{\mathbf{V}}_n = \begin{bmatrix} \mathbf{I}_{m_1+m_2} & \mathbf{R}_1 \\ & \mathbf{I}_{m_3} \end{bmatrix} \Longleftrightarrow \widetilde{\mathbf{V}}_n^{-1} = \begin{bmatrix} \mathbf{I}_{m_1+m_2} & -\mathbf{R}_1 \\ & \mathbf{I}_{m_3} \end{bmatrix}. \tag{108}$$

Then we obtain

$$\widetilde{\mathbf{V}}_n^{-1}\mathbf{B}_n\widetilde{\mathbf{V}}_n = \widetilde{\mathbf{B}}_n = \begin{bmatrix} \widetilde{\mathbf{B}}^{11} & \\ \widetilde{\mathbf{B}}^{21} & \widetilde{\mathbf{B}}^{22} \end{bmatrix} \tag{109}$$

if $\mathbf{R}_1$ solves the algebraic Riccati equation

$$\mathbf{B}^{11}\mathbf{R}_1 - \mathbf{R}_1\mathbf{B}^{22} = \mathbf{R}_1\mathbf{B}^{21}\mathbf{R}_1 - \mathbf{B}^{12}. \tag{110}$$

Theorem 2 (Kreiss et al. [43]). *Let $\mathbf{B}^{ij} \in \mathbb{R}^{m_i \times m_j}$ for $i, j = 1, 2$. Also let $\mathbf{R}^{(0)} = \mathbf{O}^{12}$ where $\mathbf{O}^{12}$ is the $m_1 \times m_2$ zero matrix. If $\|\mathbf{B}^{12}\|$ and $\|\mathbf{B}^{21}\|$ are sufficiently small, then the iteration*

$$\mathbf{B}^{11}\mathbf{R}^{(k)} - \mathbf{R}^{(k)}\mathbf{B}^{22} = \mathbf{R}^{(k-1)}\mathbf{B}^{21}\mathbf{R}^{(k-1)} - \mathbf{B}^{12}, \qquad k = 1, 2, \ldots \tag{111}$$

converges to a locally unique solution of

$$\mathbf{B}^{11}\mathbf{R} - \mathbf{R}\mathbf{B}^{22} = \mathbf{R}\mathbf{B}^{21}\mathbf{R} - \mathbf{B}^{12}. \tag{112}$$

Theorem 2 states that the algebraic Riccati equation (110) can be solved by solving a series of Sylvester equations for which we have already specified algorithms and software. We continue by letting

$$\widehat{\mathbf{V}}_n = \begin{bmatrix} \mathbf{I}_{m_1+m_2} & \\ \mathbf{S}_1 & \mathbf{I}_{m_3} \end{bmatrix} \iff \widehat{\mathbf{V}}_n^{-1} = \begin{bmatrix} \mathbf{I}_{m_1+m_2} & \\ -\mathbf{S}_1 & \mathbf{I}_{m_3} \end{bmatrix}. \tag{113}$$

Then we find that

$$\widehat{\mathbf{V}}_n^{-1}\widetilde{\mathbf{B}}_n\widehat{\mathbf{V}}_n = \widehat{\mathbf{B}}_n = \begin{bmatrix} \widehat{\mathbf{B}}^{11} & \\ & \widehat{\mathbf{B}}^{22} \end{bmatrix} \tag{114}$$

provided that $\mathbf{S}_1$ solves the Sylvester equation

$$\widetilde{\mathbf{B}}^{22}\mathbf{S}_1 - \mathbf{S}_1\widetilde{\mathbf{B}}^{11} = -\widetilde{\mathbf{B}}^{21}. \tag{115}$$

From (107), (109), and (114) we find that

$$\widehat{\mathbf{V}}_n^{-1}\widetilde{\mathbf{V}}_n^{-1}\widetilde{\mathbf{\Lambda}}_n\widetilde{\mathbf{V}}_n\widehat{\mathbf{V}}_n = \widehat{\mathbf{\Lambda}}_n = \begin{bmatrix} \widehat{\mathbf{\Lambda}}^{11} & \widehat{\mathbf{\Lambda}}^{12} & \\ \widehat{\mathbf{\Lambda}}^{21} & \widehat{\mathbf{\Lambda}}^{22} & \\ & & \widehat{\mathbf{\Lambda}}^{33} \end{bmatrix} \tag{116}$$

We now need to zero out the remaining off-diagonal blocks. We can let

$$\bar{\mathbf{V}}_n = \begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{R}_2 & \\ & \mathbf{I}_{m_2} & \\ & & \mathbf{I}_{m_3} \end{bmatrix} \iff \bar{\mathbf{V}}_n^{-1} = \begin{bmatrix} \mathbf{I}_{m_1} & -\mathbf{R}_2 & \\ & \mathbf{I}_{m_2} & \\ & & \mathbf{I}_{m_3} \end{bmatrix} \tag{117}$$

so that

$$\bar{\mathbf{V}}_n^{-1}\widehat{\mathbf{\Lambda}}_n\bar{\mathbf{V}}_n = \bar{\mathbf{\Lambda}}_n = \begin{bmatrix} \bar{\mathbf{\Lambda}}^{11} & & \\ \bar{\mathbf{\Lambda}}^{21} & \bar{\mathbf{\Lambda}}^{22} & \\ & & \bar{\mathbf{\Lambda}}^{33} \end{bmatrix} \tag{118}$$

provided that $\mathbf{R}_2$ solves the algebraic Riccati equation

$$\bar{\mathbf{\Lambda}}^{11}\mathbf{R}_2 - \mathbf{R}_2\bar{\mathbf{\Lambda}}^{22} = \mathbf{R}_2\bar{\mathbf{\Lambda}}^{21}\mathbf{R}_2 - \bar{\mathbf{\Lambda}}^{12}. \tag{119}$$

Similar to before, we can now let

$$\breve{\mathbf{V}}_n = \begin{bmatrix} \mathbf{I}_{m_1} & & \\ \mathbf{S}_2 & \mathbf{I}_{m_2} & \\ & & \mathbf{I}_{m_3} \end{bmatrix} \iff \breve{\mathbf{V}}_n^{-1} = \begin{bmatrix} \mathbf{I}_{m_1} & & \\ -\mathbf{S}_2 & \mathbf{I}_{m_2} & \\ & & \mathbf{I}_{m_3} \end{bmatrix}. \tag{120}$$

With this, we obtain

$$\breve{\mathbf{V}}_n^{-1}\bar{\mathbf{\Lambda}}_n\breve{\mathbf{V}}_n = \mathbf{\Lambda}_n = \begin{bmatrix} \mathbf{\Lambda}^{11} & & \\ & \mathbf{\Lambda}^{22} & \\ & & \mathbf{\Lambda}^{33} \end{bmatrix} \tag{121}$$

if $\mathbf{S}_2$ solves the Sylvester equation

$$\bar{\mathbf{\Lambda}}^{22}\mathbf{S}_2 - \mathbf{S}_2\bar{\mathbf{\Lambda}}^{11} = -\bar{\mathbf{\Lambda}}^{21}. \tag{122}$$

Finally, we find using (106), (116), (118), and (121) that

$$\mathbf{V}_n = \mathbf{V}_{n-1}\widetilde{\mathbf{V}}_n\widehat{\mathbf{V}}_n\bar{\mathbf{V}}_n\breve{\mathbf{V}}_n, \tag{123}$$

gives the necessary similarity transformation for (87).

There are two ways that the Riccati method can fail. The first is if the off-diagonal blocks of (106) are too large, so Theorem 2 does not apply. To fix this, we can simply decrease Δx_n until the off-diagonal blocks are small enough. Typically, we know that Δx_n is too large if too many iterations in (111) are used. The second way the Riccati method can fail is if the block structure of $\mathbf{A}(x)$ changes from x_{n-1} to x_n. That is, if one or more eigenvalues of $\mathbf{A}(x)$ go from being small to large or vice versa, then the size of the submatrices $\mathbf{\Lambda}^{11}$, $\mathbf{\Lambda}^{22}$, and $\mathbf{\Lambda}^{33}$ (i.e. their dimensions m_1, m_2, and m_3) change. When this occurs, we simply resort to the Schur method.

3.7 Small and Large Eigenvalues

A point that we have neglected thus far is what we mean by small and large eigenvalues. In order to quantify this, consider the test equation for a scalar IVP

$$\frac{dy(x)}{dx} = \lambda y(x), \qquad y(x_0) = y_0 \tag{124}$$

where $\lambda < 0$. The exact solution is $y(x) = e^{\lambda x}y_0$, but we will instead write the solution as

$$y(x_n) = e^{\lambda\Delta x_n}y(x_{n-1}). \tag{125}$$

If $\lambda\Delta x_n$ is small enough, then the trapezoidal rule

$$u_n = \frac{1 + \lambda\Delta x_n/2}{1 - \lambda\Delta x_n/2} u_{n-1} \tag{126}$$

gives an adequate approximation. If λ is very large, then it is well known that the trapezoidal rule does not give accurate results unless if Δx_n is very small. In certain cases, λ can be so large that Δx_n must be decreased to the point of impracticality. In this case, a method such as backward Euler gives better results without having to reduce Δx_n. The backward Euler method for the test equation gives

$$u_n = \frac{1}{1 - \lambda\Delta x_n} u_{n-1}. \tag{127}$$

To know whether the eigenvalue is small or large is to know whether (126) or (127) is a better approximation to (125). For both numerical methods, if $z = \lambda\Delta x_n$, we can write

$$u_n = G(z)u_{n-1} \tag{128}$$

where $G(z)$ is the growth function. We would like to have $G(z)$ emulate the exact solution as much as possible. That is, we want $|G(z) - e^z|$ to be small. In Figure 5, we plot the growth function error for both backward Euler and the trapezoidal rule.

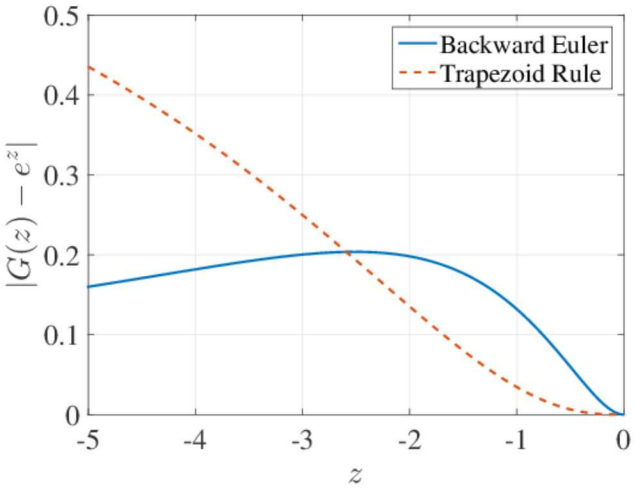

Figure 5. The growth function error as a function of $z = \lambda\Delta x_n$.

The errors are the same for $z \approx -2.59$. So the trapezoidal rule gives less error if $z > -2.59$ and backward Euler gives less error if $z < -2.59$. However, from (126), we see that if $z < -2$ the trapezoidal rule growth function is negative. This causes the numerical solution to oscillate between positive and negative. This is undesirable in problems where the solution is strictly nonnegative such as the electron transport problem. So if using backward Euler and the trapezoidal rule, we say an eigenvalue λ of (75) is small with respect to the local mesh spacing Δx_n if $|\lambda|\Delta x_n < 2$ and large if $|\lambda|\Delta x_n \geqslant 2$.

3.8 Implementing the Numerical Solution

Before stating the algorithm, a few remarks should be made. First, in Theorem 1, we need to choose constants K_i for $i = 1, 2$ and the degree of smoothness p. For choosing K_i, we simply need $K_i \Delta x \ll 1$, so the choice of K_i is problem dependent. Regardless, the smaller K_i is, the more points the algorithm will use. As for p, it has been found to be sufficient to let $p = 1$. This way, when the derivatives needed in Theorem 1 are calculated, we can use a first order finite difference approximation with only two points. The procedure is summarized in Algorithm 1.

Algorithm 1 Implementation of the Upwind Solution

1: choose a preliminary mesh $a = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = b$ that satisfies (86)
2: use the Schur method to find $\mathbf{V}(x_0)$ and $\mathbf{\Lambda}(x_0)$ and calculate $\mathbf{D}(x_0)$
3: set $n = 1$
4: **while** $x_n \neq b$ **do**
5: use the Riccati method to find $\mathbf{V}(x_n)$ and $\mathbf{\Lambda}(x_n)$
6: **if** Riccati iteration (111) does not converge **then**
7: replace Δx_n with $\Delta x_n/\sqrt{2}$, update x_n, and go to step 5
8: **else if** block structure of $\mathbf{\Lambda}(x_n)$ and $\mathbf{\Lambda}(x_{n-1})$ differs **then**
9: use the Schur method to replace $\mathbf{V}(x_{n-1})$ and $\mathbf{\Lambda}(x_{n-1})$ with block structure forced to be the same as $\mathbf{\Lambda}(x_n)$
10: **end if**
11: calculate $\mathbf{D}(x_n)$
12: check smoothness with (84) and (85)
13: **if** not smooth enough **then**
14: replace Δx_n with $\Delta x_n/\sqrt{2}$, update x_n, and go to step 5
15: **end if**
16: accept Δx_n and adjust mesh accordingly
17: calculate finite difference matrices (90) and store
18: replace n with $n + 1$
19: **end while**
20: assemble finite difference matrices (90) and boundary conditions (76) into a linear system
21: solve system to find solution $\mathbf{u}_n \approx \mathbf{y}(x_n)$ for $n = 0, 1, \ldots, N$

A few more remarks about Algorithm 1 are in order. First, the number of points $N + 1$ changes as the algorithm proceeds. Also, this algorithm leaves open the possibility that there will be an abrupt change in mesh spacing. That is, either $\Delta x_n/\Delta x_{n-1} \ll 1$ or $\Delta x_n/\Delta x_{n-1} \gg 1$. This can give spurious results in the numerical solution, but can be remedied by adding more points so that the ratio $\Delta x_n/\Delta x_{n-1} \in [\frac{1}{2}, 2]$. Lastly in step 9, we use the Schur method but alter how the blocking is done. That is, instead of setting the size of the blocks according to the size of the eigenvalues at x_{n-1}, we set the size of the blocks according to the size of the eigenvalues at x_n. It should also be pointed out that a good approximation to the eigenvalues of $\mathbf{D}(x)$ are its diagonal elements since it is essentially diagonally dominant.

4 Boundary Element Solution of a Reduced Problem

4.1 Verification of the Upwind Method

Numerical experiments help in establishing the validity of a numerical method and increases confidence that it works in general. However, these experiments do not necessarily show that the method works on real world BVPs. Hence, when it is possible to solve the problem using a different method, it is advantageous to do so. If the two solutions agree, then we can further increase our confidence that the upwind method is accurately finding the solution. However, Trucano et al. [103] shows that there are many pitfalls to doing a code comparison and caution must be taken. After all, two codes giving the same answer does not necessarily mean either code is correct. For this reason, we will find the exact solution to a simplified electron transport problem. This way, the only error introduced will be in evaluating the expression giving that solution.

The boundary element method is a numerical technique that is useful if the Green's function or a fundamental solution can be found (see Kythe [48]). If either of these can be found, then it is possible to write the exact solution as an integral. The boundary element method then becomes a matter of evaluating this integral. Unfortunately, finding the Green's function or a fundamental solution is not possible, or at least not obvious, for the electron transport equation due to the summation over species. However, if we let the atmosphere be composed of a single species, then a boundary element solution is possible, and the resulting solution can be compared to the finite difference solution. If we can determine that the two solutions agree with each other, then we can verify that the upwind method is working properly.

4.2 Preliminary Simplifications

As stated above, we will let the atmosphere be composed of a single species. Although this is not physically realistic, it allows us to make a change of variables that leads to an exact solution. This chapter is only concerned with verifying the finite difference solution, which can be applied to the full problem. The single species assumption eliminates the summation over species, and we can drop all ξ subscripts. The species we will choose is oxygen O because this is the most abundant species for most altitudes (see Figure 1). However, at low altitudes the density of oxygen is very small, so using only oxygen would eliminate the stiffness of the problem. This would defeat the purpose of using the upwind method described in the previous chapter. Even worse, many electrons would reach the ground, and we would have to replace the boundary condition at $z = z_{\text{bot}}$ to be some sort of reflection condition at $z = 0$. We can avoid both of these undesirable features if we set the oxygen density to be the sum of all densities in the upper atmosphere. This way, we can retain the essential mathematical features of the problem.

Let us begin by rewriting (61) using the above simplification. We get

$$\mu \sin\alpha \frac{\partial I(z,E,\mu)}{\partial z} = n(z)\Bigg(-\big[\sigma^{\text{tot}}(E) - f(E)\sigma^{\text{el}}(E)\big] I(z,E,\mu) + \big[1-f(E)\big]\sigma^{\text{el}}(E) \sum_{m=0}^{M-1} \frac{2m+1}{2} \chi_m^*(E) P_m(\mu) \int_{-1}^{1} P_m(\mu') I(z,E,\mu')\, d\mu' + q(z,E,\mu)\Bigg) \tag{129}$$

where $q(z,E,\mu)$ is the sum of all other terms in (61). We can write it this way because $q(z,E,\mu)$ only contains terms that depend on electron intensities at higher energies, which can be assumed to be known. It may seem like finding the Green's function or a fundamental solution is unlikely due to the complexity of the right-hand-side, but this is not the case. Let us define the scattering depth to be

$$\tau = \frac{\sigma^{\text{tot}} - f\sigma^{\text{el}}}{\sin\alpha} \int_z^{z_{\text{top}}} n(z')\, dz' \tag{130}$$

where we have dropped the E dependency. If $\hat{I}(\tau,\mu) = I(z,\mu)$, then our equation becomes

$$\mu \frac{\partial \hat{I}(\tau,\mu)}{\partial \tau} = \hat{I}(\tau,\mu) - \frac{c}{2} \sum_{m=0}^{M-1} (2m+1)\chi_m^* P_m(\mu) \int_{-1}^{1} P_m(\mu') \hat{I}(\tau,\mu')\, d\mu' + \hat{q}(\tau,\mu) \tag{131}$$

where

$$c = \frac{(1-f)\sigma^{\text{el}}}{\sigma^{\text{tot}} - f\sigma^{\text{el}}}, \tag{132}$$

$$\hat{q}(\tau,\mu) = \frac{-q(z,\mu)}{\sigma^{\text{tot}} - f\sigma^{\text{el}}}. \tag{133}$$

It is important to note that $c < 1$, which can be verified with (42). It can be shown that this is essential to using the boundary element method for this problem (see Case and Zweifel [17]). Under this change of variables, the boundary conditions become

$$\hat{I}(0,\mu) = I_{\text{top}}(\mu), \quad \mu < 0, \tag{134}$$

$$\hat{I}(\tau_{\max},\mu) = 0, \quad \mu > 0. \tag{135}$$

Notice that the "top of the atmosphere" is at $\tau = 0$ and the "bottom of the atmosphere" is at $\tau = \tau_{\max}$. Hence, scattering depth (130) is a dimensionless measure of how far an electron has penetrated into the atmosphere.

4.3 The Homogeneous Solution

In order to find the Green's function or a fundamental solution, we must be able to find the homogeneous solution of (131). The procedure we will follow is due to Case [16], who

solved a similar problem for the case of isotropic scattering. This corresponds to $M = 1$ in (131). As M is increased, the scattering becomes anisotropic. The case of $M = 2$ was treated in depth by Kuščer et al. [47], McCormick [64], and Shure and Natelson [83] and was applied to neutron transport theory in Case and Zweifel [17]. Mika [68] extended parts of Case [16] to general M, but his ideas did not gain much traction in solving actual problems. Case and Zweifel [17] also briefly discuss the general M case, but not much detail is given. We wish to leave M general, so this chapter will give all the details of extending the method given in Case [16].

Looking for a homogeneous solution, we set $\hat{q}(\tau, \mu) = 0$ in (131). Dropping the hat notation, we use separation of variables and assume $I(\tau, \mu) = \psi(\tau)\varphi(\mu)$. Substituting this into (131) and dividing by $\mu\psi(\tau)\varphi(\mu)$, we find

$$\frac{\psi'(\tau)}{\psi(\tau)} = \frac{1}{\mu} - \frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu)\frac{\int_{-1}^{1} P_m(\mu')\varphi(\mu')\,d\mu'}{\mu\varphi(\mu)} = \frac{1}{\nu} \tag{136}$$

where $1/\nu$ is the separation constant. We will call ν an eigenvalue and $\varphi(\mu, \nu)$ an eigenfunction to (131). Solving for $\psi(\tau, \nu)$ is simple, and we find that

$$\psi(\tau, \nu) = e^{\tau/\nu} \tag{137}$$

where we have set the arbitrary constant to unity.

To find $\varphi(\mu, \nu)$, multiply (136) by $\mu\varphi(\mu, \nu)$ and rearrange to find

$$\left(1 - \frac{\mu}{\nu}\right)\varphi(\mu, \nu) = \frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu)\int_{-1}^{1} P_m(\mu')\varphi(\mu', \nu)\,d\mu'. \tag{138}$$

This is a homogeneous linear Fredholm integral equation of the second kind (see Tricomi [102]). Let us define

$$R_m(\nu) = \int_{-1}^{1} P_m(\mu)\varphi(\mu, \nu)\,d\mu, \tag{139}$$

which give us

$$\left(1 - \frac{\mu}{\nu}\right)\varphi(\mu, \nu) = \frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu)R_m(\nu), \tag{140}$$

but the $R_m(\nu)$ are still unknown. Choose the normalization of $\varphi(\mu, \nu)$ such that

$$R_0(\nu) = 1. \tag{141}$$

Now, integrate (140) over μ from -1 to 1. Since $P_1(\mu) = \mu$ and the Legendre polynomials are orthogonal, we find that

$$R_1(\nu) = (1 - c\chi_0^*)\nu. \tag{142}$$

Now if we multiply (140) by $P_\ell(\mu)$ and integrate over μ, we can use the well known Legendre polynomial recursion formula to find

$$(m+1)R_{m+1}(\nu) = (2m+1)(1 - c\chi_m^*)\nu R_m(\nu) - mR_{m-1}(\nu), \qquad m = 1, 2, \ldots, M-1 \tag{143}$$

Equations (141)–(143) show that $R_m(\nu)$ is a polynomial of degree m.

Now that $R_m(\nu)$ is known for all m, we can write

$$\left(1 - \frac{\mu}{\nu}\right)\varphi(\mu,\nu) = \frac{c}{2}D(\mu,\nu) \tag{144}$$

where

$$D(\mu,\nu) = \sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu)R_m(\nu) \tag{145}$$

is a known function. It will be useful later to know that $D(-\mu,-\nu) = D(\mu,\nu)$. This is simple enough to prove using the facts $P_m(-\mu) = (-1)^m P_m(\mu)$ and $R_m(-\nu) = (-1)^m R_m(\nu)$.

It would seem that solving for $\varphi(\mu,\nu)$ in (144) is simple, but it is only simple if $\nu \notin [-1,1]$. It turns out that there are a finite number of eigenvalues with this property (the discrete eigenvalues), but every $\nu \in [-1,1]$ is also an eigenvalue (the continuum eigenvalues). We will treat each case separately.

The Discrete Eigenvalues

For the discrete eigenvalues, $\nu \notin [-1,1]$. Hence, the eigenfunctions are given by

$$\varphi(\mu,\nu) = \frac{c\nu}{2}\frac{D(\mu,\nu)}{\nu-\mu}. \tag{146}$$

Now we need to find what values the eigenvalues can take. To do this, we integrate (146) over μ. Using (139), (141), and the fact that $P_0(\mu) = 1$, we find

$$1 = \frac{c\nu}{2}\int_{-1}^{1}\frac{D(\mu,\nu)}{\nu-\mu}\,d\mu. \tag{147}$$

Let us define the so-called dispersion function

$$\Lambda(z) = 1 + \frac{cz}{2}\int_{-1}^{1}\frac{D(\mu,z)}{\mu-z}\,d\mu. \tag{148}$$

The discrete eigenvalues are the zeros of $\Lambda(z)$.

A few properties of $\Lambda(z)$ are useful. One is that $\Lambda(-z) = \Lambda(z)$. This can be easily shown using the fact that $D(-\mu,-\nu) = D(\mu,\nu)$ mentioned above. This implies that if ζ is a zero of $\Lambda(z)$, then so is $-\zeta$. Useful for calculating $\Lambda(z)$ is the relation between Legendre polynomials (Legendre functions of the first kind) and Legendre functions of the second kind. This relationship is (see MacRobert [57])

$$Q_m(z) = \frac{1}{2}\int_{-1}^{1}\frac{P_m(\mu)}{z-\mu}\,d\mu, \qquad |z| > 1 \tag{149}$$

which immediately gives

$$\Lambda(z) = 1 - cz \sum_{m=0}^{M-1} (2m+1)\chi_m^* Q_m(z) R_m(z). \tag{150}$$

This is useful because the Legendre functions of the second kind satisfy the same recursion relation as the Legendre polynomials, making (150) simple to compute.

Another useful property is the limiting value of $\Lambda(z)$ as $z \to \infty$. From MacRobert [57], we know that

$$zP_m(z)Q_m(z) = \frac{1}{2}\int_{-1}^{1} \frac{\mu P_m^2(\mu)}{z-\mu}\, d\mu + \frac{1}{2m+1}, \tag{151}$$

which gives

$$\lim_{z\to\infty} zP_m(z)Q_m(z) = \frac{1}{2m+1}. \tag{152}$$

Also, from the Legendre polynomial recursion formula and (143), we know that

$$R_m(z) \sim P_m(z) \prod_{\ell=0}^{m-1} (1 - c\chi_\ell^*) \tag{153}$$

as $z \to \infty$. Therefore, it follows that

$$\begin{aligned}
\Lambda(\infty) &= \lim_{z\to\infty} \Lambda(z) \\
&= 1 - c \sum_{m=0}^{M-1} (2m+1)\chi_m^* \lim_{z\to\infty} zQ_m(z)R_m(z) \\
&= 1 - c \sum_{m=0}^{M-1} \chi_m^* \prod_{\ell=0}^{m-1} (1 - c\chi_\ell^*) \\
&= 1 - c\chi_0^* - c \sum_{m=1}^{M-1} \chi_m^* \prod_{\ell=0}^{m-1} (1 - c\chi_\ell^*) \\
&= (1 - c\chi_0^*)\left(1 - c \sum_{m=1}^{M-1} \chi_m^* \prod_{\ell=1}^{m-1} (1 - c\chi_\ell^*)\right) \\
&\vdots \\
&= \prod_{m=0}^{M-1} (1 - c\chi_m^*).
\end{aligned} \tag{154}$$

Equation (150) works well for z close to unity. For large z however, this formula becomes inaccurate in finite precision arithmetic since $Q_m(z) \to 0$ and $R_m(z) \to \infty$ as $z \to \infty$. A better formula is given by

$$\Lambda(z) = 1 + \frac{cz}{2}\int_{-1}^{1} \frac{D(\mu,\mu)}{\mu - z}\, d\mu, \tag{155}$$

which is almost the same as (148). They are, in fact, equivalent because

$$\begin{aligned}\int_{-1}^{1} \frac{D(\mu,z)}{\mu-z}\,d\mu &= \int_{-1}^{1} \frac{D(\mu,\mu)}{\mu-z}\,d\mu - \int_{-1}^{1} \frac{D(\mu,\mu)-D(\mu,z)}{\mu-z}\,d\mu \\ &= \int_{-1}^{1} \frac{D(\mu,\mu)}{\mu-z}\,d\mu - \sum_{m=0}^{M-1}(2m+1)\chi_m^* \int_{-1}^{1} \frac{R_m(\mu)-R_m(z)}{\mu-z}P_m(\mu)\,d\mu \\ &= \int_{-1}^{1} \frac{D(\mu,\mu)}{\mu-z}\,d\mu. \end{aligned} \tag{156}$$

Here, we used the fact that $(R_m(\mu)-R_m(z))/(\mu-z)$ is a polynomial of degree $m-1$, and $P_m(\mu)$ is orthogonal to any polynomial of degree less than m over $(-1,1)$.

At this point, the question of how many discrete eigenvalues exist needs to be answered. Before we do this, however, a few concepts should be developed. First is the notion of Hölder continuity. The formal definition of this is given next.

Definition 2 (Muskhelishvili [72])**.** A function $f(\mu)$ is *Hölder continuous* on $[-1,1]$ if for any two points $\mu_1, \mu_2 \in [-1,1]$

$$|f(\mu_1)-f(\mu_2)| \leqslant A|\mu_1-\mu_2|^{\alpha} \tag{157}$$

where A and α are positive constants.

For the remainder of this chapter, we will restrict ourselves to Hölder continuous functions unless otherwise specified. This is not a very restrictive requirement since any function that is differentiable or Lipschitz continuous is also Hölder continuous. Also, we will be dealing with functions that are holomorphic everywhere in the complex plane except on the cut from -1 to 1. An example of one such function is $\Lambda(z)$. We have to place a cut in the complex plane from -1 to 1 because the integral in (155) diverges at those points.

We will, however, be dealing with functions such as $\Lambda(z)$ evaluated at points just above and below the cut. These points will be denoted as

$$\Lambda^{\pm}(\mu) = \lim_{\varepsilon\to 0^+} \Lambda(\mu \pm i\varepsilon), \qquad \mu \in [-1,1] \tag{158}$$

where $i=\sqrt{-1}$. Further, we will be dealing with Cauchy principal value integrals. These are defined as

$$\fint_{-1}^{1} \frac{f(\mu)}{\mu-\nu}\,d\mu = \lim_{\varepsilon\to 0^+}\left(\int_{-1}^{\nu-\varepsilon} + \int_{\nu+\varepsilon}^{1}\right)\frac{f(\mu)}{\mu-\nu}\,d\mu \tag{159}$$

for $-1<\nu<1$. If $f(\mu)$ is Hölder continuous, then the above principal value integral exists

because

$$\begin{aligned}
\fint_{-1}^{1} \frac{f(\mu)}{\mu-\nu}\,d\mu &= \int_{-1}^{1} \frac{f(\mu)-f(\nu)}{\mu-\nu}\,d\mu + f(\nu)\fint_{-1}^{1} \frac{1}{\mu-\nu}\,d\mu \\
&= \int_{-1}^{1} \frac{f(\mu)-f(\nu)}{\mu-\nu}\,d\mu + f(\nu)\log\left(\frac{1-\nu}{1+\nu}\right) \\
&\leqslant \left|\int_{-1}^{1} \frac{f(\mu)-f(\nu)}{\mu-\nu}\,d\mu\right| + \left|f(\nu)\log\left(\frac{1-\nu}{1+\nu}\right)\right| \\
&\leqslant \int_{-1}^{1} \left|\frac{f(\mu)-f(\nu)}{\mu-\nu}\right| d\mu + \left|f(\nu)\log\left(\frac{1-\nu}{1+\nu}\right)\right| \\
&\leqslant \int_{-1}^{1} \frac{A|\mu-\nu|^{\alpha}}{|\mu-\nu|}\,d\mu + \left|f(\nu)\log\left(\frac{1-\nu}{1+\nu}\right)\right| \\
&= A\int_{-1}^{1} |\mu-\nu|^{\alpha-1}\,d\mu + \left|f(\nu)\log\left(\frac{1-\nu}{1+\nu}\right)\right|. \qquad (160)
\end{aligned}$$

The integral on the final line above exists because $\alpha > 0$ (i.e. at worst ν is an integrable singularity).

The following theorem is often referred to as the Sokhotski-Plemelj or simply the Plemelj theorm. It will be vital for the remainder of this chapter.

Theorem 3 (Muskhelishvili [72]). *Let $f(\mu)$ be Hölder continuous on $[-1,1]$. If*

$$\Phi(z) = \frac{1}{2\pi i}\int_{-1}^{1} \frac{f(\mu)}{\mu-z}\,d\mu, \qquad (161)$$

then

$$\Phi^{+}(\nu) + \Phi^{-}(\nu) = \frac{1}{i\pi}\fint_{-1}^{1} \frac{f(\mu)}{\mu-\nu}\,d\mu, \qquad (162)$$

$$\Phi^{+}(\nu) - \Phi^{-}(\nu) = \frac{1}{2}f(\nu) \qquad (163)$$

for any $-1 < \nu < 1$. Further,

$$\Phi(z) \to \frac{f(-1)}{2\pi i}\log\frac{1}{z+1} + g(z) \qquad (164)$$

as $z \to -1$ and

$$\Phi(z) \to -\frac{f(1)}{2\pi i}\log\frac{1}{z-1} + h(z) \qquad (165)$$

as $z \to 1$. Here, both $g(z)$ and $h(z)$ approach definite limits for -1 and 1, respectively, along any path.

Now we are in a position to answer how many discrete eigenvalues or how many zeros of $\Lambda(z)$ exist. Since the zeros of $\Lambda(z)$ come in plus/minus pairs, let us denote the number of

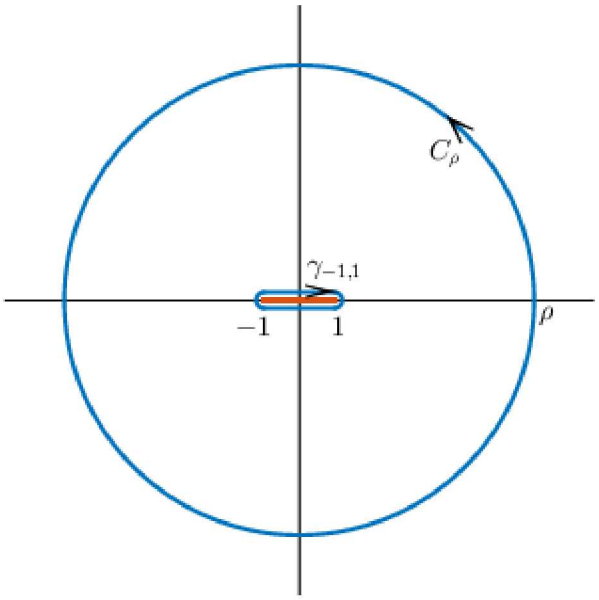

Figure 6. The contour $C_\rho + \gamma_{-1,1}$.

zeros as $2N$. Since $\Lambda(z)$ is holomorphic everywhere except on the cut from -1 to 1, we can use the argument principle of complex variables (see Ablowitz and Fokas [1]) to determine N. We have

$$2N = \frac{1}{2\pi i} \lim_{\rho\to\infty} \oint_{C_\rho+\gamma_{-1,1}} \frac{\Lambda'(z)}{\Lambda(z)}\, dz = \frac{1}{2\pi i} \lim_{\rho\to\infty} \oint_{C_\rho+\gamma_{-1,1}} \frac{d}{dz} \log \Lambda(z)\, dz \tag{166}$$

where C_ρ is the circle of radius ρ oriented in the counter-clockwise direction and $\gamma_{-1,1}$ is the contour surrounding the cut oriented in the clockwise direction (see Figure 6). Writing $\Lambda(z) = |\Lambda(z)|e^{i\Theta(z)}$, the argument principle becomes

$$2N = \frac{1}{2\pi} \lim_{\rho\to\infty} \Theta(z)\big|_{C_\rho+\gamma_{-1,1}} \tag{167}$$

which is the change in argument of $\Lambda(z)$ around the contour $C_\rho + \gamma_{-1,1}$ (hence the name argument principle). From (154), we see $\lim_{\rho\to\infty} \Theta(z)\big|_{C_\rho} = 0$ so that

$$2N = \frac{1}{2\pi}\Theta(z)\big|_{\gamma_{-1,1}}, \tag{168}$$

which is the argument of $\Lambda(z)$ around the cut.

Assume that $\Lambda^{\pm}(\nu) \neq 0$ for any $-1 < \nu < 1$. From Theorem 3 we have

$$\Lambda^{\pm}(\nu) = 1 + \frac{cz}{2} \fint_{-1}^{1} \frac{D(\mu,\mu)}{\mu-\nu}\, d\mu \pm \frac{i\pi c\nu}{2} D(\nu,\nu), \tag{169}$$

so this assumption requires that the real and imaginary parts do not simultaneously vanish (this certainly is possible for special values of c and χ_m^*, but in practice it doesn't occur – see Mika [68] to see how the following arguments change in case if it does). Also, notice that $\Lambda^+(\nu)$ and $\Lambda^-(\nu)$ are complex conjugates, which means $\Theta^+(\nu) = -\Theta^-(\nu)$. Further, it can be shown that $\Theta^+(\nu) = -\Theta^+(-\nu)$. With this, the argument principle becomes

$$\begin{aligned} N &= \frac{1}{4\pi}\Theta(z)\Big|_{\gamma_{-1,1}} \\ &= \frac{1}{4\pi}\Big[\Theta^+(\nu)\big|_{-1}^{1} + \Theta^-(\nu)\big|_{1}^{-1}\Big] \\ &= \frac{1}{2\pi}\Theta^+(\nu)\big|_{-1}^{1} \\ &= \frac{1}{\pi}\Theta^+(1). \end{aligned} \tag{170}$$

Now we can write

$$\Theta^+(\nu) = \tan^{-1}\left(\frac{\frac{\pi c\nu}{2}D(\nu,\nu)}{1 + \frac{c\nu}{2}\fint_{-1}^{1}\frac{D(\mu,\mu)}{\mu-\nu}\,d\mu}\right). \tag{171}$$

Since $D(\nu,\nu)$ is an even polynomial of degree $(M-1)^2$, it can have at most $M-1$ zeros between 0 and 1. Further, there is a possibility of a 2π increment in $\Theta^+(\nu)$ only when $\nu D(\nu,\nu) = 0$. Hence, there are at most M zeros of $\Lambda(z)$. This shows that the number of discrete eigenvalues is finite. Finally, Kuščer and Vidav [46] have shown that for $c < 1$, all eigenvalues are real.

The Continuum Eigenvalues

For the continuum eigenvalues, we have $\nu \in [-1,1]$. In this case, the eigenfunctions are distributions. The solution to (144) is now given by

$$\phi(\mu,\nu) = \frac{c\nu}{2}\mathcal{P}\frac{D(\mu,\nu)}{\nu-\mu} + \lambda(\nu)\delta(\nu-\mu) \tag{172}$$

where $\mathcal{P}$ denotes Cauchy principal value when the integral is absent (the dashed integral notation will be used otherwise). The Dirac delta term makes the expression as general as possible because $x\delta(x) = 0$. That is, (172) satisfies (144) for every $\nu \in [-1,1]$. All that is left to do is to solve for $\lambda(\nu)$.

To do this, simply integrate (172) over μ. After some rearranging, we find

$$\lambda(\nu) = 1 + \frac{c\nu}{2}\fint_{-1}^{1}\frac{D(\mu,\nu)}{\mu-\nu}\,d\mu \tag{173}$$

which is the same as (148) except we now have a principal value integral. Using arguments similar to those leading to (155), we can also write

$$\lambda(\nu) = 1 + \frac{c\nu}{2}\fint_{-1}^{1}\frac{D(\mu,\mu)}{\mu-\nu}\,d\mu \tag{174}$$

which is related to $\Lambda(z)$ through (169). The relation is

$$\Lambda^{\pm}(\nu) = \lambda(\nu) \pm \frac{i\pi c\nu}{2} D(\nu, \nu). \tag{175}$$

Now that we have the eigenvalues and eigenfunctions, we can write the homogeneous solution to (131) as a linear combination

$$I(\tau, \mu) = \sum_{\substack{j=-N \\ j\neq 0}}^{N} a_j \varphi(\mu, \nu_j) e^{\tau/\nu_j} + \int_{-1}^{1} A(\nu)\varphi(\mu, \nu) e^{\tau/\nu}\, d\nu \tag{176}$$

where the a_j are arbitrary constants and $A(\nu)$ is an arbitrary Hölder continuous function on $[-1, 1]$. Here, we denote the discrete eigenvalues by ν_j with $\nu_{-j} = -\nu_j$ and the continuum eigenvalues by $\nu \in [-1, 1]$.

4.4 Orthogonality and Normalization

Let ν' and ν'' be any two eigenvalues. From (144), we can write

$$\left(1 - \frac{\mu}{\nu'}\right)\varphi(\mu, \nu') = \frac{c}{2} D(\mu, \nu'), \tag{177}$$

$$\left(1 - \frac{\mu}{\nu''}\right)\varphi(\mu, \nu'') = \frac{c}{2} D(\mu, \nu''). \tag{178}$$

Now, let us multiply (177) by $\varphi(\mu, \nu'')$, (178) by $\varphi(\mu, \nu')$, subtract, and integrate over μ. We find that

$$\begin{aligned}
&\left(\frac{1}{\nu'} - \frac{1}{\nu''}\right) \int_{-1}^{1} \mu\varphi(\mu, \nu')\varphi(\mu, \nu'')\, d\mu \\
&\qquad = \frac{c}{2} \int_{-1}^{1} \big[D(\mu, \nu'')\varphi(\mu, \nu') - D(\mu, \nu')\varphi(\mu, \nu'')\big]\, d\mu \\
&\qquad = \frac{c}{2} \sum_{m=0}^{M-1} (2m+1)\chi_m^* \int_{-1}^{1} P_m(\mu)\big[R_m(\nu'')\varphi(\mu, \nu') - R_m(\nu')\varphi(\mu, \nu'')\big]\, d\mu \\
&\qquad = \frac{c}{2} \sum_{m=0}^{M-1} (2m+1)\chi_m^* \big[R_m(\nu'')R_m(\nu') - R_m(\nu')R_m(\nu'')\big] \\
&\qquad = 0
\end{aligned} \tag{179}$$

where we have used (139) and (145). Thus, for discrete eigenvalues $\nu_i \neq \nu_j$ and continuum eigenvalues ν, we have

$$\int_{-1}^{1} \mu\varphi(\mu, \nu_i)\varphi(\mu, \nu_j)\, d\mu = 0, \tag{180}$$

$$\int_{-1}^{1} \mu\varphi(\mu, \nu)\varphi(\mu, \nu_j)\, d\mu = 0. \tag{181}$$

This shows that the eigenfunction are orthogonal over $(-1,1)$ with weight function μ.

In future sections, we will come across the expression

$$\psi(\mu) = \sum_{\substack{j=-N \\ j\neq 0}}^{N} a_j \varphi(\mu,\nu_j) + \int_{-1}^{1} A(\nu)\varphi(\mu,\nu)\,d\nu \tag{182}$$

where $\psi(\mu)$ is a known function. This comes from evaluating (176) at some τ, which is what happens when the boundary conditions are applied. Regardless, in order to find the a_j, we multiply (182) by $\mu\varphi(\mu,\nu_i)$ and integrate over μ. We then use (180) and (181) to find that

$$\int_{-1}^{1} \mu\varphi(\mu,\nu_j)\psi(\mu)\,d\mu = a_j \int_{-1}^{1} \mu\varphi^2(\mu,\nu_j)\,d\mu. \tag{183}$$

We use this to define

$$N_j = \int_{-1}^{1} \mu\varphi^2(\mu,\nu_j)\,d\mu. \tag{184}$$

Similarly, to find $A(\nu)$, we multiply (182) by $\mu\varphi(\mu,\nu')$ where $\nu' \in [-1,1]$ and integrate over μ. We again use (181) to find that

$$\int_{-1}^{1} \mu\varphi(\mu,\nu)\psi(\mu)\,d\mu = \int_{-1}^{1} \mu\varphi(\mu,\nu)\bigg(\int_{-1}^{1} A(\nu')\varphi(\mu,\nu')\,d\nu'\bigg)d\mu. \tag{185}$$

We now define

$$N(\nu) = \frac{1}{A(\nu)}\int_{-1}^{1} \mu\varphi(\mu,\nu)\bigg(\int_{-1}^{1} A(\nu')\varphi(\mu,\nu')\,d\nu'\bigg)d\mu. \tag{186}$$

Equations (184) and (186) form what are called the normalization relations. Thus, it is necessary to find N_j and $N(\nu)$. We begin by finding N_j.

Discrete Eigenfunction Normalization

Let us consider the symmetric function

$$H(z,z') = \frac{cz}{2}\frac{cz'}{2}\int_{-1}^{1} \mu\frac{D(\mu,z)}{z-\mu}\frac{D(\mu,z')}{z'-\mu}\,d\mu, \tag{187}$$

which is (184) when $z = z' = \nu_j$. However, we will leave z and z' arbitrary for now and take the appropriate limit to ν_j later. If we perform a partial fraction expansion on $H(z,z')$ and use (145), we find that

$$H(z,z') = \frac{cz}{2}\frac{cz'}{2}\frac{1}{z-z'}\sum_{m=0}^{M-1}(2m+1)\chi_m^*\big[R_m(z)\hat{K}_m(z') - R_m(z')\hat{K}_m(z)\big] \tag{188}$$

where

$$\hat{K}_m(z) = \int_{-1}^{1} \mu P_m(\mu)\frac{D(\mu,z)}{z-\mu}\,d\mu. \tag{189}$$

We can write

$$\hat{K}_m(z) = \int_{-1}^{1} \mu P_m(\mu) \frac{D(\mu, z) - D(\mu, \mu)}{z - \mu} \, d\mu + \int_{-1}^{1} \mu P_m(\mu) \frac{D(\mu, \mu)}{z - \mu} \, d\mu \tag{190}$$

where we recognize that the first integral is a polynomial of degree no greater than $M - 2$ since $D(\mu, z)$ is a polynomial of degree $M - 1$ in both arguments which makes $(D(\mu, z) - D(\mu, \mu))/(z - \mu)$ a polynomial of degree $M - 2$.

The second integral is not a polynomial. We can rewrite it using (175) as

$$\begin{aligned}
\int_{-1}^{1} \mu P_m(\mu) \frac{D(\mu, \mu)}{z - \mu} \, d\mu &= -\frac{1}{i\pi c} \int_{-1}^{1} P_m(\mu) \frac{\Lambda^+(\mu) - \Lambda^-(\mu)}{\mu - z} \, d\mu \\
&= -\frac{1}{i\pi c} \oint_{\gamma_{-1,1}} \frac{P_m(\zeta)\Lambda(\zeta)}{\zeta - z} \, d\zeta \\
&= -\frac{1}{i\pi c} \lim_{\rho \to \infty} \left(\oint_{C_\rho + \gamma_{-1,1}} - \oint_{C_\rho} \right) \frac{P_m(\zeta)\Lambda(\zeta)}{\zeta - z} \, d\zeta \\
&= \frac{1}{i\pi c} \lim_{\rho \to \infty} \oint_{C_\rho} \frac{P_m(\zeta)\Lambda(\zeta)}{\zeta - z} \, d\zeta - \frac{2}{c} P_m(z)\Lambda(z)
\end{aligned} \tag{191}$$

where we have used Cauchy's integral theorem (see Ablowitz and Fokas [1]) to get from the third line to the fourth line. Using the substitution $\zeta = \rho e^{i\theta}$, the geometric series, and a series representation of the Legendre polynomials (see Koornwinder et al. [39]), we can simplify this further to

$$\begin{aligned}
\int_{-1}^{1} \mu P_m(\mu) \frac{D(\mu, \mu)}{z - \mu} \, d\mu &= \frac{\Lambda(\infty)}{\pi c} \lim_{\rho \to \infty} \int_0^{2\pi} P_m(\rho e^{i\theta}) \sum_{\ell=0}^{\infty} \frac{z^\ell}{\rho^\ell e^{i\ell\theta}} \, d\theta - \frac{2}{c} P_m(z)\Lambda(z) \\
&= \frac{\Lambda(\infty)}{\pi c} \lim_{\rho \to \infty} \int_0^{2\pi} 2^m \sum_{k=0}^{m} \binom{m}{k} \binom{\frac{m+k-1}{2}}{m} \rho^k e^{ik\theta} \sum_{\ell=0}^{\infty} \frac{z^\ell}{\rho^\ell e^{i\ell\theta}} \, d\theta \\
&\quad - \frac{2}{c} P_m(z)\Lambda(z) \\
&= \frac{2^{m+1}\Lambda(\infty)}{c} \sum_{k=0}^{m} \binom{m}{k} \binom{\frac{m+k-1}{2}}{m} z^k - \frac{2}{c} P_m(z)\Lambda(z)
\end{aligned} \tag{192}$$

where we have used

$$\int_0^{2\pi} e^{in\theta} \, d\theta = \begin{cases} 2\pi, & n = 0 \\ 0, & \text{otherwise} \end{cases}. \tag{193}$$

Substituting (192) into (190), we find

$$\hat{K}_m(z) = K_m(z) - \frac{2}{c} P_m(z)\Lambda(z) \tag{194}$$

where

$$K_m(z) = \int_{-1}^{1} \mu P_m(\mu) \frac{D(\mu,z) - D(\mu,\mu)}{z-\mu}\, d\mu + \frac{2^{m+1}\Lambda(\infty)}{c} \sum_{k=0}^{m} \binom{m}{k} \binom{\frac{m+k-1}{2}}{m} z^k. \tag{195}$$

Let $z = \nu \in [-1,1]$ and use (194) to write

$$\begin{aligned} K_m(\nu) &= \fint_{-1}^{1} \mu P_m(\mu) \frac{D(\mu,\nu)}{\nu-\mu}\, d\mu + \frac{2}{c} P_m(\nu) \frac{\Lambda^+(\nu) + \Lambda^-(\nu)}{2} \\ &= \frac{2}{c\nu} \fint_{-1}^{1} \mu P_m(\mu) \frac{c\nu}{2} \frac{D(\mu,\nu)}{\nu-\mu}\, d\mu + \frac{2}{c} P_m(\nu)\lambda(\nu) \\ &= \frac{2}{c\nu} \int_{-1}^{1} \mu P_m(\mu) \left(\frac{c\nu}{2} \mathcal{P} \frac{D(\mu,\nu)}{\nu-\mu} + \lambda(\nu)\delta(\nu-\mu) \right) d\mu \\ &= \frac{2}{c\nu} \int_{-1}^{1} \mu P_m(\mu) \varphi(\mu,\nu)\, d\mu. \end{aligned} \tag{196}$$

We can now use the Legendre polynomial recursion formula and (139) to obtain

$$K_m(\nu) = \frac{2}{c}(1 - c\chi_m^*) R_m(\nu). \tag{197}$$

Since (195) shows that $K_m(z)$ is a polynomial of degree no greater than $\max(M-2, m)$ for complex z and (197) shows that $K_m(\nu)$ is a polynomial of degree m for $\nu \in [-1,1]$, we conclude that

$$K_m(z) = \frac{2}{c}(1 - c\chi_m^*) R_m(z) \tag{198}$$

is a polynomial of degree m for all $z \in \mathbb{C}$.

Now if we substitute (194) and (198) into (188), we find after some simplification

$$H(z,z') = \frac{czz'}{2} \frac{D(z,z')\Lambda(z) - D(z',z)\Lambda(z')}{z-z'}. \tag{199}$$

Finally, if we let both z and z' approach a discrete eigenvalues ν_j, we obtain

$$N_j = \frac{c\nu_j^2}{2} D(\nu_j,\nu_j)\Lambda'(\nu_j). \tag{200}$$

Continuum Eigenfunction Normalization

Before we consider normalization of the continuum eigenfunctions, we need to address the issue of permuting the order of integration for double principal value integrals. Let us consider the integral

$$\fint_{-1}^{1} \frac{1}{\nu-\mu} \left(\fint_{-1}^{1} \frac{f(\mu,\nu')}{\nu'-\mu}\, d\nu' \right) d\mu \tag{201}$$

where $\nu \in [-1,1]$ and $f(\mu,\nu')$ is Hölder continuous in both arguments. Usually, to change the order of integration, Fubini's theorem is invoked. In this case, Fubini's theorem does not apply because the integrand is not absolutely integrable. We can, however, switch the order of integration using the Poincaré-Bertrand transposition formula which is normally written (see Jacobs and McInerney [37])

$$\fint_{-1}^{1} \frac{1}{\nu-\mu}\left(\fint_{-1}^{1} \frac{f(\mu,\nu')}{\nu'-\mu}\,d\nu'\right)d\mu = \pi^2 f(\nu,\nu) + \fint_{-1}^{1}\left(\fint_{-1}^{1} \frac{f(\mu,\nu')}{(\nu-\mu)(\nu'-\mu)}\,d\mu\right)d\nu'. \tag{202}$$

This formula is not completely defined, though, because the integral on the right-hand-side does not converge (even in the Cauchy principal value sense) when $\nu' = \nu$. In order to circumvent this difficulty, you can define

$$\mathcal{P}\frac{1}{\nu-\mu}\mathcal{P}\frac{1}{\nu'-\mu} = \frac{1}{\nu'-\nu}\left(\mathcal{P}\frac{1}{\nu-\mu} - \mathcal{P}\frac{1}{\nu'-\mu}\right) \tag{203}$$

which seems quite natural. However, this is merely a convention because the product of two distributions is not well defined. Furthur, it turns out to not be the most convenient convention for our purposes. This was pointed out by Kuščer and McCormick [45]. Instead, we write the Poincaré-Bertrand transposition formula as

$$\fint_{-1}^{1} \frac{1}{\nu-\mu}\left(\fint_{-1}^{1} \frac{f(\mu,\nu')}{\nu'-\mu}\,d\nu'\right)d\mu = \fint_{-1}^{1}\left(\fint_{-1}^{1} \frac{f(\mu,\nu')}{(\nu-\mu)(\nu'-\mu)}\,d\mu\right)d\nu' \tag{204}$$

with the convention

$$\mathcal{P}\frac{1}{\nu-\mu}\mathcal{P}\frac{1}{\nu'-\mu} = \pi^2\delta(\nu-\mu)\delta(\nu'-\mu) + \frac{1}{\nu'-\nu}\left(\mathcal{P}\frac{1}{\nu-\mu} - \mathcal{P}\frac{1}{\nu'-\mu}\right). \tag{205}$$

Notice that if using either (202) and (203) or (204) and (205), we end up with the well-defined formula

$$\begin{aligned}\fint_{-1}^{1} \frac{1}{\nu-\mu}\left(\fint_{-1}^{1} \frac{f(\mu,\nu')}{\nu'-\mu}\,d\nu'\right)d\mu &= \pi^2 f(\nu,\nu) \\ &\quad + \int_{-1}^{1} \frac{1}{\nu'-\nu}\left[\fint_{-1}^{1} f(\mu,\nu')\left(\frac{1}{\nu-\mu} - \frac{1}{\nu'-\mu}\right)d\mu\right]d\nu'.\end{aligned} \tag{206}$$

Using (204) and (205), though, we can formally permute the order of integration in any double Cauchy principal value integral. This will be necessary later on.

Now, the continuum eigenfunction normalization can be found. From the definition of

$N(\nu)$ (186), we find that

$$\begin{aligned}
N(\nu) &= \frac{1}{A(\nu)} \int_{-1}^{1} \mu\varphi(\mu,\nu)\bigg(\int_{-1}^{1} A(\nu')\varphi(\mu,\nu')\,d\nu' \bigg) d\mu \\
&= \frac{1}{A(\nu)} \bigg[\fint_{-1}^{1} \mu \frac{c\nu}{2} \frac{D(\mu,\nu)}{\nu-\mu} \bigg(\fint_{-1}^{1} A(\nu') \frac{c\nu'}{2} \frac{D(\mu,\nu')}{\nu'-\mu}\,d\nu' \bigg) d\mu \\
&\quad + \fint_{-1}^{1} \mu \frac{c\nu}{2} \frac{D(\mu,\nu)}{\nu-\mu} \bigg(\int_{-1}^{1} A(\nu')\lambda(\nu')\delta(\nu'-\mu)\,d\nu' \bigg) d\mu \\
&\quad + \int_{-1}^{1} \mu\lambda(\nu)\delta(\nu-\mu) \bigg(\fint_{-1}^{1} A(\nu') \frac{c\nu'}{2} \frac{D(\mu,\nu')}{\nu'-\mu}\,d\nu' \bigg) d\mu \\
&\quad + \int_{-1}^{1} \mu\lambda(\nu)\delta(\nu-\mu) \bigg(\int_{-1}^{1} \lambda(\nu')\delta(\nu'-\mu)\,d\nu' \bigg) d\mu \bigg]
\end{aligned} \tag{207}$$

where we have simply used the definition of the eigenfunctions. We can now use (204) and (205) to obtain

$$\begin{aligned}
N(\nu) = \frac{1}{A(\nu)} \bigg[& \int_{-1}^{1} \frac{A(\nu')}{\nu'-\nu} \frac{c\nu}{2} \frac{c\nu'}{2} \bigg(\fint_{-1}^{1} \mu D(\mu,\nu) D(\mu,\nu') \bigg(\frac{1}{\nu-\mu} - \frac{1}{\nu'-\mu} \bigg) d\mu \bigg) d\nu' \\
& + \fint_{-1}^{1} A(\nu') \frac{c\nu\nu'}{2} \frac{D(\nu',\nu)}{\nu-\nu'} \lambda(\nu')\,d\nu' + \fint_{-1}^{1} A(\nu') \frac{c\nu\nu'}{2} \frac{D(\nu,\nu')}{\nu'-\nu} \lambda(\nu)\,d\nu' \\
& + A(\nu)\nu\lambda^2(\nu) + A(\nu)\nu \bigg(\frac{\pi c\nu D(\nu,\nu)}{2} \bigg)^2 \bigg]
\end{aligned} \tag{208}$$

For the first integral of (208), we have

$$\begin{aligned}
&\frac{c\nu}{2} \frac{c\nu'}{2} \fint_{-1}^{1} \frac{\mu D(\mu,\nu) D(\mu,\nu')}{\nu-\mu}\,d\mu \\
&\qquad = \frac{c\nu}{2} \frac{c\nu'}{2} \sum_{m=0}^{M-1} (2m+1)\chi_m^* R_m(\nu') \fint_{-1}^{1} \mu P_m(\mu) \frac{D(\mu,\nu)}{\nu-\mu}\,d\mu \\
&\qquad = \frac{c\nu}{2} \frac{c\nu'}{2} \sum_{m=0}^{M-1} (2m+1)\chi_m^* R_m(\nu') \bigg(K_m(\nu) - \frac{2}{c} P_m(\nu)\lambda(\nu) \bigg) \\
&\qquad = \frac{c\nu}{2} \frac{c\nu'}{2} \sum_{m=0}^{M-1} (2m+1)\chi_m^* R_m(\nu') \bigg(\frac{2}{c}(1-c\chi_m^*) R_m(\nu) - \frac{2}{c} P_m(\nu)\lambda(\nu) \bigg) \\
&\qquad = \frac{c\nu\nu'}{2} \big[S(\nu,\nu') - D(\nu,\nu')\lambda(\nu) \big]
\end{aligned} \tag{209}$$

where

$$S(\nu,\nu') = S(\nu',\nu) = \sum_{m=0}^{M-1} (2m+1)(1-c\chi_m^*)\chi_m^* R_m(\nu) R_m(\nu'). \tag{210}$$

With this, we find

$$\begin{aligned}\int_{-1}^{1} \frac{A(\nu')}{\nu'-\nu}\frac{c\nu}{2}\frac{c\nu'}{2}\left(\fint_{-1}^{1} \mu D(\mu,\nu)D(\mu,\nu')\left(\frac{1}{\nu-\mu}-\frac{1}{\nu'-\mu}\right)d\mu\right)d\nu' \\ = -\fint_{-1}^{1} A(\nu')\frac{c\nu\nu'}{2}\frac{D(\nu',\nu)}{\nu-\nu'}\lambda(\nu')\,d\nu' - \fint_{-1}^{1} A(\nu')\frac{c\nu\nu'}{2}\frac{D(\nu,\nu')}{\nu'-\nu}\lambda(\nu)\,d\nu'\end{aligned} \tag{211}$$

which shows that the first term of (208) cancels with the second and third terms. Thus, the normalization for the continuum eigenfunctions is

$$N(\nu) = \nu\Lambda^+(\nu)\Lambda^-(\nu) \tag{212}$$

since (175) gives

$$\Lambda^+(\nu)\Lambda^-(\nu) = \lambda^2(\nu) + \left(\frac{\pi c\nu D(\nu,\nu)}{2}\right)^2. \tag{213}$$

4.5 Eigenfunction Completeness

The homogeneous solution to (131) is given by (176). If we are also able to find a particular solution, then all that remains is to find arbitrary constants a_j and function $A(\nu)$ in (176). This, of course, is done by applying the boundary conditions. Using the orthogonality and normalization conditions found in the previous section, this does not present any difficulties. That is, suppose we have

$$\sum_{\substack{j=-N\\ j\neq 0}}^{N} a_j\varphi(\mu,\nu_j) + \int_{-1}^{1} A(\nu)\varphi(\mu,\nu)\,d\nu = \psi(\mu). \tag{214}$$

To find the a_j, we would multiply (214) by $\mu\varphi(\mu,\nu_i)$ and apply (180), (181), and (200) to find

$$a_j = \frac{1}{N_j}\int_{-1}^{1} \mu\varphi(\mu,\nu_j)\psi(\mu)\,d\mu. \tag{215}$$

Then we can multiply (214) by $\mu\varphi(\mu,\nu')$ for $\nu' \in [-1,1]$ and apply (180), (181), and (212) to find

$$A(\nu) = \frac{1}{N(\nu)}\int_{-1}^{1} \mu\varphi(\mu,\nu)\psi(\mu)\,d\mu. \tag{216}$$

In order for this procedure to be valid, however, the eigenfunctions have to be complete over some function space.

The first to show the completeness of the eigenfunctions was Mika [68], where it was shown that this is equivalent to solving a certain singular integral equation. The problem was solved by following a general prescription from Muskhelishvili [72] (also found in Polyanin and Manzhirov [76]). Thus, the eigenfunctions were shown to be complete, but it was not shown how to relate the solution of the singular integral equation to the solution found

through orthogonality (215) and (216). This section aims to show that these solutions are the same by taking advantage of the structure of the integral equation.

Suppose $\psi(\mu)$ is Hölder continuous on $[-1,1]$. We will show that there exist constants a_j and a function $A(\nu)$ such that

$$\sum_{\substack{j=-N \\ j\neq 0}}^{N} a_j \varphi(\mu,\nu_j) + \int_{-1}^{1} A(\nu)\varphi(\mu,\nu)\,d\nu = \psi(\mu). \tag{217}$$

For now, let us assume that

$$\tilde{\psi}(\mu) = \psi(\mu) - \sum_{\substack{j=-N \\ j\neq 0}}^{N} a_j \varphi(\mu,\nu_j) \tag{218}$$

is a known function and write

$$\fint_{-1}^{1} A(\nu)\frac{c\nu}{2}\frac{D(\mu,\nu)}{\nu-\mu}\,d\nu + A(\mu)\lambda(\mu) = \tilde{\psi}(\mu) \tag{219}$$

where we have used (172). This is a singular integral equation of Cauchy type.

One way to solve this singular integral equation is to introduce a function

$$\Phi(z,\zeta) = \frac{D(z,\zeta)}{2\pi i D(z,z)}\int_{-1}^{1} A(\nu)\frac{c\nu}{2}\frac{D(z,\nu)}{\nu-z}\,d\nu \tag{220}$$

where ζ is a parameter that will be determined later. From Theorem 3, we have

$$\Phi^+(\mu,\zeta) + \Phi^-(\mu,\zeta) = \frac{D(\mu,\zeta)}{i\pi D(\mu,\mu)}\fint_{-1}^{1} A(\nu)\frac{c\nu}{2}\frac{D(\mu,\nu)}{\nu-\mu}\,d\nu, \tag{221}$$

$$\Phi^+(\mu,\zeta) - \Phi^-(\mu,\zeta) = A(\mu)\frac{c\mu D(\mu,\zeta)}{2} \tag{222}$$

both of which can be substituted into (219). Doing so gives

$$\frac{i\pi D(\mu,\mu)}{D(\mu,\zeta)}\big[\Phi^+(\mu,\zeta)+\Phi^-(\mu,\zeta)\big] + \frac{2\lambda(\mu)}{c\mu D(\mu,\zeta)}\big[\Phi^+(\mu,\zeta)-\Phi^-(\mu,\zeta)\big] = \tilde{\psi}(\mu). \tag{223}$$

Using (175), this simplifies to

$$\Lambda^+(\mu)\Phi^+(\mu,\zeta) - \Lambda^-(\mu)\Phi^-(\mu,\zeta) = \frac{c\mu}{2}D(\mu,\zeta)\tilde{\psi}(\mu). \tag{224}$$

Let us define a function $\Psi(z,\zeta) = \Lambda(z)\Phi(z,\zeta)$. This gives us

$$\Psi^+(\mu,\zeta) - \Psi^-(\mu,\zeta) = \frac{c\mu}{2}D(\mu,\zeta)\tilde{\psi}(\mu) \tag{225}$$

which immediately from Theorem 3 gives

$$\Psi(z,\zeta) = \frac{1}{2\pi i}\int_{-1}^{1} \frac{c\mu}{2}\frac{D(\mu,\zeta)}{\mu - z}\tilde{\psi}(\mu)\,d\mu \tag{226}$$

or equivalently

$$\Phi(z,\zeta) = \frac{1}{\Lambda(z)}\frac{1}{2\pi i}\int_{-1}^{1} \frac{c\mu}{2}\frac{D(\mu,\zeta)}{\mu - z}\tilde{\psi}(\mu)\,d\mu. \tag{227}$$

Upon inspection of the integrand, we see that it resembles one of the eigenfunctions if we let $\zeta = z$. Define $\Phi(z) = \Phi(z,z)$ so that

$$\Phi(z) = \frac{1}{\Lambda(z)}\frac{1}{2\pi i}\int_{-1}^{1} \frac{c\mu}{2}\frac{D(\mu,z)}{\mu - z}\tilde{\psi}(\mu)\,d\mu. \tag{228}$$

We can make the same substitution in (220) to find

$$\Phi(z) = \frac{1}{2\pi i}\int_{-1}^{1} A(\nu)\frac{c\nu}{2}\frac{D(z,\nu)}{\nu - z}\,d\nu. \tag{229}$$

Clearly, (229) shows that $\Phi(z)$ is holomorphic in the complex plane cut on $[-1,1]$, but (228) is not because $\Lambda(z)$ has zeros at $z = \nu_j$ for $j = \pm1, \pm2, \ldots, \pm N$. Hence, (228) is not true for general $\tilde{\psi}(\mu)$, but from (218) $\tilde{\psi}(\mu)$ has a special form that allows (228). Let $z = \nu_j$ for some j and substitute (218) into (228) to find

$$\begin{aligned}
\Phi(\nu_j) &= \frac{1}{\Lambda(\nu_j)}\frac{1}{2\pi i}\int_{-1}^{1} \frac{c\mu}{2}\frac{D(\mu,\nu_j)}{\mu - \nu_j}\left(\psi(\mu) - \sum_{\substack{i=-N\\ i\neq 0}}^{N} a_i\varphi(\mu,\nu_i)\right)d\mu \\
&= -\frac{1}{\Lambda(\nu_j)}\frac{1}{2\pi i\nu_j}\int_{-1}^{1} \mu\frac{c\nu_j}{2}\frac{D(\mu,\nu_j)}{\nu_j - \mu}\left(\psi(\mu) - \sum_{\substack{i=-N\\ i\neq 0}}^{N} a_i\varphi(\mu,\nu_i)\right)d\mu \\
&= -\frac{1}{\Lambda(\nu_j)}\frac{1}{2\pi i\nu_j}\int_{-1}^{1} \mu\varphi(\mu,\nu_j)\left(\psi(\mu) - \sum_{\substack{i=-N\\ i\neq 0}}^{N} a_i\varphi(\mu,\nu_i)\right)d\mu.
\end{aligned} \tag{230}$$

In order to make $\Phi(z)$ holomorphic in the cut plane, the integral in (230) must vanish. That is, we must have

$$\sum_{\substack{i=-N\\ i\neq 0}}^{N} a_i\int_{-1}^{1} \mu\varphi(\mu,\nu_j)\varphi(\mu,\nu_i)\,d\mu = \int_{-1}^{1} \mu\varphi(\mu,\nu_j)\psi(\mu)\,d\mu. \tag{231}$$

From (180) and (200), this simplifies to

$$a_j = \frac{1}{N_j}\int_{-1}^{1} \mu\varphi(\mu,\nu_j)\psi(\mu)\,d\mu \tag{232}$$

which is precisely what we obtained before in (215).

Now all that is left is to find $A(\nu)$. From (222), we see that

$$A(\nu) = \frac{2}{c\nu D(\nu,\nu)}\big[\Phi^+(\nu) - \Phi^-(\nu)\big] \tag{233}$$

where we have set $\zeta = z = \nu$. Then, applying Theorem 3 to (228), we find

$$\begin{aligned}\Phi^{\pm}(\nu) &= \frac{1}{\Lambda^{\pm}(\nu)}\left(\frac{1}{2\pi i}\int_{-1}^{1}\frac{c\mu}{2}\frac{D(\mu,\nu)}{\mu-\nu}\tilde{\psi}(\mu)\,d\mu \pm \frac{1}{2}\frac{c\nu}{2}D(\nu,\nu)\tilde{\psi}(\nu)\right)\\ &= \frac{1}{\Lambda^{\pm}(\nu)}\left(-\frac{1}{2\pi i\nu}\int_{-1}^{1}\mu\frac{c\nu}{2}\frac{D(\mu,\nu)}{\nu-\mu}\tilde{\psi}(\mu)\,d\mu\right.\\ &\quad\left.\pm\frac{1}{2\nu}\,\frac{c\nu}{2}D(\nu,\nu)\int_{-1}^{1}\mu\delta(\nu-\mu)\tilde{\psi}(\mu)\,d\mu\right).\end{aligned} \tag{234}$$

It is now straightforward to use this with (175) and (212) to find

$$\Phi^+(\nu) - \Phi^-(\nu) = \frac{1}{N(\nu)}\frac{c\nu}{2}D(\nu,\nu)\int_{-1}^{1}\mu\varphi(\mu,\nu)\tilde{\psi}(\mu)\,d\mu. \tag{235}$$

By orthogonality and (233), we can replace $\tilde{\psi}(\mu)$ with $\psi(\mu)$. Finally, we arrive at

$$A(\nu) = \frac{1}{N(\nu)}\int_{-1}^{1}\mu\varphi(\mu,\nu)\psi(\mu)\,d\mu\text{`} \tag{236}$$

which is precisely the same as (216). Thus, the eigenfunctions are complete in the space of Hölder continuous functions on $[-1,1]$, and the arbitrary coefficients are given uniquely by (215) and (216). From now on, instead of solving a singular integral equation such as (219), we can simply apply the orthogonality and normalization relationships from the previous section.

4.6 A Fundamental Solution

We are now ready to find a fundamental solution to (131). A fundamental solution is any distribution $F(\tau,\mu;\tau_0,\mu_0)$ such that

$$\begin{aligned}&\mu\frac{\partial F(\tau,\mu;\tau_0,\mu_0)}{\partial\tau} - F(\tau,\mu;\tau_0,\mu_0)\\ &\quad+\frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^*P_m(\mu)\int_{-1}^{1}P_m(\mu')F(\tau,\mu';\tau_0,\mu_0)\,d\mu' = \delta(\tau-\tau_0)\delta(\mu-\mu_0).\end{aligned} \tag{237}$$

A fundamental solution is not unique because no domain is specified. We have the freedom to choose any domain and boundary conditions. We will specify this momentarily. For now,

we can use the results from the previous sections to write

$$F(\tau,\mu;\tau_0,\mu_0)=\begin{cases}\displaystyle\sum_{\substack{j=-N\\ j\neq 0}}^{N} b_j\varphi(\mu,\nu_j)e^{\tau/\nu_j}+\int_{-1}^{1}B(\nu)\varphi(\mu,\nu)e^{\tau/\nu}\,d\nu, & \tau>\tau_0\\ \displaystyle\sum_{\substack{j=-N\\ j\neq 0}}^{N} c_j\varphi(\mu,\nu_j)e^{\tau/\nu_j}+\int_{-1}^{1}C(\nu)\varphi(\mu,\nu)e^{\tau/\nu}\,d\nu, & \tau<\tau_0\end{cases} \tag{238}$$

If we were trying to find the Green's functions $G(\tau,\mu;\tau_0,\mu_0)$, then the boundary conditions would be (see Duffy [23])

$$G(0,\mu;\tau_0,\mu_0)=0,\quad \mu<0, \tag{239}$$

$$G(\tau_{\max},\mu;\tau_0,\mu_0)=0,\quad \mu>0. \tag{240}$$

It is possible to find a closed form for the Green's function, but it is a very convoluted process. In addition to the relationships we found in the previous sections, we have to find orthogonality relationships (actually biorthogonality relationships) over the half interval $[-1,0]$. Doing so involves solving a more complicated singular integral equation than what we had above. The general procedure can be found in Case and Zweifel [17] for $M=1$ and McCormick and Kuščer [65] for general M. However, this is unnecessary because we can use a fundamental solution along with the boundary element method to solve (131).

Since we have freedom in choosing the boundary conditions, we will choose

$$\lim_{\tau\to+\infty}F(\tau,\mu;\tau_0,\mu_0)=0,\qquad \mu>0, \tag{241}$$

$$\lim_{\tau\to-\infty}F(\tau,\mu;\tau_0,\mu_0)=0,\qquad \mu<0. \tag{242}$$

This gives us

$$F(\tau,\mu;\tau_0,\mu_0)=\begin{cases}\displaystyle\sum_{j=-N}^{-1} a_j\varphi(\mu,\nu_j)e^{\tau/\nu_j}+\int_{-1}^{0}A(\nu)\varphi(\mu,\nu)e^{\tau/\nu}\,d\nu, & \tau>\tau_0\\ \displaystyle-\sum_{j=1}^{N} a_j\varphi(\mu,\nu_j)e^{\tau/\nu_j}-\int_{0}^{1}A(\nu)\varphi(\mu,\nu)e^{\tau/\nu}\,d\nu, & \tau<\tau_0\end{cases} \tag{243}$$

where $a_j=b_j$ for $j<0$, $a_j=-c_j$ for $j>0$, $A(\nu)=B(\nu)$ for $\nu<0$, and $A(\nu)=-C(\nu)$ for $\nu>0$. To find the remaining coefficients, we need another condition. This is found by integrating (237) over τ from $\tau_0-\varepsilon$ to $\tau_0+\varepsilon$ where $\varepsilon>0$ and taking the limit as $\varepsilon\to 0$. Doing so gives the jump condition

$$F(\tau_0^+,\mu;\tau_0,\mu_0)-F(\tau_0^-,\mu;\tau_0,\mu_0)=\frac{\delta(\mu-\mu_0)}{\mu} \tag{244}$$

from which we immediately find that

$$\sum_{\substack{j=-N\\ j\neq 0}}^{N} a_j\varphi(\mu,\nu_j)e^{\tau_0/\nu_j}+\int_{-1}^{1}A(\nu)\varphi(\mu,\nu)e^{\tau_0/\nu}\,d\nu=\frac{\delta(\mu-\mu_0)}{\mu}. \tag{245}$$

Using the orthogonality relations (180) and (181) is straightforward and gives us

$$a_j = \frac{1}{N_j}\varphi(\mu_0, \nu_j)e^{-\tau_0/\nu_j}, \tag{246}$$

$$A(\nu) = \frac{1}{N(\nu)}\varphi(\mu_0, \nu)e^{-\tau_0/\nu}. \tag{247}$$

There is a problem with using orthogonality to find these coefficients. In Earlier, we required that $\psi(\mu)$ and $A(\nu)$ be Hölder continuous. Here, we have $\psi(\mu) = \delta(\mu - \mu_0)/\mu$ and $A(\nu)$ is given by (247). Distributions such as the Dirac delta and Cauchy principal values are not Hölder continuous. However, this can be mitigated by noting that

$$\delta(x) = \frac{1}{2\pi i}\lim_{\varepsilon\to 0^+}\left(\frac{1}{x - i\varepsilon} - \frac{1}{x + i\varepsilon}\right), \tag{248}$$

$$\mathcal{P}\frac{1}{x} = \frac{1}{2}\lim_{\varepsilon\to 0^+}\left(\frac{1}{x - i\varepsilon} - \frac{1}{x + i\varepsilon}\right). \tag{249}$$

We can simply replace all delta functions and Cauchy principal values with these expressions and take the limit only after (246) and (247) have been found.

We also point out that in order to find (247), it was necessary to use (204) and (205) because in the process of finding $N(\nu)$, we assumed that we could formally switch the order of integration. Using (202) and (203) would require us to include other terms that would end up complicating our expression for $A(\nu)$. This was pointed out in McInerney [66] and is why we use (204) and (205). Substituting (246) and (247) into (243), we obtain

$$F(\tau, \mu; \tau_0, \mu_0) = \begin{cases} \displaystyle\sum_{j=-N}^{-1} \frac{1}{N_j}\varphi(\mu, \nu_j)\varphi(\mu_0, \nu_j)e^{-(\tau_0-\tau)/\nu_j} \\ \qquad \displaystyle + \int_{-1}^{0} \frac{1}{N(\nu)}\varphi(\mu, \nu)\varphi(\mu_0, \nu)e^{-(\tau_0-\tau)/\nu}\, d\nu, \quad \tau > \tau_0 \\ \displaystyle -\sum_{j=1}^{N} \frac{1}{N_j}\varphi(\mu, \nu_j)\varphi(\mu_0, \nu_j)e^{-(\tau_0-\tau)/\nu_j} \\ \qquad \displaystyle - \int_{0}^{1} \frac{1}{N(\nu)}\varphi(\mu, \nu)\varphi(\mu_0, \nu)e^{-(\tau_0-\tau)/\nu}\, d\nu, \quad \tau < \tau_0 \end{cases} \tag{250}$$

Something that will serve us later is the so-called reciprocity relation. To find this, we first need the adjoint differential equation. If we define an inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty}\int_{-1}^{1} f(\tau, \mu)g(\tau, \mu)\, d\mu\, d\tau, \tag{251}$$

then the adjoint is defined with

$$\langle F^*, \mathcal{L}F\rangle = \langle \mathcal{L}^*F^*, F\rangle \tag{252}$$

where $\mathcal{L}$ is the operator such that

$$\begin{aligned}\mathcal{L}F =& \mu\frac{\partial F(\tau,\mu;\tau_0,\mu_0)}{\partial\tau} - F(\tau,\mu;\tau_0,\mu_0)\\ &+\frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu)\int_{-1}^{1}P_m(\mu')F(\tau,\mu';\tau_0,\mu_0)\,d\mu'\end{aligned} \tag{253}$$

and $\mathcal{L}^*$ is the adjoint operator. Using the inner product definition, we find that

$$\begin{aligned}\mathcal{L}^*F^* =& -\mu\frac{\partial F^*(\tau,\mu;\tau_1,\mu_1)}{\partial\tau} - F^*(\tau,\mu;\tau_1,\mu_1)\\ &+\frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu)\int_{-1}^{1}P_m(\mu')F^*(\tau,\mu';\tau_1,\mu_1)\,d\mu'\end{aligned} \tag{254}$$

with adjoint boundary conditions

$$\lim_{\tau\to+\infty} F^*(\tau,\mu;\tau_1,\mu_1) = 0, \qquad \mu<0, \tag{255}$$

$$\lim_{\tau\to-\infty} F^*(\tau,\mu;\tau_1,\mu_1) = 0, \qquad \mu>0. \tag{256}$$

We have changed the subscripts on the source variables τ_1 and μ_1 because they are not in general the same as τ_0 and μ_0. For our problem, we have

$$\mathcal{L}F = \delta(\tau-\tau_0)\delta(\mu-\mu_0). \tag{257}$$

Suppose we also have

$$\mathcal{L}^*F^* = \delta(\tau-\tau_1)\delta(\mu-\mu_1). \tag{258}$$

Then (251) and (252) immediately yield $F(\tau_1,\mu_1;\tau_0,\mu_0) = F^*(\tau_0,\mu_0;\tau_1,\mu_1)$. This is the reciprocity relation. We can now see that a fundamental solution solves the adjoint equation in the source variables τ_0 and μ_0. That is,

$$\begin{aligned}&-\mu_0\frac{\partial F(\tau,\mu;\tau_0,\mu_0)}{\partial\tau_0} - F(\tau,\mu;\tau_0,\mu_0)\\ &\quad+\frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu_0)\int_{-1}^{1}P_m(\mu_0')F(\tau,\mu';\tau_0,\mu_0)\,d\mu_0' = \delta(\tau-\tau_0)\delta(\mu-\mu_0),\end{aligned} \tag{259}$$

which will be useful to us in the next section.

4.7 The Boundary Element Method

To derive equations for the boundary element method, we start by rewriting (131) as

$$\begin{aligned}&\mu_0\frac{\partial I(\tau_0,\mu_0)}{\partial\tau_0} - I(\tau_0,\mu_0)\\ &\quad+\frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu_0)\int_{-1}^{1}P_m(\mu_0')I(\tau_0,\mu_0')\,d\mu_0' = q(\tau_0,\mu_0).\end{aligned} \tag{260}$$

Then, we multiply this equation by $F(\tau,\mu;\tau_0,\mu_0)$ and integrate over τ_0 from 0 to $\tau_{\max}$ and over μ_0 from -1 to 1. Doing so, we find

$$\begin{aligned}
&\int_0^{\tau_{\max}}\int_{-1}^{1} F(\tau,\mu;\tau_0,\mu_0)q(\tau_0,\mu_0)\,d\mu_0\,d\tau_0 \\
&\qquad= \int_0^{\tau_{\max}}\int_{-1}^{1} F(\tau,\mu;\tau_0,\mu_0)\bigg(\mu_0\frac{\partial I(\tau_0,\mu_0)}{\partial\tau_0} - I(\tau_0,\mu_0) \\
&\qquad\quad + \frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu_0)\int_{-1}^{1} P_m(\mu_0')I(\tau_0,\mu_0')\,d\mu_0'\bigg)d\mu_0\,d\tau_0 \\
&\qquad= \int_{-1}^{1}\mu_0 F(\tau,\mu;\tau_0,\mu_0)I(\tau_0,\mu_0)\Big|_{\tau_0=0}^{\tau_{\max}}d\mu_0 \\
&\qquad\quad + \int_0^{\tau_{\max}}\int_{-1}^{1} I(\tau_0,\mu_0)\bigg(-\mu_0\frac{\partial F(\tau,\mu;\tau_0,\mu_0)}{\partial\tau_0} - F(\tau,\mu;\tau_0,\mu_0) \\
&\qquad\quad + \frac{c}{2}\sum_{m=0}^{M-1}(2m+1)\chi_m^* P_m(\mu_0)\int_{-1}^{1} P_m(\mu_0')F(\tau,\mu;\tau_0,\mu_0')\,d\mu_0'\bigg)d\mu_0\,d\tau_0 \\
&\qquad= \int_{-1}^{1}\mu_0 F(\tau,\mu;\tau_0,\mu_0)I(\tau_0,\mu_0)\Big|_{\tau_0=0}^{\tau_{\max}}d\mu_0 + I(\tau,\mu)
\end{aligned} \tag{261}$$

where the final equality follows from (259). Upon applying the boundary conditions (134) and (135), we find that

$$\begin{aligned}
I(\tau,\mu) = &\int_0^{\tau_{\max}}\int_{-1}^{1} F(\tau,\mu;\tau_0,\mu_0)q(\tau_0,\mu_0)\,d\mu_0\,d\tau_0 \\
&+ \int_{-1}^{0}\mu_0\Big[F(\tau,\mu;0,\mu_0) \quad -F(\tau,\mu;\tau_{\max},\mu_0)\Big]\begin{bmatrix} I_{\text{top}}(\mu_0) \\ I(\tau_{\max},\mu_0)\end{bmatrix}d\mu_0 \\
&+ \int_{0}^{1}\mu_0\Big[F(\tau,\mu;0,\mu_0) \quad -F(\tau,\mu;\tau_{\max},\mu_0)\Big]\begin{bmatrix} I(0,\mu_0) \\ 0\end{bmatrix}d\mu_0
\end{aligned} \tag{262}$$

where the boundary term has been split into two integrals and those integrands have been written in matrix form.

If we had used the Green's function instead of a fundamental solution, then every term on the right-hand-side would be known because from (239), (240), and the reciprocity relation we would have $G(\tau,\mu;\tau_{\max},\mu_0) = 0$ for $\mu_0 < 0$ and $G(\tau,\mu;0,\mu_0) = 0$ for $\mu_0 > 0$. However, since we only found a fundamental solution, we must also solve for $I(\tau_{\max},\mu)$ for $\mu < 0$ and for $I(0,\mu)$ for $\mu > 0$. To do this, we can evaluate (262) at $\tau = 0+\varepsilon$ and $\tau = \tau_{\max}-\varepsilon$ for $\varepsilon > 0$ and let $\varepsilon \to 0$. Doing so, we find

$$\begin{aligned}
\begin{bmatrix} I(0,\mu) \\ I(\tau_{\max},\mu)\end{bmatrix} = &\int_0^{\tau_{\max}}\int_{-1}^{1}\begin{bmatrix} F(0^+,\mu;\tau_0,\mu_0) \\ F(\tau_{\max}^-,\mu;\tau_0,\mu_0)\end{bmatrix}q(\tau_0,\mu_0)\,d\mu_0\,d\tau_0 \\
&+ \int_{-1}^{0}\mu_0\begin{bmatrix} F(0^+,\mu;0,\mu_0) & -F(0^+,\mu;\tau_{\max},\mu_0) \\ F(\tau_{\max}^-,\mu;0,\mu_0) & -F(\tau_{\max}^-,\mu;\tau_{\max},\mu_0)\end{bmatrix}\begin{bmatrix} I_{\text{top}}(\mu_0) \\ I(\tau_{\max},\mu_0)\end{bmatrix}d\mu_0 \\
&+ \int_{0}^{1}\mu_0\begin{bmatrix} F(0^+,\mu;0,\mu_0) & -F(0^+,\mu;\tau_{\max},\mu_0) \\ F(\tau_{\max}^-,\mu;0,\mu_0) & -F(\tau_{\max}^-,\mu;\tau_{\max},\mu_0)\end{bmatrix}\begin{bmatrix} I(0,\mu_0) \\ 0\end{bmatrix}d\mu_0.
\end{aligned} \tag{263}$$

These are two coupled linear Fredholm integral equations of the second kind. Since $\tau_{\max} \approx 10^7$ is large, it follows from (250) that $F(0, \mu; \tau_{\max}, \mu_0) \approx 0$ and $F(\tau_{\max}, \mu; 0, \mu_0) \approx 0$. This approximation is unnecessary, but it is a very good approximation (in fact, it is exact in double precision arithmetic) and decouples our integral equations. We now have

$$\begin{aligned} I(0,\mu) = &\int_0^{\tau_{\max}} \int_{-1}^{1} F(0^+, \mu; \tau_0, \mu_0) q(\tau_0, \mu_0)\, d\mu_0\, d\tau_0 \\ &+ \int_{-1}^{0} \mu_0 F(0^+, \mu; 0, \mu_0) I_{\text{top}}(\mu_0)\, d\mu_0 \\ &- \int_{0}^{1} \mu_0 F(0^+, \mu; 0, \mu_0) I(0, \mu_0)\, d\mu_0, \qquad \mu > 0 \end{aligned} \tag{264}$$

and

$$\begin{aligned} I(\tau_{\max},\mu) = &\int_0^{\tau_{\max}} \int_{-1}^{1} F(\tau_{\max}^-, \mu; \tau_0, \mu_0) q(\tau_0, \mu_0)\, d\mu_0\, d\tau_0 \\ &- \int_{-1}^{0} \mu_0 F(\tau_{\max}^-, \mu; \tau_{\max}, \mu_0) I(\tau_{\max}, \mu)\, d\mu_0, \qquad \mu < 0. \end{aligned} \tag{265}$$

Numerically solving a linear Fredholm integral equation of the second kind is a straightforward procedure. There are many methods, but we suggest the Nyström method due to its simplicity. In fact, Delves and Mohamed [20] opine that the method is so simple that it is "enough to make a numerical analyst weep". Suppose we have an integral equation

$$f(x) = g(x) + \lambda \int_a^b K(x,y) f(y)\, dy, \qquad a \leqslant x \leqslant b. \tag{266}$$

Here, $g(x)$ and $K(x,y)$ are known functions and λ is a known constant. Replace the integral with an appropriate quadrature rule to obtain

$$f(x) = g(x) + \lambda \sum_{j=1}^{N} w_j K(x, x_j) f(x_j), \qquad a \leqslant x \leqslant b \tag{267}$$

where x_j are the quadrature nodes and w_j are the quadrature weights. Now, evaluate (267) at the quadrature nodes x_i for $i = 1, 2, \ldots, N$ to obtain the linear system of equations

$$f(x_i) = g(x_i) + \lambda \sum_{j=1}^{N} w_j K(x_i, x_j) f(x_j), \qquad i = 1, 2, \ldots N. \tag{268}$$

Solving this system poses no difficulties for a well-behaved kernel function $K(x,y)$. The only requirement is that λ is not a characteristic value. After solving for $f(x)$ at the quadrature nodes, we can evaluate $f(x)$ at any point $a \leqslant x \leqslant b$ using (267). This can be considered to be a "smart" interpolation because it preserves the order of accuracy of the quadrature rule. A different interpolation scheme, such as linear interpolation, would not.

Thus, we can use the Nyström method to solve (264) and (265). It is then a straightforward matter to evaluate (262). Since (262) is an exact solution, the only approximation is from the numerical solution of (264) and (265) and from numerically evaluating the integrals in (262). We point out that some of the integrals can have "near-singularities" at the endpoints of integration. A near-singularity arises when an integrand has a singularity just outside the region of integration, creating a boundary layer-like gradient. See Monegato and Scuderi [70] to see how to mitigate this issue. Also, the singularity in Cauchy principal value integrals can be evaluated numerically with the relation

$$\begin{aligned} \fint_{-1}^{1} \frac{f(\mu)}{\mu-\nu}\,d\mu &= \int_{-1}^{1} \frac{f(\mu)-f(\nu)}{\mu-\nu}\,d\mu + f(\nu)\fint_{-1}^{1} \frac{1}{\mu-\nu}\,d\mu \\ &= \int_{-1}^{1} \frac{f(\mu)-f(\nu)}{\mu-\nu}\,d\mu + f(\nu)\log\left(\frac{1-\nu}{1+\nu}\right). \end{aligned} \tag{269}$$

5 Auroral Light Emissions

5.1 Light Production of the Aurora

By itself, the electron intensity $I(z, E, \mu)$ does not give much useful information about the aurora. However, it is the quantity that we solve for because it is a measurable quantity, and it can be used to calculate excitation and ionization rates as a function of altitude. These rates are given by

$$r_\xi^\eta(z) = 2\pi n_\xi(z) \int_{T_\xi^\eta}^{\infty} \int_{-1}^{1} \sigma_\xi^\eta(E) I(z, E, \mu)\, d\mu\, dE, \qquad (\mathrm{cm}^{-3}\ \mathrm{s}^{-1}) \tag{270}$$

where $n_\xi(z)$ and $\sigma_\xi^\eta(E)$ are the number density and cross section for species ξ and channel η. These rates are used as inputs to a kinetic model for light emissions that we will describe in the next section. The light emissions are important because no experiment can be designed without first understanding the light emission process.

As described earlier, the excited states of the atmospheric species all decay back to their ground states in some way, and this decay is what produces the auroral light by the Planck-Einstein relation (6). In theory, we could calculate the spectra of all light emissions, but we will only concern ourselves with visible light emissions. This greatly reduces the complexity of the model because most of the excited states do not contribute visible light. Further, much of the visible light is very faint, which allows us to focus on a subset of the visible light.

We will focus on three particular wavelengths (or lines as they are called): 6300.3 Å, 5577.3 Å, and 4278.1 Å. These wavelengths are the strongest red, green, and blue lines, respectively (see Rees [78]). The relative amount of these lines varies with altitude. Figure 7 (reproduced from Baranoski et al. [9]) shows the brightness profile for a typical aurora. We see that the green line is the brightest followed by the blue line. Both these lines peak at about the same altitude, which is much lower than the red line. The red line peaks at higher altitudes.

The red line at 6300.3 Å results from the oxygen state O(^{1}D) decaying to the ground state O(^{3}P$_2$). The symbols inside the parentheses are called term symbols. For our purposes, it is not important to understand their meaning, but it is only important to know that O(^{1}D) results in the 6300.3 Å red line. Similarly, the green line at 5577.3 Å results from the oxygen state O(^{1}S) decaying to the excited state O(^{1}D). This means that for every green emission the number density of O(^{1}D) increases, which yields more red emissions.

To make this more concrete, it is helpful to have an energy level diagram. Partial energy level diagrams for the lower states of neutral and singly ionized oxygen and nitrogen are shown in Figures 8–11. Here, the black lines represent the energy threshold T_ξ^η for the state written next to them, and the higher the black line, the larger the excitation threshold. The blue arrows show how these states decay, and the numbers along the arrows are the

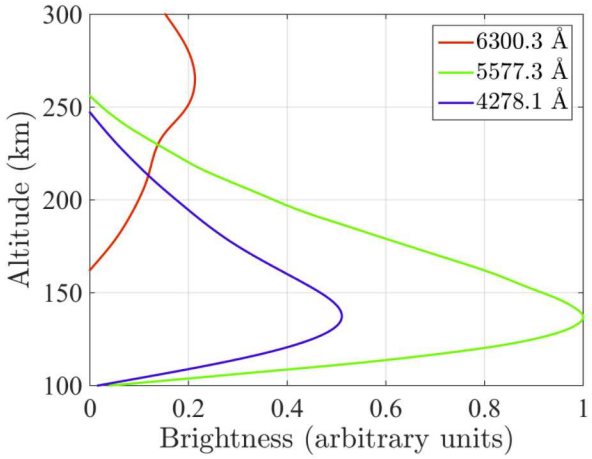

Figure 7. Brightness profile for a typical aurora (see Baranoski et al. [9]).

wavelengths of the transitions in angstroms. Since more than one transition can occur from a given state, it is important to know the branching ratios for each transition. For instance, the O(^{1}D) state in Figure 8 has three arrows showing wavelengths of 6300.3, 6363.8, and 6391.7 Å. This is due to the so-called fine structure of O(^{3}P). That is, there are three states clustered around the ground state. However, about 75% of the radiative decay is through the 6300.3 Å transition and about 25% through the 6363.8 Å transition. For this reason, these two lines are called the red doublet of oxygen (see Rees [78]). Only a very small fraction of 1% is through the 6391.7 Å transition. These branching ratios can be calculated from what are called the Einstein coefficients, which can be found in Kramida et al. [40]. For instance, the Einstein coefficient for the 6300.3 Å line is 5.6511×10^{-3} s^{-1} and 1.82339×10^{-3} s^{-1} for the 6363.8 Å line. The Einstein coefficient for the 6391.7 Å line, by contrast, is only 8.6×10^{-7} s^{-1}. The branching ratio is found by dividing one of these coefficients by their sum. Similarly, about 95% of the radiative decay from O(^{1}S) is through the 5577.3 Å transition. The other two wavelengths are not visible, and we will not be concerned with them.

Figures 9–11 are included here for a couple of reasons. The first is to show that much of the auroral spectrum is not visible. For instance, some of the wavelengths in Figure 9 are non-visible. The visible spectrum is roughly 3800 Å $\leqslant \lambda \leqslant$ 7500 Å. There are authors who are interested in parts of the non-visible spectrum (see Strickland et al. [99] for example), and the procedure described below could be used to calculate the non-visible emissions. However, we will confine ourselves to the three wavelengths stated above.

The other reason for including Figures 9–11 is to point out some visible wavelengths for which there is no need to calculate the emissions. The reason for this is that a relatively small amount of light is produced. For example, from Figure 10 we see two visible wavelengths at 5197.9 and 5200.3 Å. Both of these lines have very long lifetimes (about 13.66 and 36.74

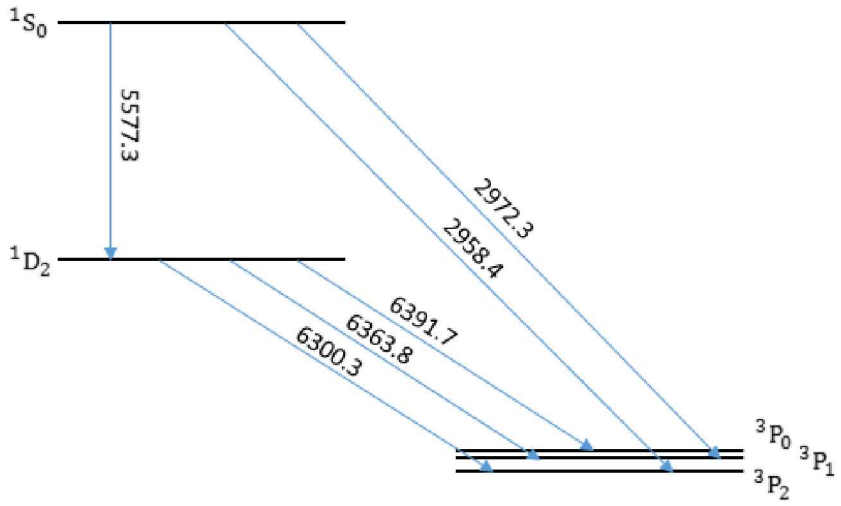

Figure 8. Partial energy level diagram for oxygen O.

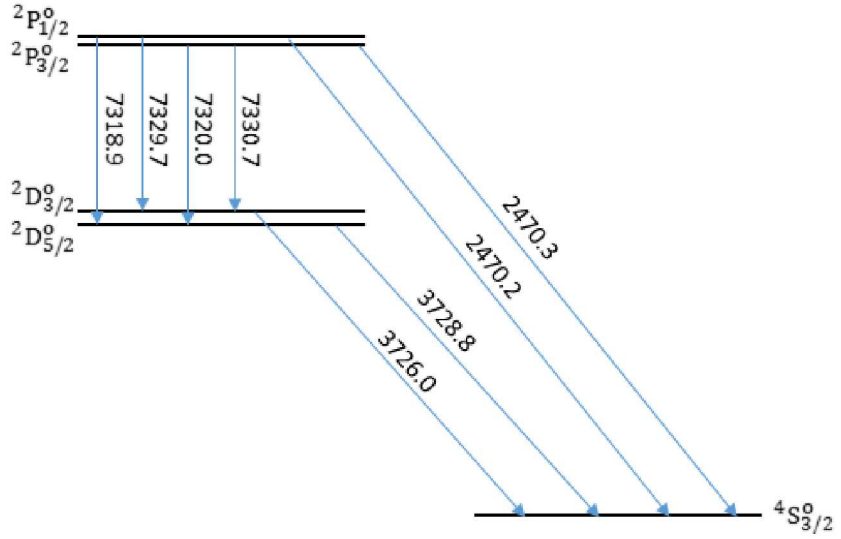

Figure 9. Partial energy level diagram for singly ionized oxygen O^+.

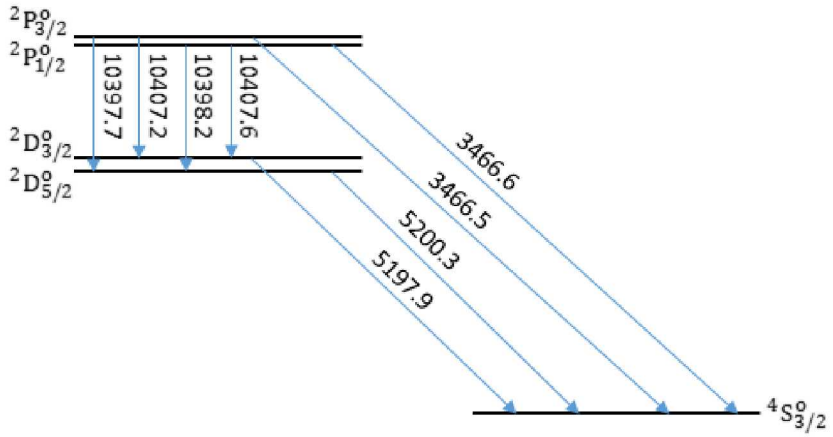

Figure 10. Partial energy level diagram for nitrogen N.

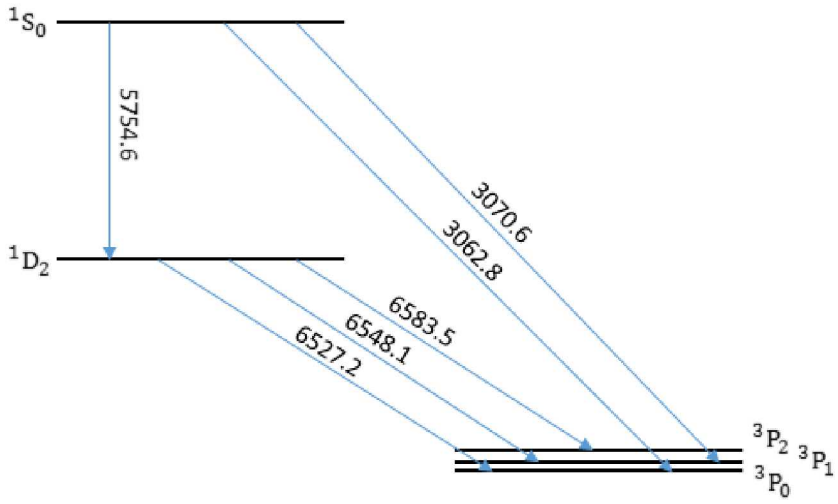

Figure 11. Partial energy level diagram for singly ionized nitrogen N^+.

hours, respectively), so the N(^{2}D^o) states mainly decay through collisional quenching. As a comparison, the 5577.3 line has a lifetime of about 0.8 seconds. The visible wavelengths from N^+ shown in Figure 11, however, have comparable lifetimes to those of O. We do not calculate those emission lines, though, because N^+ is much less abundant than O.

In order to explain where the blue line at 4278.1 Å originates, it is necessary to explain the radiative decay of molecules because it is more complex than it is for atoms. A partial energy level diagram of neutral and singly ionized dinitrogen is shown in Figures 12 and 13. Now, instead of wavelengths beside the blue arrows, there is some designation (usually called a group or system). For instance, one of the labels in Figure 13 is "first negative". This is the designation for the decay group from $N_2^+(B\,^2\Sigma_u^+)$ to $N_2^+(X\,^2\Sigma_g^+)$. The first negative group is often abbreviated 1NG.

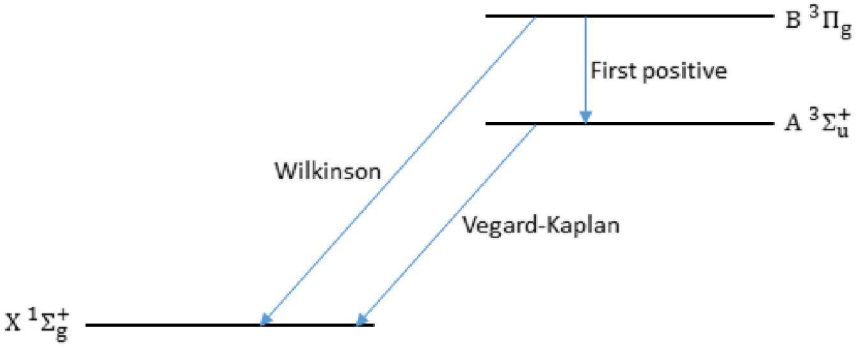

Figure 12. Partial energy level diagram for dinitrogen N_2.

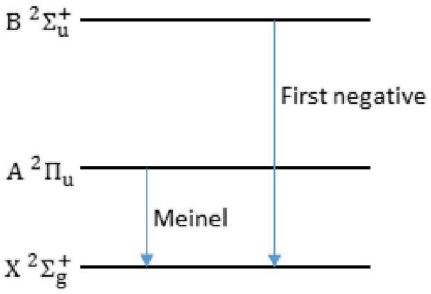

Figure 13. Partial energy level diagram for singly ionized dinitrogen N_2^+.

Within 1NG, there are many wavelengths due to the vibrational structure of dinitrogen. The state of N_2^+ or any other molecule is not completely specified until a vibrational quantum number v is given. These quantum numbers are integers beginning at 0 and increase in energy

to some maximum quantum number. Increasing the energy further causes the molecule to dissociate. The lines of 1NG consist of all possible combinations (v', v'') where v' is the quantum number for the excited state $N_2^+(B\,^2\Sigma_u^+)$ and v'' is the quantum number for the ground state $N_2^+(X\,^2\Sigma_g^+)$. A partial energy level diagram for the first negative group of singly ionized dinitrogen is shown in Figure 14. Here, the vibrational quantum numbers are shown along with the wavelength for each transition. The blue line at 4278.1 Å results from the $(0, 1)$ transition and has a lifetime of about 270 nanoseconds.

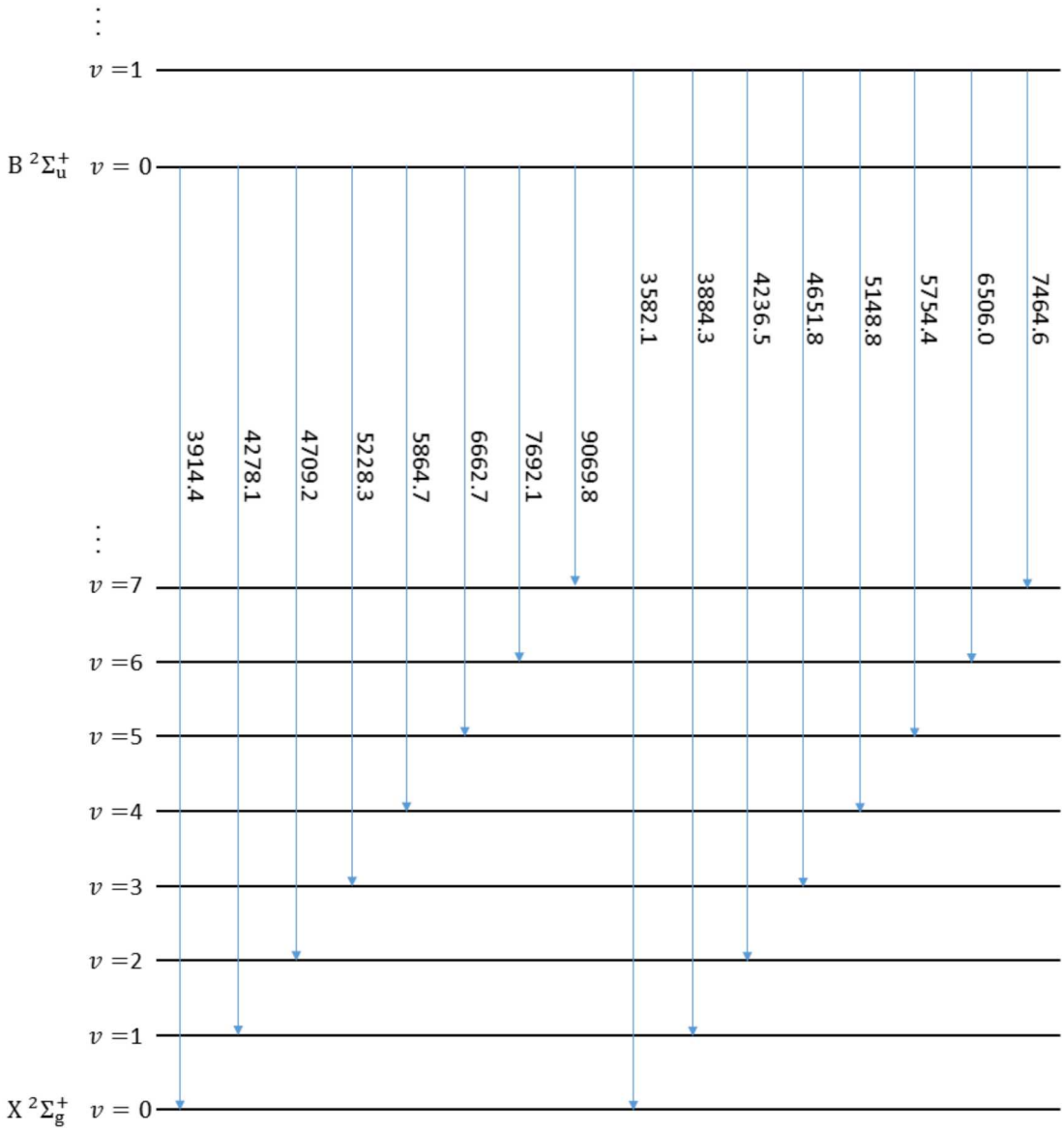

Figure 14. Partial energy level diagram for the first negative group of singly ionized dinitrogen N_2^+.

Figure 14 shows that each molecular group contributes many lines to the auroral spectrum. These individual lines can be found in references such as Krupenie [44], Lofthus and Krupenie [53], and Gilmore et al. [24] or calculated from the tables in Laher and Gilmore [50]. However, most of the lines are not visible and are very faint. For instance, the $(0, 7)$ transition of 1NG results in the non-visible line at 9069.8 Å. Further, the branching ratio for this transition is a very small fraction of 1% and makes a negligible contribution to the non-visible auroral spectrum.

5.2 Kinetic Model for the Light Emissions

In order to quantify the above light emissions, we need to know the populations of the excited states that lead to the emissions (i.e. the number densities of $O(^1D)$, $O(^1S)$, and $N_2^+(B\ ^2\Sigma_u^+, v = 0)$). We already know that electron transport through the atmosphere leads to these excited states, but there are other sources as well. We saw how every 5577.3 Å emission yields an additional $O(^1D)$ atom (see Figure 8). It turns out that there are other sources due to the naturally occurring dynamics of the upper atmosphere. For instance, another source of $O(^1D)$ is the reaction

$$O_2^+ + e_t^- \rightarrow O(^1D) + O^* \tag{271}$$

where e_t^- denotes a "thermal" or ambient electron and the asterisk on O* denotes an unspecified state of oxygen. Thus, in order to know the number density of $O(^1D)$, it is necessary to know the number densities of O_2^+ and e_t^- as well as the rate at which this reaction occurs. However, this is not sufficient because there are many other sources of $O(^1D)$. For each source, there is a reaction similar to (271) for which it is necessary to know other number densities and reaction rates.

Knowing all the necessary number densities and reaction rates would account for the sources of $O(^1D)$, but we also need to account for the sinks. We already know that for every 6300.3 Å emission, we lose one $O(^1D)$ atom. Again, this is not the only sink. It turns out that $O(^1D)$ can also decay to $O(^3P_1)$ for a 6363.8 Å emission and to a lesser extent $O(^3P_0)$ for a 6391.7 Å emission. Other sinks include quenching of the excited states as was discussed earlier. An example reaction would be

$$O(^1D) + N_2 \rightarrow O(^3P) + N_2. \tag{272}$$

In order to find the number densities of species such as O_2^+, we also need to know the reactions that lead to their production and loss. We see from Figure 2 that the ground state ion number densities are known for an unperturbed atmosphere. However, as auroral electrons stream through the atmosphere, more ions are created and many neutrals enter excited states. The rate at which this occurs can be calculated with (270).

For each species, we have the differential equation

$$\frac{dn_\xi^\eta(z)}{dz} = P_\xi^\eta(z) - L_\xi^\eta(z, n_\xi^\eta(z)) \tag{273}$$

where $P_\xi^\eta(z)$ and $L_\xi^\eta(z, n_\xi^\eta(z))$ are the sums of the production and loss rates, respectively. The production rate $P_\xi^\eta(z)$ can include excitation rates calculated through (270) and the other source rates. For example, the rate of (271) is given by

$$k n_{\mathrm{O}_2^+}(z) n_{\mathrm{e_t^-}}(z) \tag{274}$$

where k is the reaction coefficient. The loss rate $L_\xi^\eta(z, n_\xi^\eta(z))$ includes the rates for spontaneous emission and terms similar to (274). The rate for spontaneous emission is given by

$$A n_\xi^\eta(z) \tag{275}$$

where A is the Einstein coefficient.

The model that we used contains 15 species for which the number densities are calculated. For each species, there is an equation (273) giving a nonlinear, coupled IVP. The 15 computed species are $\mathrm{O}(^1\mathrm{D})$, $\mathrm{O}(^1\mathrm{S})$, $\mathrm{O}^+(^4\mathrm{S^o})$, $\mathrm{O}^+(^2\mathrm{D^o})$, $\mathrm{O}^+(^2\mathrm{P^o})$, $\mathrm{O}^-(^2\mathrm{P^o})$, $\mathrm{N_2}(\mathrm{A}\,^3\Sigma_\mathrm{u}^+)$, $\mathrm{N_2^+}(\mathrm{X}\,^2\Sigma_\mathrm{g}^+)$, $\mathrm{O_2^+}(\mathrm{X}\,^2\Pi_\mathrm{g})$, $\mathrm{N}(^4\mathrm{S^o})$, $\mathrm{N}(^2\mathrm{D^o})$, $\mathrm{N}(^2\mathrm{P^o})$, $\mathrm{N}^+(^3\mathrm{P})$, $\mathrm{NO}(\mathrm{X}\,^2\Pi)$, and $\mathrm{NO}^+(\mathrm{X}\,^1\Sigma^+)$. In addition at every time step, the ambient electron density is updated by assuming the atmosphere is net neutral. The number densities of $\mathrm{O}(^3\mathrm{P})$, $\mathrm{N_2}(\mathrm{X}\,^1\Sigma_\mathrm{g}^+)$, and $\mathrm{O_2}(\mathrm{X}\,^3\Sigma_\mathrm{g}^-)$ (the ground states) are held constant. Allowing these to vary would necessitate the inclusion of many more species into the kinetic model. Further, from Figure 1 these number densities are greater than the above 15 species, so holding them constant is a reasonable assumption.

The initial condition for this IVP is given by the unperturbed densities given in Figures 1 and 2 for the ground states. The initial conditions for the excited states are set to zero. This IVP is stiff, so a stiff solver should be used. Fortunately, stiff IVP solvers are much more available than stiff BVP solvers, so the stiffness does not present any difficulties. All production and loss reactions for the 15 species as well as reaction coefficients and Einstein coefficients are given in Appendix B.

Finally, we point out that this model does not include transport. That is, it does not allow for an individual atom or molecule to change altitude over time. This restricts the validity of the model to altitudes below about 300 km. Including transport would change the system of ODEs into a system of 15 hyperbolic PDEs. This is not necessary, though, because almost all auroral light originates below 300 km (see Rees [78]).

6 Conclusions

The purpose of this SAND report is to show how the electron transport equation (61) can be solved with a stiff solver. The problem is very stiff with eigenvalues that span over 16 orders of magnitude. We showed that we can use an upwind method that essentially decouples the fast and slow modes and applies an appropriate finite difference method to each mode. The trapezoidal rule could be used for the slow modes and backward Euler for the fast modes, giving a low order method. Further, the method generates an a priori mesh that is guaranteed to resolve the solution. It is also possible to extend this to higher orders using different finite differences. This has been outlined in Brown and Lorenz [13]. However, there is enough uncertainty in the model data (i.e. atmospheric densities, cross sections, and incident electron intensity) that using a higher order method likely will not produce better results.

We also showed that if we consider a single species atmosphere, then a representation for the exact solution can be derived. The homogeneous solution is found using separation of variables, and this is used to find a fundamental solution using the orthogonality of the eigenfunctions. This, in turn, is used in the boundary element method to find an integral representation for the exact solution. The only approximation in this formulation is in the numerical evaluation of the solution. This solution could be compared to the upwind method solution for the same single species problem. If they were found to agree with each other, this would help verify that the upwind method is correctly solving the problem.

Finally, we showed how the output of the electron transport equation can be used to predict the light output of an aurora. A kinetic model was described that would calculate the number densities of 15 different atmospheric constituents. These number densities could then be multiplied by the appropriate Einstein coefficients to determine the volume emission rate for various wavelengths of light.

It is possible to validate the model in a number of ways. For instance, if enough funding (and patience) was available, a rocket could be sent up to measure the electron intensity on a night where the aurora was present. The downward electron intensity from this data could have been used as the boundary condition at the top of the atmosphere. We could then calculate the upward electron intensity at the top of the atmosphere and compare it to the rocket measurements. If agreement was found, this would serve as a validation to our electron transport model. Further, if light emissions were measured, then we could have compared that to the output of the kinetic model for light emissions. This would validate the kinetic light emission model.

For any scientist studying the aurora, using the upwind method from this SAND report would be wise because it avoids using spurious boundary conditions and unorthodox numerical methods. Certainly, scientists will continue to study the aurora and electron transport, and it will be important for them to have reliable mathematical tools at their disposal.

References

[1] M. J. Ablowitz and A. S. Fokas. *Complex Variables: Introduction and Applications*. Cambridge University Press, New York, NY, 2nd edition, 2011.

[2] L. L. Alves. The IST-LISBON database on LXCat. *JPCS*, 565(1):1–10, July 2014.

[3] U. Ascher and R. Weiss. Collocation for singular perturbation problems I: First order systems with constant coefficients. *SIAM J. Numer. Anal.*, 20(3):537–557, June 1983.

[4] U. Ascher and R. Weiss. Collocation for singular perturbation problems II: Linear first order systems without turning points. *Math. Comp.*, 43(167):157–187, July 1984.

[5] U. M. Ascher and L. R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1998.

[6] U. M. Ascher, R. M. Mattheij, and R. D. Russell. *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1995.

[7] P. M. Banks and A. F. Nagy. Concerning the influence of elastic scattering upon photoelectron transport and escape. *J. Geophys. Res.*, 75(10):1902–1910, Apr. 1970.

[8] P. M. Banks, A. F. Chappell, and A. F. Nagy. A new model for the interaction of auroral electrons with the atmosphere: Spectra degradation, backscatter, optical emission, and ionization. *J. Geophys. Res.*, 79(10):1459–1470, Apr. 1974.

[9] G. V. G. Baranoski, J. G. Rokne, P. Shirley, T. S. Trondsen, and R. Bastos. Simulating the aurora. *J. Visual. Comput. Animat.*, 14(1):43–59, Feb. 2003.

[10] B. Basu, J. R. Jasperse, D. J. Strickland, and R. E. Jr. Daniell. Transport-theoretic model for the electron-proton-hydrogen atom aurora. I: Theory. *J. Geophys. Res.*, 98 (A12):21517–21532, Dec. 1993.

[11] S. F. Biagi. Monte carlo simulation of electron drift and diffusion in counting gases under the influence of electric and magnetic fields. *Nucl. Instr. Meth. Phys. Res. A*, 421(2):234–240, Jan. 1999.

[12] D. Bilitza and B. W. Reinisch. International reference ionosphere 2007: Improvements and new parameters. *Adv. Space Res.*, 42(4):599–609, Aug. 2008.

[13] D. L. Brown and J. Lorenz. A high-order method for stiff boundary value problems with turning points. *SIAM J. Sci. Stat. Comput.*, 8(5):790–805, Sept. 1987.

[14] D. A. Bryant. *Electron Acceleration in the Aurora and Beyond*. Institute of Physics Publishing Ltd, Philadelphia, PA, 1999.

[15] D. C. Cartwright, M. J. Brunger, L. Campbell, B. Mojarrabi, and P. J. O. Teubner. Electron impact excitation of nitric oxide under auroral conditions. *Geophys. Res. Lett.*, 25(9):1495–1498, May 1998.

[16] K. M. Case. Elementary solution of the transport equation and their applications. *Ann. Phys.*, 9(1):1–23, Jan. 1960.

[17] K. M. Case and P. F. Zweifel. *Linear Transport Theory.* Addison-Wesley Publishing Company, Inc., Reading, MA, 1967.

[18] A. Chulliat, S. Macmillan, P. Alken, C. Beggan, M. Nair, B. Hamilton, A. Woods, V. Ridley, S. Maus, and A. Thomson. The US/UK world magnetic model for 2015-2020. Technical Report 10.7289/V5TB14V7, National Geophysical Data Center, NOAA, Boulder, CO, Mar. 2015.

[19] R. J. Cicerone, W. E. Swartz, R. S. Stolarski, A. F. Nagy, and J. S. Nisbet. Thermalization and transport of photoelectrons: A comparison of theoretical approaches. *J. Geophys. Res.*, 78(28):6709–6728, Oct. 1973.

[20] L. M. Delves and J. L. Mohamed. *Computational Methods for Integral Equations.* Cambridge University Press, Cambridge, UK, 1985.

[21] L. Dieci, M. R. Osborne, and R. D. Russell. A Riccati transformation method for solving linear BVPs. I: Theoretical aspects. *SIAM J. Numer. Anal.*, 25(5):1055–1073, Oct. 1988.

[22] L. Dieci, M. R. Osborne, and R. D. Russell. A Riccati transformation method for solving linear BVPs. II: Computational aspects. *SIAM J. Numer. Anal.*, 25(5):1074–1092, Oct. 1988.

[23] D. Duffy. *Green's Functions with Applications.* CRC Press LLC, Boca Raton, FL, 2nd edition, 2015.

[24] F. R. Gilmore, R. R. Laher, and P. J. Espy. Franck-Condon factors, r-centroids, electron transition moments, and Einstein coefficients for many nitrogen and oxygen band systems. *J. Phys. Chem. Ref. Data*, 21(5):1005–1107, Sept. 1992.

[25] G. H. Golub, S. Nash, and C. Van Loan. A Hessenberg-Schur method for the problem $AX + XB = C$. *IEEE Trans. Autom. Control*, AC-24(6):909–913, Dec. 1979.

[26] D. J. Griffiths. *Introduction to Quantum Mechanics.* Pearson Prentice Hall, London, UK, 2nd edition, 2004.

[27] S. L. Guberman. The production of $O(^1D)$ from dissociative recombination of O_2^+. *Planet. Space Sci.*, 36(1):47–53, Jan. 1988.

[28] M. Hayashi. *Nonequilibrium Processes in Partially Ionized Gases*, volume 220 of *NATO ASI Series*, chapter 21, pages 333–340. Springer Springer+Business Media, LLC, New York, NY, 1990.

[29] J. H. Hecht, A. B. Christensen, D. J. Strickland, and R. R. Meier. Deducing composition and incident electron spectra from ground-based auroral optical measurements: Variations in oxygen density. *J. Geophys. Res.*, 94(A10):13553–13563, Oct. 1989.

[30] A. E. Hedin. MSIS-86 thermospheric model. *J. Geophys. Res.*, 92(A5):4649–4662, May 1987.

[31] A. E. Hedin. Extension of the MSIS thermospheric model into the middle and lower atmosphere. *J. Geophys. Res.*, 96(A2):1159–1172, Feb. 1991.

[32] M. H. Holmes. *Introduction to Perturbation Methods*. Springer, New York, NY, 2nd edition, 2013.

[33] A. A. Ionin, I. V. Kochetov, A. P. Napartovich, and N. N. Yuryshev. Physics and engineering of singlet delta oxygen production in low-temperature plasma. *J. Phys. D: Appl. Phys.*, 40(2):R25–R61, Jan. 2007.

[34] Y. Itikawa. Cross sections for electron collisions with nitrogen molecules. *J. Phys. Chem. Ref. Data*, 35(1):31–53, Mar. 2006.

[35] Y. Itikawa. Cross sections for electron collisions with oxygen molecules. *J. Phys. Chem. Ref. Data*, 38(1):1–20, Mar. 2009.

[36] Y. Itikawa and A. Ichimura. Cross sections for collisions of electrons and photons with atomic oxygen. *J. Phys. Chem. Ref. Data*, 19(3):637–651, May 1990.

[37] A. M. Jacobs and J. J. McInerney. On the Green's function of monoenergetic neutron transport theory. *Nucl. Sci. Eng.*, 22(1):119–120, May 1965.

[38] Y. K. Kim, K. K. Irikura, M. E. Rudd, M. A. Ali, P. M. Stone, J. Chang, J. S. Coursey, R. A. Dragoset, A. R. Kishore, K. J. Olsen, A. M. Sansonetti, G. G. Wiersma, D. S. Zucker, and M. A. Zucker. *Electron-Impact Cross Sections for Ionization and Excitation, Version 3.0, SRD 107*. National Institute of Standards and Technology, 2004.

[39] T. H. Koornwinder, R. Wong, R. Koekoek, and R. F. Swarttouw. *NIST Handbook of Mathematical Functions*, chapter 18, pages 435–484. Cambridge University Press, New York, NY, 2010.

[40] A. Kramida, Y. Ralchenko, and J. Reader. *NIST Atomic Spectra Database, Version 5.3, SRD 78*. National Institute of Standards and Technology, 2015.

[41] B. Kreiss and H. O. Kreiss. Numerical methods for singular perturbation problems. *SIAM J. Numer. Anal.*, 18(2):262–276, Apr. 1981.

[42] H. O. Kreiss. Central difference schemes and stiff boundary value problems. *BIT Numer. Math.*, 24(4):560–567, Dec. 1984.

[43] H. O. Kreiss, N. K. Nichols, and D. L. Brown. Numerical methods for stiff two-point boundary value problems. *SIAM J. Numer. Anal.*, 23(2):325–368, Apr. 1986.

[44] P. H. Krupenie. The spectrum of molecular oxygen. *J. Phys. Chem. Ref. Data*, 1(2): 423–534, Apr. 1972.

[45] I. Kuščer and N. J. McCormick. On the use of the Poincaré-Bertrand formula in neutron transport theory. *Nucl. Sci. Eng.*, 23(4):404, Dec. 1965.

[46] I. Kuščer and I. Vidav. On the spectrum of relaxation lengths of neutron distributions in a moderator. *J. Math. Anal. Appl.*, 25(1):80–92, Jan. 1969.

[47] I. Kuščer, N. J. McCormick, and G. C. Summerfield. Orthogonality of Case's eigenfunctions in one-speed transport theory. *Ann. Phys.*, 30(3):411–421, Dec. 1964.

[48] P. K. Kythe. *An Introduction to Boundary Element Methods*. CRC Press LLC, Boca Raton, FL, 1995.

[49] R. R. Laher and F. R. Gilmore. Updated excitation and ionization cross sections for electron impact on atomic oxygen. *J. Phys. Chem. Ref. Data*, 19(1):277–305, Jan. 1990.

[50] R. R. Laher and F. R. Gilmore. Improved fits for the vibrational and rotational cconstant of many states of nitrogen and oxygen. *J. Phys. Chem. Ref. Data*, 20(4): 685–712, July 1991.

[51] B. Lanchester and B. Gustavsson. *Imaging of Aurora to Estimate the Energy and Flux of Electron Precipitation, in Auroral Phenomenology and Magnetospheric Processes: Earth And Other Planets*, volume 197 of *Geophysical Monograph Series*. American Geophysical Union, Washington, DC, 2012.

[52] R. Link. Feautrier solution of the electron transport equation. *J. Geophys. Res.*, 97 (A1):159–169, Jan. 1992.

[53] A. Lofthus and P. H. Krupenie. The spectrum of molecular nitrogen. *J. Phys. Chem. Ref. Data*, 6(1):113–307, Jan. 1977.

[54] D. Lummerzheim. *Electron Transport and Optical Emissions in the Aurora*. PhD thesis, University of Alaska Fairbanks, Dec. 1987.

[55] D. Lummerzheim and J. Lilensten. Electron transport and energy degradation in the ionosphere: Evaluation of the numerical solution, comparison with laboratory experiments and auroral observations. *Ann. Geophys.*, 12(10/11):1039–1051, Aug. 1994.

[56] D. Lummerzheim, M. H. Rees, and H. R. Anderson. Angular dependent transport of auroral electrons in the upper atmosphere. *Planet. Space Sci.*, 37(1):109–129, Jan. 1989.

[57] T. M. MacRobert. On Neumann's formula for the Legendre functions. *Glasg. Math. J.*, 1(1):10–12, Jan. 1952.

[58] G. P. Mantas. *Electron Collision Processes in the Ionosphere*. PhD thesis, University of Illinois at Urbana-Champaign, Sept. 1973.

[59] G. P. Mantas. Theory of photoelectron thermalization and transport in the ionosphere. *Planet. Space Sci.*, 23(2):337–354, Feb. 1975.

[60] G. P. Mantas. Photoelectron thermalization and transport in the ionosphere at low energies. *Planet. Space Sci.*, 29(12):1319–1323, Dec. 1981.

[61] G. P. Mantas and S. A. Bowhill. Calculated photoelectron pitch angle and energy spectra. *Planet. Space Sci.*, 23(2):355–375, Feb. 1975.

[62] G. P. Mantas and J. C. G. Walker. The penetration of soft electrons into the ionosphere. *Planet. Space Sci.*, 24(5):409–423, May 1976.

[63] G. P. Mantas, H. C. Carlson, and V. B. Wickwar. Photoelectron flux buildup in the plasmasphere. *J. Geophys. Res.*, 83(A1):1–15, Jan. 1978.

[64] N. J. McCormick. *One-Speed Neutron Transport Problems in Plane Geometry.* PhD thesis, The University of Michigan, Feb. 1965.

[65] N. J. McCormick and I. Kuščer. Bi-orthogonality relations for solving half-space transport problems. *J. Math. Phys.*, 7(11):2036–2045, Nov. 1966.

[66] J. J. McInerney. A solution of the space-energy-angle-dependent neutron slowing-down problem. *Nucl. Sci. Eng.*, 22(2):215–234, June 1965.

[67] R. R. Meier, D. J. Strickland, J. H. Hecht, and A. B. Christensen. Deducing composition and incident electron spectra from ground-based auroral optical measurements: A study of auroral red line processes. *J. Geophys. Res.*, 94(A10):13541–13552, Oct. 1989.

[68] J. R. Mika. Neutron transport with anisotropic scattering. *Nucl. Sci. Eng.*, 11(4): 415–427, Dec. 1961.

[69] Q. L. Min, D. Lummerzheim, M. H. Rees, and K. Stamnes. Effects of a parallel electric field and the geomagnetic field in the topside ionosphere on auroral and photoelectron energy distributions. *J. Geophys. Res.*, 98(11):19223–19234, Nov. 1993.

[70] G. Monegato and L. Scuderi. Numerical integration of functions with boundary singularities. *J. Comput. Appl. Math.*, 112(2):201–214, Nov. 1999.

[71] W. L. Morgan. Electron collision data for plasma chemistry modeling. *Adv. At., Mol., Opt. Phys.*, 43(1):79–110, Jan. 2000.

[72] N. I. Muskhelishvili. *Singular Integral Equations: Boundary Problems of Function Theory and Their Application to Mathematical Physics.* Dover Publications, Inc., Mineola, NY, 2nd edition, 2008. Translated by Radok, J. R. M.

[73] A. F. Nagy and P. M. Banks. Photoelectron fluxes in the ionosphere. *J. Geophys. Res.*, 75(31):6260–6270, Nov. 1970.

[74] S. Pancheshnyi, S. Biagi, M. C. Bordage, G. J. M. Hagelaar, W. L. Morgan, A. V. Phelps, and L. C. Pitchford. The LXCat project: Electron scattering cross sections and swarm parameters for low temperature plasma modeling. *Chem. Phys.*, 398(4): 148–153, Apr. 2012.

[75] A. V. Phelps. Collision cross sections for electrons with atmospheric species. *Ann. Géophys.*, 28(3):611–625, July 1972.

[76] A. D. Polyanin and A. V. Manzhirov. *Handbook of Integral Equations*. CRC Press LLC, Boca Raton, FL, 1998.

[77] H. S. Porter, F. Varosi, and H. G. Mayr. Iterative solution of the multistream electron transport equation. I: Comparison with laboratory beam injection experiments. *J. Geophys. Res.*, 92(A6):5933–5959, June 1987.

[78] M. H. Rees. *Physics and Chemistry of the Upper Atmosphere*. Cambridge Atmospheric and Space Science. Cambridge University Press, New York, NY, 1989.

[79] C. Ringhofer. On collocation schemes for quasilinear singular perturbed boundary value problems. *SIAM J. Numer. Anal.*, 21(5):864–882, Oct. 1984.

[80] J. J. Sakurai and J. Napolitano. *Modern Quantum Mechanics*. Addison-Wesley, London, UK, 2nd edition, 2010.

[81] F. Salvat, A. Jablonski, and C. J. Powell. ELSEPA - Dirac partial-wave calculation of elastic scattering of electrons and positrons by atoms, positive ions and molecules. *Comput. Phys. Commun.*, 165(2):157–190, Jan. 2005.

[82] R. D. Sharma, J. H. Brown, A. Berk, P. K. Acharya, J. Gruninger, J. W. Duff, and R. L. Sundberg. User's manual for SAMM, SHARC and MODTRAN merged. Technical Report PL-TR-96-2090, Phillips Laboratory, Hanscom AFB, MA, Apr. 1996.

[83] F. Shure and M. Natelson. Anisotropic scattering in half-space transport problems. *Ann. Phys.*, 26(2):274–291, Feb. 1964.

[84] S. C. Solomon. Auroral electron transport using the Monte Carlo method. *Geophys. Res. Lett.*, 20(2):185–188, Feb. 1993.

[85] S. C. Solomon. Auroral particle transport using Monte Carlo and hybrid methods. *J. Geophys. Res.*, 106(A1):107–116, Jan. 2001.

[86] K. Stamnes. Analytic approach to photoelectron transport: Extensions of Stolarski's (1972) work. *J. Geophys. Res.*, 82(16):2391–2395, June 1977.

[87] K. Stamnes. *A Theoretical Investigation of the Interaction of Auroral Electrons with the Atmosphere*. PhD thesis, University of Colorado at Boulder, May 1978.

[88] K. Stamnes. Analytic approach to auroral electron transport and energy degradation. *Planet. Space Sci.*, 28(4):427–441, Apr. 1980.

[89] K. Stamnes. On the two-stream approach to electron transport and thermalization. *J. Geophys. Res.*, 86(A4):2405–2410, Apr. 1981.

[90] K. Stamnes. The theory of multiple scattering of radiation in plane parallel atmospheres. *Rev. Geophys.*, 24(2):299–310, May 1986.

[91] K. Stamnes and H. Dale. A new look at the discrete ordinate method for radiative transfer calculations in anisotropically scattering atmospheres. II: Intensity computations. *J. Atmos. Sci.*, 38(12):2696–2706, Dec. 1981.

[92] K. Stamnes and M. H. Rees. Inelastic scattering effects on photoelectron spectra and ionospheric electron temperature. *J. Geophys. Res.*, 88(A8):6301–6309, Aug. 1983.

[93] K. Stamnes and R. A Swanson. A new look at the discrete ordinate method for radiative transfer calculations in anisotropically scattering atmospheres. *J. Atmos. Sci.*, 38(2): 387–399, Feb. 1981.

[94] K. Stamnes, O. Lie-Svendsen, and M. H. Rees. The linear Boltzmann equation in slab geometry: Development and verification of a reliable and efficient solution. *Planet. Space Sci.*, 39(10):1435–1463, Oct. 1991.

[95] R. S. Stolarski. Analytic approach to photoelectron transport. *J. Geophys. Res.*, 77 (16):2862–2870, June 1972.

[96] D. J. Strickland, D. L. Book, T. P. Coffey, and J. A. Fedder. Transport equation techniques for the doposition of auroral electrons. *J. Geophys. Res.*, 81(16):2755–2764, June 1976.

[97] D. J. Strickland, R. R. Meier, J. H. Hecht, and A. B. Christensen. Deducing composition and incident electron spectra from ground-based auroral optical measurements: Theory and model results. *J. Geophys. Res.*, 94(A10):13527–13539, Oct. 1989.

[98] D. J. Strickland, R. E. Jr. Daniell, J. R. Jasperse, and B. Basu. Transport-theoretic model for the electron-proton-hydrogen atom aurora. II: Model results. *J. Geophys. Res.*, 98(A12):21533–21548, Dec. 1993.

[99] D. J. Strickland, J. Bishop, J. S. Evans, T. Majeed, P. M. Shen, R. J. Cox, R. Link, and R. E. Huffman. Atmospheric ultraviolet radiance integrated code (AURIC): Theory, software architechture, inputs, and selected results. *J. Quant. Spectrosc. Radiat. Transfer*, 62(6):689–742, Aug. 1999.

[100] W. E. Swartz. Thermalization and transport of photoelectrons: A comparison of theoretical approaches. II: Transport details for isotropic scattering. *J. Geophys. Res.*, 81(1):183–192, Jan. 1976.

[101] W. E. Swartz and K. Stamnes. Thermalization and transport of photoelectrons: A comparison of theoretical approaches. III: Addendum to details for isotropic scattering. *J. Geophys. Res.*, 82(16):2401–2404, June 1977.

[102] F. G. Tricomi. *Integral Equations*. Dover Publications, Inc., Mineola, NY, 1985.

[103] T. G. Trucano, M. Pilch, and W. L. Oberkampf. On the role of code comparison in verification and validation. Technical Report SAND2003-2752, Sandia National Laboratories, Albuquerque, NM, Aug. 2003.

[104] W. J. Wiscombe. The delta-M method: Rapid yet accurate radiative flux calculations for strongly asymmetric phase functions. *J. Atmos. Sci.*, 34(9):1408–1422, Sept. 1977.

A Electron Impact Cross Sections

All cross sections $\sigma(E)$ used in the solution of the electron transport equation (61) are shown in Figures A.1–A.14. The corresponding elastic cross section is given in each figure as a frame of reference. The energy thresholds, extrapolation formulas, and data references are given in Tables A.1–A.8. To better understand these figures and tables, a few remarks are in order. First, the hieroglyphics (also called term symbols) throughout the figures and tables are the designations of the various electronic states. It is not important what these symbols mean for our purposes. Second, there are a number of states written as a sum. This means that these states are very close to each other in excitation threshold, so their cross sections are merely added together. Third, the molecules have states such as $J = 0 \rightarrow 2$ and $v = 0 \rightarrow 1$. A different value of J designates a different rotational mode for the ground state of the molecule whereas v designates a different vibrational mode. Technically, all molecular states have various rotational and vibrational modes, but the inclusion of all these states is not necessary. The ground state rotational and vibrational cross sections are given because these states are important for energy exchange for low energy electrons. Fourth, some of the states for molecules are designated as predissociation states. This means that the molecule will jump to the shown state and then dissociate. Fifth, all ionization cross sections are for a single ionization to the ion ground state. Finally, the extrapolation formulas are provided because in each reference, the cross sections are only given up to a certain energy. For example, the ionization cross section for oxygen O is only given up to 1000 eV in Laher and Gilmore [49]. The extrapolation formula allows us to calculate the cross sections for higher energies.

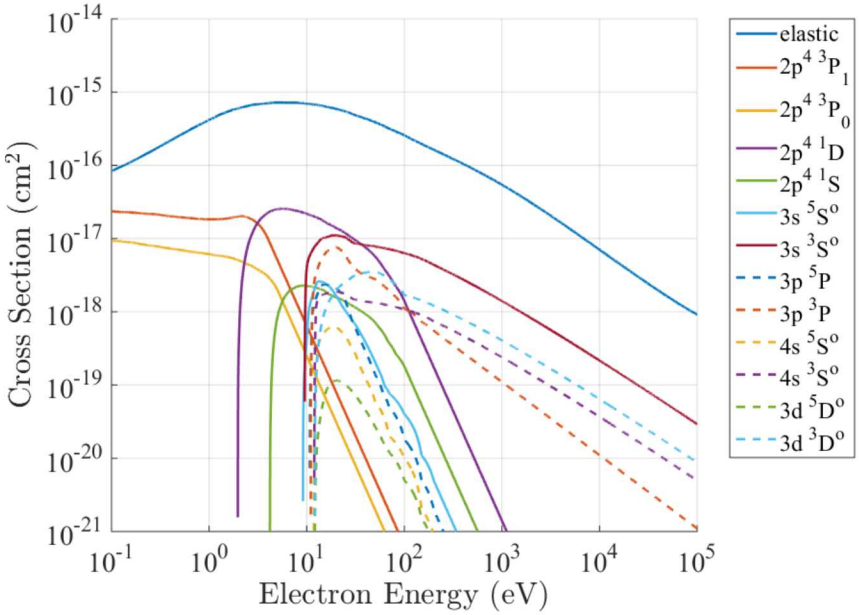

Figure A.1. Electron-oxygen cross sections.

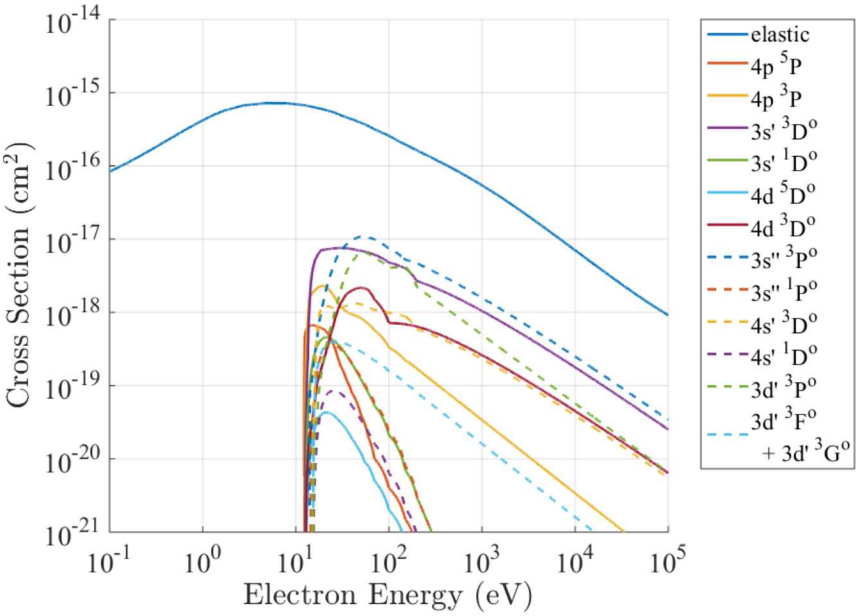

Figure A.2. Electron-oxygen cross sections (continued from Figure A.1).

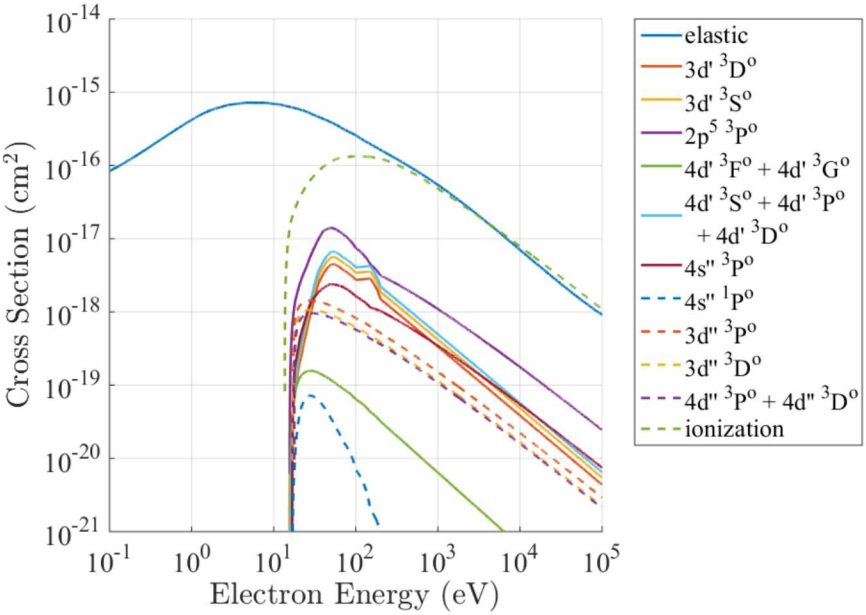

Figure A.3. Electron-oxygen cross sections (continued from Figure A.2).

Table A.1. List of electron-oxygen impact cross sections.

State	Threshold	Extrapolation Formula	Reference
$2p^4\ ^3P_2$ (elastic)	0	$(-5.89 + 1.304 \log E)E^{-1} \times 10^{-14}$	[36, 81]
$2p^4\ ^3P_1$	0.019622	$6.4E^{-3} \times 10^{-16}$	[36]
$2p^4\ ^3P_0$	0.028142	$2.5E^{-3} \times 10^{-16}$	[36]
$2p^4\ ^1D$	1.96736	$1.4E^{-3} \times 10^{-12}$	[74, 71]
$2p^4\ ^1S$	4.18975	$1.8E^{-3} \times 10^{-13}$	[74, 71]
$3s\ ^5S^o$	9.14609	$3.98E^{-3} \times 10^{-14}$	[36, 49]
$3s\ ^3S^o$	9.52136	$(-8.69 + 3.29 \log E)E^{-1} \times 10^{-16}$	[49]
$3p\ ^5P$	10.74023	$1.58E^{-3} \times 10^{-14}$	[49]
$3p\ ^3P$	10.98879	$1.1E^{-1} \times 10^{-16}$	[49]
$4s\ ^5S^o$	11.83761	$7.24E^{-3} \times 10^{-15}$	[49]
$4s\ ^3S^o$	11.93039	$(-1.48 + 0.56 \log E)E^{-1} \times 10^{-16}$	[49]
$3d\ ^5D^o$	12.07862	$4.79E^{-3} \times 10^{-15}$	[49]
$3d\ ^3D^o$	12.08702	$(-2.98 + 1.02 \log E)E^{-1} \times 10^{-16}$	[49]
$4p\ ^5P$	12.28596	$4.79E^{-3} \times 10^{-15}$	[49]
$4p\ ^3P$	12.35886	$3.39E^{-1} \times 10^{-17}$	[49]
$3s'\ ^3D^o$	12.53919	$(-1.145 + 0.317 \log E)E^{-1} \times 10^{-15}$	[36, 49]
$3s'\ ^1D^o$	12.72847	$2.4E^{-3} \times 10^{-14}$	[49]
$4d\ ^5D^o$	12.75370	$2.4E^{-3} \times 10^{-15}$	[49]
$4d\ ^3D^o$	12.75901	$(-3.06 + 0.82 \log E)E^{-1} \times 10^{-16}$	[49]
$3s''\ ^3P^o$	14.12316	$(-1.17 + 0.397 \log E)E^{-1} \times 10^{-15}$	[49]
$3s''\ ^1P^o$	14.37202	$2.4E^{-3} \times 10^{-14}$	[49]
$4s'\ ^3D^o$	15.17811	$(-2.53 + 0.7 \log E)E^{-1} \times 10^{-16}$	[49]
$4s'\ ^1D^o$	15.22497	$7.5E^{-3} \times 10^{-15}$	[49]
$3d'\ ^3P^o$	15.28685	$(2.71 + 0.326 \log E)E^{-1} \times 10^{-16}$	[49]
$3d'\ ^3F^o + 3d'\ ^3G^o$	15.40052	$1.6E^{-1} \times 10^{-17}$	[49]
$3d'\ ^3D^o$	15.40468	$(1.83 + 0.22 \log E)E^{-1} \times 10^{-16}$	[49]
$3d'\ ^3S^o$	15.41574	$(2.29 + 0.275 \log E)E^{-1} \times 10^{-16}$	[49]
$2p^5\ ^3P^o$	15.65510	$(-9.14 + 2.91 \log E)E^{-1} \times 10^{-16}$	[49]
$4d'\ ^3F^o + 4d'\ ^3G^o$	16.07665	$6.4E^{-1} \times 10^{-18}$	[49]
$4d'\ ^3S^o + 4d'\ ^3P^o$ + $4d'\ ^3D^o$	16.07980	$(2.72 + 0.324 \log E)E^{-1} \times 10^{-16}$	[49]
$4s''\ ^3P^o$	16.82239	$(-2.55 + 0.87 \log E)E^{-1} \times 10^{-16}$	[49]
$4s''\ ^1P^o$	16.90560	$8.0E^{-3} \times 10^{-15}$	[49]
$3d''\ ^3P^o$	17.10319	$(-5.9 + 3.05 \log E)E^{-1} \times 10^{-17}$	[49]
$3d''\ ^3D^o$	17.10517	$(-4.4 + 2.3 \log E)E^{-1} \times 10^{-17}$	[49]
$4d''\ ^3P^o + 4d''\ ^3D^o$	17.77480	$(-4.4 + 2.2 \log E)E^{-1} \times 10^{-17}$	[49]
$2p^3\ ^4S^o$ (ionization)	13.61806	$(-4.31 + 1.32 \log E)E^{-1} \times 10^{-14}$	[49]

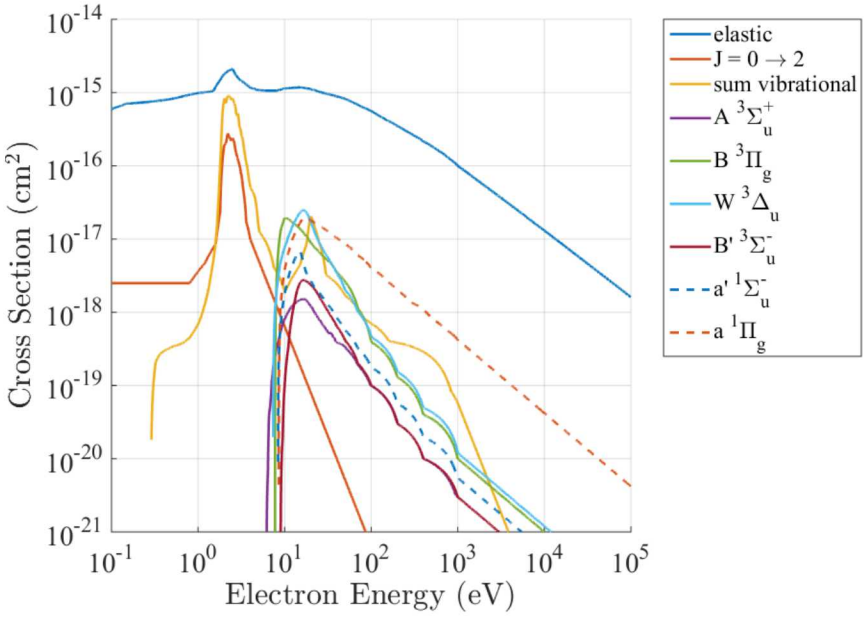

Figure A.4. Electron-dinitrogen cross sections.

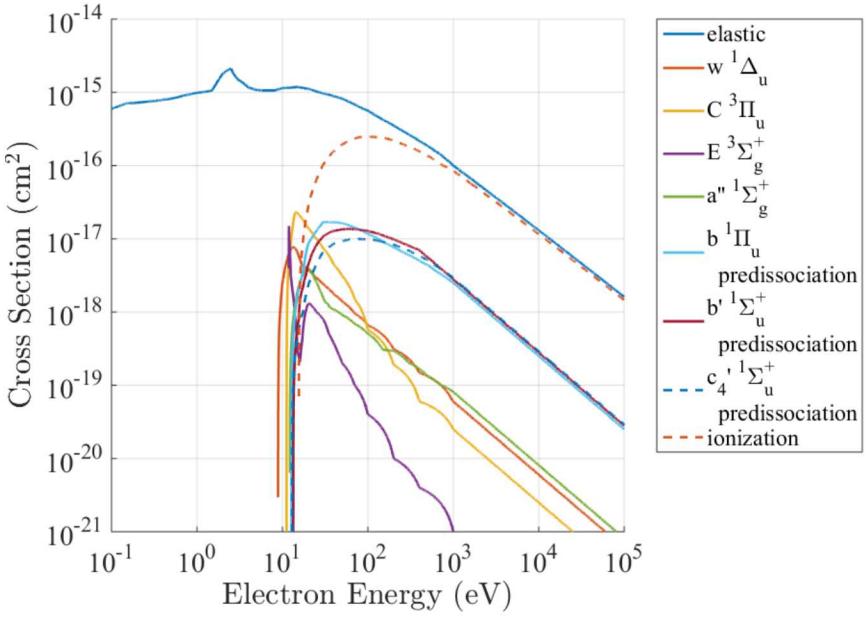

Figure A.5. Electron-dinitrogen cross sections (continued from Figure A.4).

Table A.2. List of electron-dinitrogen impact cross sections.

State	Threshold	Extrapolation Formula	Reference
X $^1\Sigma_g^+$ (elastic)	0	$(0.8818 + 1.32 \log E)E^{-1} \times 10^{-14}$	[34]
$J = 0 \rightarrow 2$	0.00148	$6.4E^{-3} \times 10^{-16}$	[74, 75]
$v = 0 \rightarrow 1$	0.2889	$4.0E^{-3} \times 10^{-11}$	[74, 11]
$v = 0 \rightarrow 2$	0.5742	$1.0E^{-3} \times 10^{-11}$	[74, 11]
$v = 0 \rightarrow 3$	0.8559	$3.0E^{-3} \times 10^{-12}$	[74, 11]
$v = 0 \rightarrow 4$	1.1342	$4.0E^{-3} \times 10^{-12}$	[74, 11]
$v = 0 \rightarrow 5$	1.4088	$1.5E^{-3} \times 10^{-13}$	[74, 11]
$v = 0 \rightarrow 6$	1.6801	$1.5E^{-3} \times 10^{-13}$	[74, 11]
$v = 0 \rightarrow 7$	1.9475	$9.0E^{-3} \times 10^{-14}$	[74, 11]
$v = 0 \rightarrow 8$	2.2115	$2.0E^{-3} \times 10^{-14}$	[74, 11]
$v = 0 \rightarrow 9$	2.4718	$7.0E^{-3} \times 10^{-15}$	[74, 11]
$v = 0 \rightarrow 10$	2.7284	$2.0E^{-3} \times 10^{-15}$	[74, 11]
$v = 0 \rightarrow 11$	2.9815	$8.0E^{-3} \times 10^{-16}$	[74, 11]
$v = 0 \rightarrow 12$	3.2310	$1.5E^{-3} \times 10^{-16}$	[74, 11]
$v = 0 \rightarrow 13$	3.4769	$7.0E^{-3} \times 10^{-17}$	[74, 11]
$v = 0 \rightarrow 14$	3.7191	$4.0E^{-3} \times 10^{-17}$	[74, 11]
$v = 0 \rightarrow 15$	3.9576	$2.5E^{-3} \times 10^{-17}$	[74, 11]
A $^3\Sigma_u^+$	6.1688	$3.0E^{-1} \times 10^{-18}$	[74, 2]
B $^3\Pi_g$	7.3532	$1.0E^{-1} \times 10^{-17}$	[74, 11]
W $^3\Delta_u$	7.3622	$1.2E^{-1} \times 10^{-17}$	[74, 2]
B′ $^3\Sigma_u^-$	8.1647	$3.0E^{-1} \times 10^{-18}$	[74, 11]
a′ $^1\Sigma_u^-$	8.3986	$5.5E^{-1} \times 10^{-18}$	[74, 2]
a $^1\Pi_g$	8.5488	$4.2E^{-1} \times 10^{-16}$	[74, 2]
w $^1\Delta_u$	8.8895	$6.0E^{-1} \times 10^{-17}$	[74, 2]
C $^3\Pi_u$	11.032	$2.5E^{-1} \times 10^{-17}$	[74, 11]
E $^3\Sigma_g^+$	11.875	$1.0E^{-1} \times 10^{-18}$	[74, 11]
a″ $^1\Sigma_g^+$	12.255	$8.1E^{-1} \times 10^{-17}$	[74, 2]
b $^1\Pi_u$ (predissociation)	12.5	$2.5E^{-1} \times 10^{-15}$	[74, 11]
b′ $^1\Sigma_u^+$ (predissociation)	12.854	$2.8E^{-1} \times 10^{-15}$	[74, 11, 34]
c_4' $^1\Sigma_u^+$ (predissociation)	12.935	$2.9E^{-1} \times 10^{-15}$	[74, 11]
X $^2\Sigma_g^+$ (ionization)	15.581	$(-0.6182 + 1.32 \log E)E^{-1} \times 10^{-14}$	[74, 11]

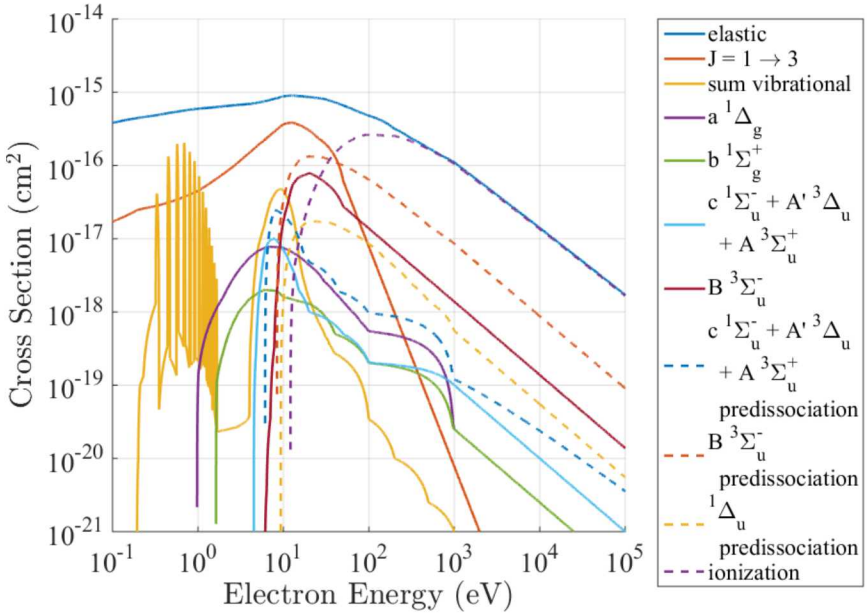

Figure A.6. Electron-dioxygen cross sections.

Table A.3. List of electron-dioxygen impact cross sections.

State	Threshold	Extrapolation Formula	Reference
X $^3\Sigma_g^-$ (elastic)	0	$(1.8818 + 1.32 \log E)E^{-1} \times 10^{-14}$	[35]
$J = 1 \rightarrow 3$	0.0002	$7.75E^{-3} \times 10^{-12}$	[74, 33]
$v = 0 \rightarrow 1$	0.193	$5.0E^{-1} \times 10^{-19}$	[74, 2]
$v = 0 \rightarrow 2$	0.386	$5.0E^{-1} \times 10^{-19}$	[74, 2]
$v = 0 \rightarrow 3$	0.579	$6.4E^{-1} \times 10^{-17}$	[74, 2]
$v = 0 \rightarrow 4$	0.772	$8.0E^{-1} \times 10^{-17}$	[74, 2]
a $^1\Delta_g$	0.977	$2.5E^{-1} \times 10^{-17}$	[74, 11]
b $^1\Sigma_g^+$	1.627	$2.5E^{-1} \times 10^{-17}$	[74, 11]
c $^1\Sigma_u^-$ + A′ $^3\Delta_u$ + A $^3\Sigma_u^+$	4.217	$1.0E^{-1} \times 10^{-16}$	[74, 11]
B $^3\Sigma_u^-$	6.12	$1.38E^{-1} \times 10^{-15}$	[35]
c $^1\Sigma_u^-$ + A′ $^3\Delta_u$ + A $^3\Sigma_u^+$ (predissociation)	6.10	$(-2.32 + 0.51 \log E)E^{-1} \times 10^{-16}$	[74, 11]
B $^3\Sigma_u^-$ (predissociation)	8.40	$(7.664 + 0.121 \log E)E^{-1} \times 10^{-15}$	[74, 11]
$^1\Delta_u$ (predissociation)	9.30	$5.5E^{-1} \times 10^{-16}$	[74, 11]
X $^2\Pi_g$	12.072	$(1.3818 + 1.32 \log E)E^{-1} \times 10^{-14}$	[74, 11]

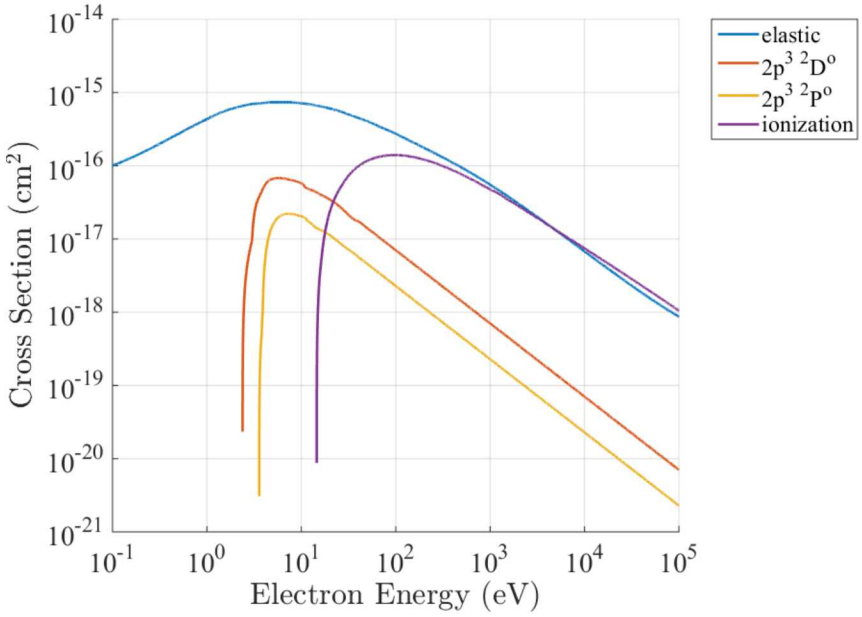

Figure A.7. Electron-nitrogen cross sections.

Table A.4. List of electron-nitrogen impact cross sections.

State	Threshold	Extrapolation Formula	Reference
$2p^3\ {}^4S^o$ (elastic)	0	$(-1.3704+0.1943\log E)E^{-1}\times10^{-13}$	[81]
$2p^3\ {}^2D^o$	2.833530	$7.02E^{-1}\times10^{-16}$	[74, 2]
$2p^3\ {}^2P^o$	3.575570	$2.28E^{-1}\times10^{-16}$	[74, 2]
$2p^2\ {}^3P$ (ionization)	14.53413	$(-4.7107+1.32\log E)E^{-1}\times10^{-14}$	[38]

Table A.5. List of electron-hydrogen impact cross sections.

State	Threshold	Extrapolation Formula	Reference
$1s\ {}^2S$ (elastic)	0	$(-6.117+0.845\log E)E^{-1}\times10^{-15}$	[74, 2, 81]
$2p\ {}^2P^o$	10.19881	$(-7.461+2.886\log E)E^{-1}\times10^{-15}$	[38]
$2s\ {}^2S$	10.19881	$(-3.77+1.73\log E)E^{-1}\times10^{-16}$	[74, 2]
$3p\ {}^2P^o + 3s\ {}^2S + 3d\ {}^2D$	12.08749	$(-1.112+0.453\log E)E^{-1}\times10^{-15}$	[74, 2]
$4p\ {}^2P^o + 4s\ {}^2S + 4d\ {}^2D + 4f\ {}^2F^o$	12.74853	$(-5.0+1.98\log E)E^{-1}\times10^{-16}$	[74, 2]
$5p\ {}^2P^o + 5s\ {}^2S + 5d\ {}^2D + 5f\ {}^2F^o + 5g\ {}^2G$	13.05450	$(-2.61+1.01\log E)E^{-1}\times10^{-16}$	[74, 2]
p^+ (ionization)	13.59843	$(2.711+1.4\log E)E^{-1}\times10^{-15}$	[38]

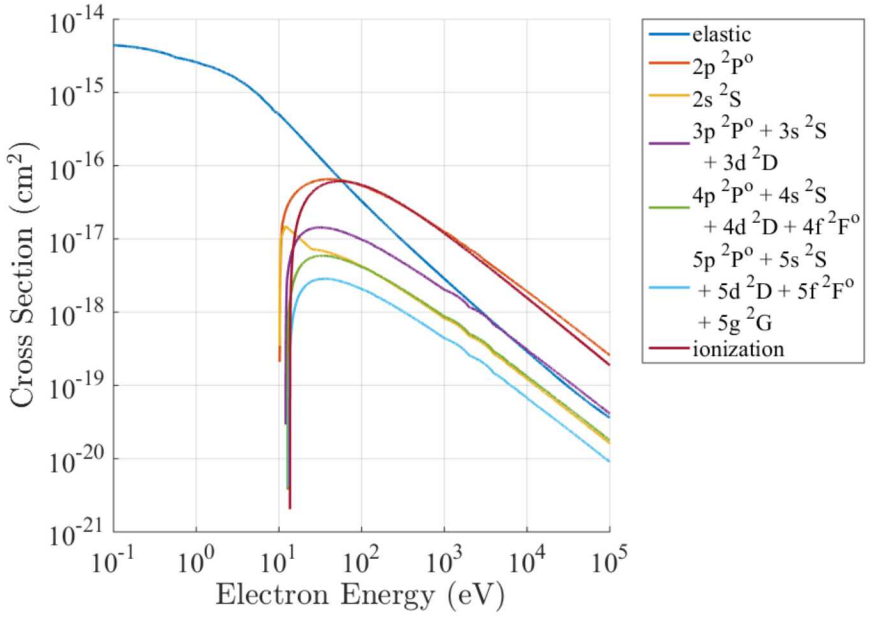

Figure A.8. Electron-hydrogen cross sections.

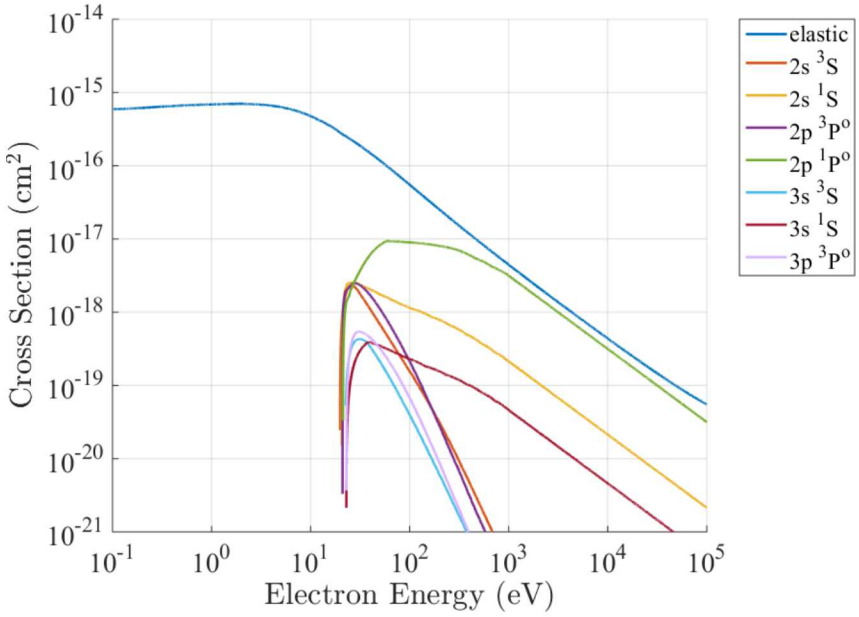

Figure A.9. Electron-helium cross sections.

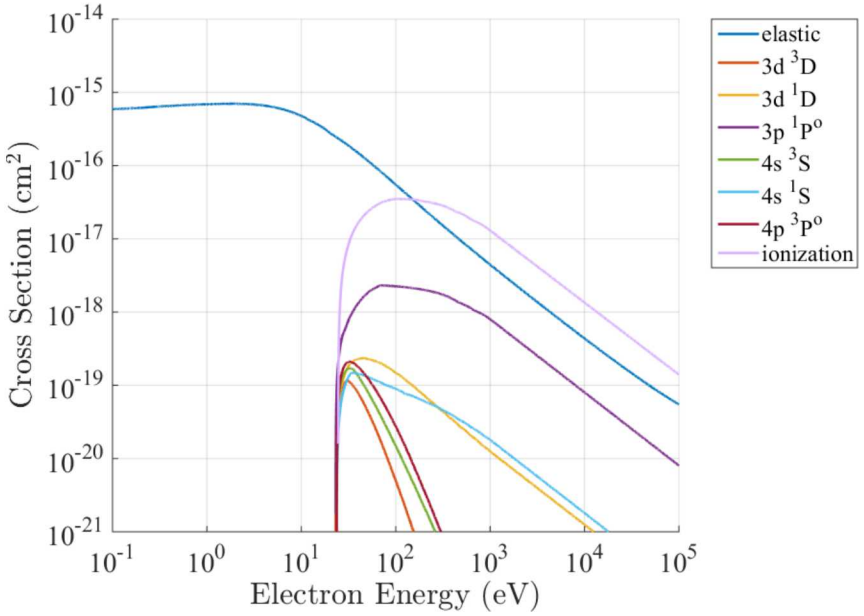

Figure A.10. Electron-helium cross sections (continued from Figure A.9).

Table A.6. List of electron-helium impact cross sections.

State	Threshold	Extrapolation Formula	Reference
$1s^2\ ^1S$ (elastic)	0	$(-9.238 + 1.279 \log E)E^{-1} \times 10^{-15}$	[74, 71, 81]
2s ^{3}S	19.81961	$3.25E^{-3} \times 10^{-13}$	[74, 11]
2s ^{1}S	20.61577	$2.13E^{-1} \times 10^{-16}$	[74, 11]
2p ^{3}P^o	20.96409	$1.75E^{-3} \times 10^{-13}$	[74, 11]
2p ^{1}P^o	21.21802	$3.15E^{-1} \times 10^{-15}$	[74, 11]
3s ^{3}S	22.71847	$5.6E^{-3} \times 10^{-14}$	[74, 11]
3s ^{1}S	22.92032	$4.6E^{-1} \times 10^{-17}$	[74, 11]
3p ^{3}P^o	23.00707	$5.0E^{-3} \times 10^{-14}$	[74, 11]
3d ^{3}D	23.07365	$1.6E^{-3} \times 10^{-15}$	[74, 11]
3d ^{1}D	23.07407	$1.25E^{-1} \times 10^{-17}$	[74, 11]
3p ^{1}P^o	23.08702	$8.0E^{-1} \times 10^{-16}$	[74, 11]
4s ^{3}S	23.59396	$1.9E^{-3} \times 10^{-14}$	[74, 11]
4s ^{1}S	23.67357	$1.8E^{-1} \times 10^{-17}$	[74, 11]
4p ^{3}P^o	23.70789	$2.0E^{-3} \times 10^{-14}$	[74, 11]
1s ^{2}S (ionization)	24.58739	$(1.1567 + 0.0207 \log E)E^{-1} \times 10^{-14}$	[74, 11]

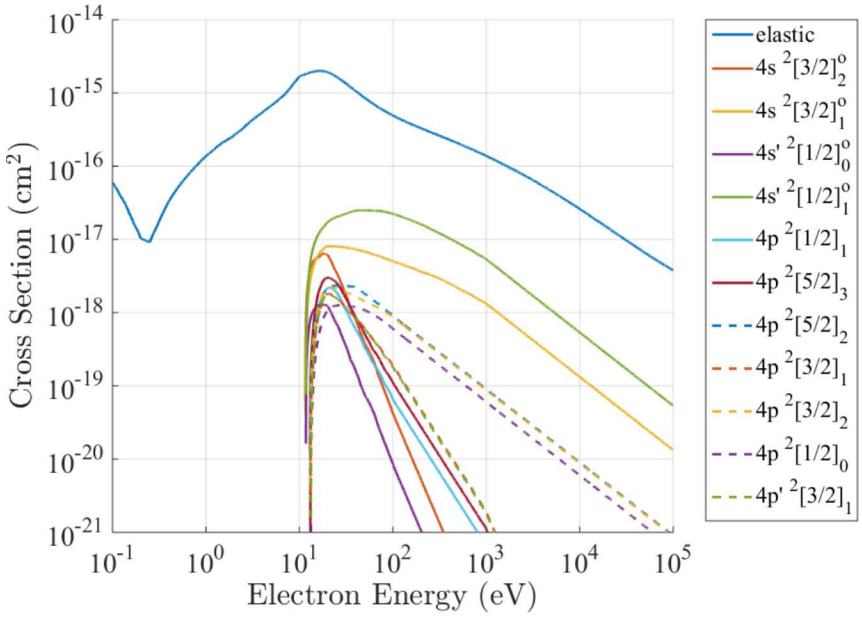

Figure A.11. Electron-argon cross sections.

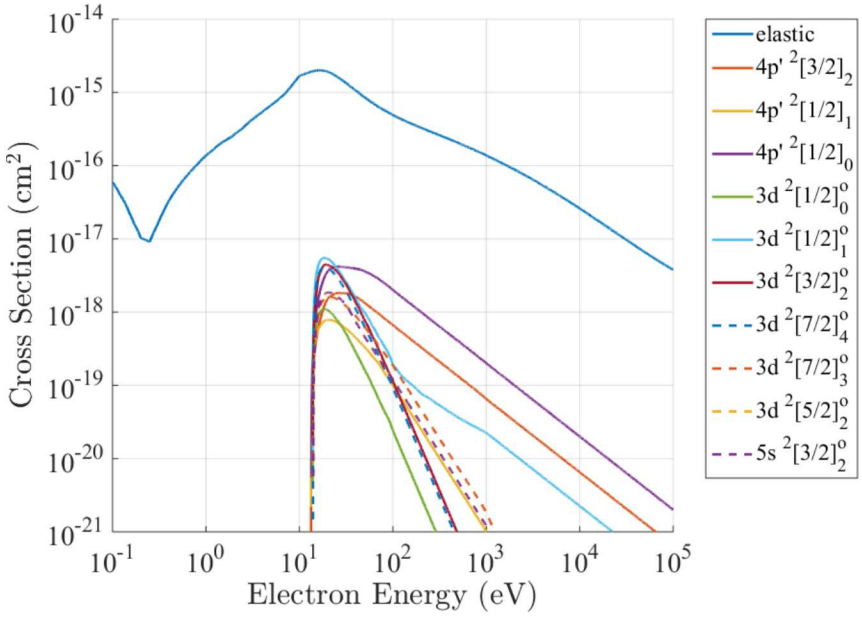

Figure A.12. Electron-argon cross sections (continued from Figure A.11).

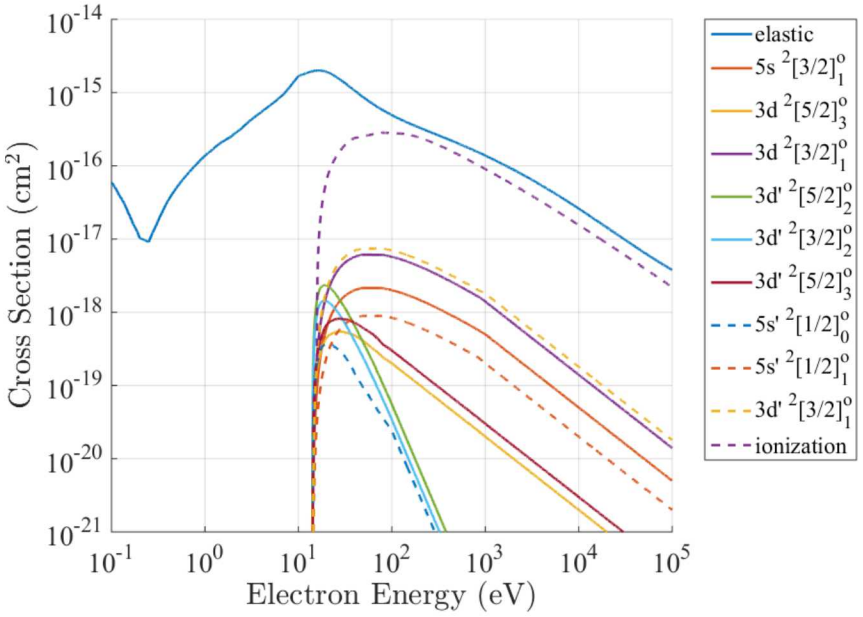

Figure A.13. Electron-argon cross sections (continued from Figure A.12).

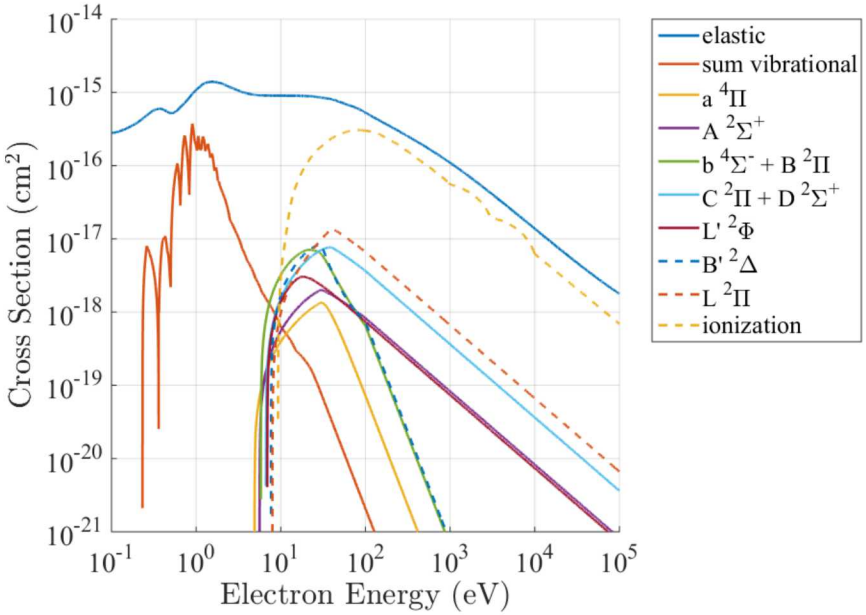

Figure A.14. Electron-nitric oxide cross sections.

Table A.7. List of electron-argon impact cross sections.

State	Threshold	Extrapolation Formula	Reference
$3p^6\ {}^1S$ (elastic)	0	$(-7.275+0.95983\log E)E^{-1}\times10^{-13}$	[74, 2, 81]
$4s\ {}^2[3/2]^o_2$	11.54835	$4.0E^{-3}\times10^{-14}$	[74, 11]
$4s\ {}^2[3/2]^o_1$	11.62359	$1.325E^{-1}\times10^{-15}$	[74, 11]
$4s'\ {}^2[1/2]^o_0$	11.72316	$8.5E^{-3}\times10^{-15}$	[74, 11]
$4s'\ {}^2[1/2]^o_1$	11.82807	$3.35E^{-1}\times10^{-15}$	[74, 11]
$4p\ {}^2[1/2]_1$	12.90702	$6.5E^{-3}\times10^{-13}$	[74, 11]
$4p\ {}^2[5/2]_3$	13.07572	$1.1E^{-3}\times10^{-12}$	[74, 11]
$4p\ {}^2[5/2]_2$	13.09487	$9.0E^{-1}\times10^{-17}$	[74, 11]
$4p\ {}^2[3/2]_1$	13.15314	$1.9E^{-3}\times10^{-12}$	[74, 11]
$4p\ {}^2[3/2]_2$	13.17178	$8.75E^{-1}\times10^{-17}$	[74, 11]
$4p\ {}^2[1/2]_0$	13.27304	$6.0E^{-1}\times10^{-17}$	[74, 11]
$4p'\ {}^2[3/2]_1$	13.28264	$1.8E^{-3}\times10^{-12}$	[74, 11]
$4p'\ {}^2[3/2]_2$	13.30223	$6.5E^{-1}\times10^{-17}$	[74, 11]
$4p'\ {}^2[1/2]_1$	13.32786	$1.0E^{-3}\times10^{-12}$	[74, 11]
$4p'\ {}^2[1/2]_0$	13.47989	$2.0E^{-1}\times10^{-16}$	[74, 11]
$3d\ {}^2[1/2]^o_0$	13.84504	$2.35E^{-3}\times10^{-14}$	[74, 11]
$3d\ {}^2[1/2]^o_1$	13.86367	$2.25E^{-1}\times10^{-17}$	[74, 11]
$3d\ {}^2[3/2]^o_2$	13.90345	$1.1E^{-3}\times10^{-13}$	[74, 11]
$3d\ {}^2[7/2]^o_4$	13.97924	$9.3E^{-3}\times10^{-14}$	[74, 11]
$3d\ {}^2[7/2]^o_3$	14.01274	$1.85E^{-3}\times10^{-12}$	[74, 11]
$3d\ {}^2[5/2]^o_2$	14.06303	$1.2E^{-3}\times10^{-12}$	[74, 11]
$5s\ {}^2[3/2]^o_2$	14.06830	$1.2E^{-3}\times10^{-12}$	[74, 11]
$5s\ {}^2[3/2]^o_1$	14.08997	$5.0E^{-1}\times10^{-16}$	[74, 11]
$3d\ {}^2[5/2]^o_3$	14.09906	$2.0E^{-1}\times10^{-17}$	[74, 11]
$3d\ {}^2[3/2]^o_1$	14.15251	$1.4E^{-1}\times10^{-15}$	[74, 11]
$3d'\ {}^2[5/2]^o_2$	14.21367	$5.5E^{-1}\times10^{-14}$	[74, 11]
$3d'\ {}^2[3/2]^o_2$	14.23402	$3.4E^{-1}\times10^{-14}$	[74, 11]
$3d'\ {}^2[5/2]^o_3$	14.23611	$3.0E^{-1}\times10^{-17}$	[74, 11]
$5s'\ {}^2[1/2]^o_0$	14.24103	$2.4E^{-3}\times10^{-14}$	[74, 11]
$5s'\ {}^2[1/2]^o_1$	14.25509	$2.0E^{-1}\times10^{-16}$	[74, 11]
$3d'\ {}^2[3/2]^o_1$	14.30367	$1.8E^{-1}\times10^{-15}$	[74, 11]
$3p^5\ {}^2P^o$ (ionization)	15.75961	$(-1.0577+0.2834\log E)E^{-1}\times10^{-13}$	[74, 11]

Table A.8. List of electron-nitric oxide impact cross sections.

State	Threshold	Extrapolation Formula	Reference
X $^2\Pi$ (elastic)	0	$(-3.1305+0.4264\log E)E^{-1}\times10^{-13}$	[74, 28, 81]
$v=0\rightarrow1$	0.232	$1.918E^{-3}\times10^{-17}$	[74, 28]
$v=0\rightarrow2$	0.455	$6.473E^{-3}\times10^{-17}$	[74, 28]
$v=0\rightarrow3$	0.677	$4.215E^{-3}\times10^{-18}$	[74, 28]
$v=0\rightarrow4$	0.896	$4.939E^{-3}\times10^{-17}$	[74, 28]
$v=0\rightarrow5$	1.110	$1.935E^{-3}\times10^{-15}$	[74, 28]
a $^4\Pi$	4.88025	$7.197E^{-3}\times10^{-14}$	[74, 15]
A $^2\Sigma^+$	5.60717	$8.271E^{-1}\times10^{-17}$	[74, 15]
b $^4\Sigma^-$ + B $^2\Pi$	5.80523	$6.489E^{-3}\times10^{-13}$	[74, 15]
C $^2\Pi$ + D $^2\Sigma^+$	6.66992	$3.634E^{-1}\times10^{-16}$	[74, 15]
L′ $^2\Phi$	6.90551	$7.363E^{-1}\times10^{-17}$	[74, 15]
B′ $^2\Delta$	7.66341	$7.302E^{-3}\times10^{-13}$	[74, 15]
L $^2\Pi$	7.9341	$6.631E^{-1}\times10^{-16}$	[74, 15]
X $^1\Sigma^+$ (ionization)	9.260	$(3.6696+0.2791\log E)E^{-1}\times10^{-14}$	[74, 28]

B Chemical Reactions of the Upper Atmosphere

All chemical reactions used in the kinetic model for auroral light emissions are listed in Tables B.9–B.38. A table is given for reactions that lead to production or loss for each species included in the model. As a result, some reactions are listed multiple times. For instance, the reaction

$$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^1D) + O(^1S) \tag{276}$$

is listed three times because it is source for both $O(^1D)$ and $O(^1S)$ and a sink for $O_2^+(X\,^2\Pi_g)$. The rates that are given for each reaction come from [99] and have units of $cm^3\ s^{-1}$ except where noted. Lastly, e_t^- denotes a "thermal" or ambient electron whereas e^- denotes a streaming or auroral electron.

Table B.9. Sources for $O(^1D)$.

Reaction	Rate
$O(^3P) + e^- \rightarrow O(^1D) + e^-$	Equation (270)
$O_2(X\,^3\Sigma_g^-) + e^- \rightarrow O(^1D) + O(^3P) + e^-$	Equation (270)
$O_2(X\,^3\Sigma_g^-) + e^- \rightarrow O(^1D) + O(^1D) + e^-$	Equation (270)
$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^1D) + O(^3P)$	$1.2207 \times 10^{-7}(T_e/300)^{-0.7}$, $T_e \leqslant 1200$ K†
	$1.0078 \times 10^{-7}(T_e/300)^{-0.56}$, $T_e > 1200$ K†
$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^1D) + O(^1D)$	$5.655 \times 10^{-8}(T_e/300)^{-0.7}$, $T_e \leqslant 1200$ K†
	$4.669 \times 10^{-8}(T_e/300)^{-0.56}$, $T_e > 1200$ K†
$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^1D) + O(^1S)$	$1.638 \times 10^{-8}(T_e/300)^{-0.7}$, $T_e \leqslant 1200$ K†
	$1.3524 \times 10^{-8}(T_e/300)^{-0.56}$, $T_e > 1200$ K†
$O_2(X\,^3\Sigma_g^-) + O(^1S) \rightarrow O(^1D) + O_2(X\,^3\Sigma_g^-)$	$7.192 \times 10^{-13} e^{-(6750-0.0151T_n^2)/8.314T_n}$
$O_2(X\,^3\Sigma_g^-) + N(^2D^o) \rightarrow O(^1D) + NO(X\,^2\Pi)$	6.0×10^{-13}
$O_2(X\,^3\Sigma_g^-) + N(^2P^o) \rightarrow O(^1D) + NO(X\,^2\Pi)$	2.2×10^{-12}
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow O(^1D) + NO^+(X\,^1\Sigma^+)$	1.8361×10^{-10}
$NO(X\,^2\Pi) + O(^1S) \rightarrow O(^1D) + NO(X\,^2\Pi)$	5.12×10^{-11}
$O(^1S) + e_t^- \rightarrow O(^1D) + e_t^-$	8.5×10^{-9}
$O(^1S) \rightarrow O(^1D) + h\nu(5577.3\ \text{Å})$	$1.26\ s^{-1}$‡

† Guberman [27]
‡ Kramida et al. [40]

Table B.10. Sinks for $O(^1D)$.

Reaction	Rate
$O(^1D) + O(^3P) \rightarrow O(^3P) + O(^3P)$	7.0×10^{-12}
$O(^1D) + N_2(X\,^1\Sigma_g^+) \rightarrow O(^3P) + N_2(X\,^1\Sigma_g^+)$	$1.8 \times 10^{-11} e^{107/T_n}$
$O(^1D) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + O_2(X\,^3\Sigma_g^-)$	$3.2 \times 10^{-11} e^{67/T_n}$
$O(^1D) + NO(X\,^2\Pi) \rightarrow O(^3P) + NO(X\,^2\Pi)$	1.5×10^{-10}
$O(^1D) + e_t^- \rightarrow O(^3P) + e_t^-$	$1.6 \times 10^{-12} T_e^{0.91}$
$O(^1D) \rightarrow O(^3P) + h\nu(6300.3, 6363.8, 6391.7\ \text{Å})^\dagger$	$7.47535 \times 10^{-3}\ \text{s}^{-1\ddagger}$

† Multiple wavelengths are due to fine structure splittings (see Figure 8).
‡ Kramida et al. [40]

Table B.11. Sources for $O(^1S)$.

Reaction	Rate	
$O(^3P) + e^- \rightarrow O(^1S) + e^-$	Equation (270)	
$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^1S) + O(^1D)$	$1.638 \times 10^{-8}(T_e/300)^{-0.7}$,	$T_e \leqslant 1200\ \text{K}^\dagger$
	$1.3524 \times 10^{-8}(T_e/300)^{-0.56}$,	$T_e > 1200\ \text{K}^\dagger$
$O(^3P) + N_2(A\,^3\Sigma_u^+) \rightarrow O(^1S) + N_2(X\,^1\Sigma_g^+)$	3.375×10^{-11}	

† Guberman [27]

Table B.12. Sinks for $O(^1S)$.

Reaction	Rate
$O(^1S) + O(^3P) \rightarrow O(^3P) + O(^3P)$	2.0×10^{-14}
$O(^1S) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + O_2(X\,^3\Sigma_g^-)$	$1.6008 \times 10^{-12} e^{-(6750-0.0151T_n^2)/8.314T_n}$
$O(^1S) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^1D) + O_2(X\,^3\Sigma_g^-)$	$7.192 \times 10^{-13} e^{-(6750-0.0151T_n^2)/8.314T_n}$
$O(^1S) + NO(X\,^2\Pi) \rightarrow O(^3P) + NO(X\,^2\Pi)$	2.88×10^{-11}
$O(^1S) + NO(X\,^2\Pi) \rightarrow O(^1D) + NO(X\,^2\Pi)$	5.12×10^{-11}
$O(^1S) + e_t^- \rightarrow O(^3P) + e_t^-$	$7.3 \times 10^{-13} T_e^{0.94}$
$O(^1S) + e_t^- \rightarrow O(^1D) + e_t^-$	8.5×10^{-9}
$O(^1S) \rightarrow O(^3P) + h\nu(2958.4, 2972.3\ \text{Å})^\dagger$	$7.5642 \times 10^{-2}\ \text{s}^{-1\ddagger}$
$O(^1S) \rightarrow O(^1D) + h\nu(5577.3\ \text{Å})$	$1.26\ \text{s}^{-1\ddagger}$

† Multiple wavelengths are due to fine structure splittings (see Figure 8).
‡ Kramida et al. [40]

Table B.13. Sources for $O^+(^4S^o)$.

Reaction	Rate	
$O(^3P) + e^- \rightarrow O^+(^4S^o) + e^- + e^-$	Equation (270)	
$O(^3P) + O^+(^2D^o) \rightarrow O^+(^4S^o) + O(^3P)$	5.0×10^{-12}	
$O(^3P) + N_2^+(X\,^2\Sigma_g^+) \rightarrow O^+(^4S^o) + N_2(X\,^1\Sigma_g^+)$	$9.8 \times 10^{-12}(T_i/300)^{-0.23}$,	$T_i \leqslant 1500$ K
	$3.64 \times 10^{-12}(T_i/300)^{0.41}$,	$T_i > 1500$ K
$O(^3P) + N^+(^3P) \rightarrow O^+(^4S^o) + N(^4S^o)$	2.2×10^{-12}	
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow O^+(^4S^o) + NO(X\,^2\Pi)$	3.66×10^{-11}	
$O^+(^2D^o) + e_t^- \rightarrow O^+(^4S^o) + e_t^-$	$6.6 \times 10^{-8}(T_e/300)^{-0.5}$	
$O^+(^2P^o) + e_t^- \rightarrow O^+(^4S^o) + e_t^-$	$3.31 \times 10^{-8}(T_e/300)^{-0.5}$	
$O^+(^2D^o) \rightarrow O^+(^4S^o) + h\nu(3726.0, 3728.8$ Å$)^\dagger$	1.0409×10^{-4} s^{-1}‡	
$O^+(^2P^o) \rightarrow O^+(^4S^o) + h\nu(2470.2, 2470.3$ Å$)^\dagger$	3.67×10^{-2} s^{-1}‡	

† Multiple wavelengths are due to fine structure splittings (see Figure 9).
‡ Kramida et al. [40]

Table B.14. Sinks for $O^+(^4S^o)$.

Reaction	Rate
$O^+(^4S^o) + e_t^- \rightarrow O(^3P)$	$3.7432 \times 10^{-12}(T_e/300)^{-0.5}$
$O^+(^4S^o) + N_2(X\,^1\Sigma_g^+) \rightarrow N(^4S^o) + NO^+(X\,^1\Sigma^+)$	$1.2 \times 10^{-12}(T_n/300)^{-0.5}(1 - e^{-3394/T_n})$
	$+8.0 \times 10^{-11} e^{-6788/T_n}$
$O^+(^4S^o) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + O_2^+(X\,^2\Pi_g)$	$1.7 \times 10^{-11}(T_n/300)^{-0.77}$
	$+8.54 \times 10^{-11} e^{-3464/T_n}$
$O^+(^4S^o) + NO(X\,^2\Pi) \rightarrow O(^3P) + NO^+(X\,^1\Sigma^+)$	8.0×10^{-13}
$O^+(^4S^o) + N(^2D^o) \rightarrow O(^3P) + N^+(^3P)$	1.3×10^{-10}
$O^+(^4S^o) + O^-(^2P^o) \rightarrow O(^3P) + O^{*\dagger}$	1.0×10^{-7}

† O* denotes $O(^5S^o)$, $O(^3S^o)$, $O(^5P)$, or $O(^3P)$.

Table B.15. Sources for $O^+(^2D^o)$.

Reaction	Rate
$O(^3P) + e^- \rightarrow O^+(^2D^o) + e^- + e^-$	Equation (270)
$O(^3P) + O^+(^2P^o) \rightarrow O^+(^2D^o) + O(^3P)$	5.2×10^{-11}
$O^+(^2P^o) + e_t^- \rightarrow O^+(^2D^o) + e_t^-$	$1.39 \times 10^{-7}(T_e/300)^{-0.5}$
$O^+(^2P^o) \rightarrow O^+(^2D^o) + h\nu(7318.9, 7320.0$ Å$)^\dagger$	7.5485×10^{-2} s^{-1}‡
$O^+(^2P^o) \rightarrow O^+(^2D^o) + h\nu(7329.7, 7330.7$ Å$)^\dagger$	7.006×10^{-2} s^{-1}‡

† Multiple wavelengths are due to fine structure splittings (see Figure 9).
‡ Kramida et al. [40]

Table B.16. Sinks for $O^+(^2D^o)$.

Reaction	Rate
$O^+(^2D^o) + O(^3P) \rightarrow O(^3P) + O^+(^4S^o)$	5.0×10^{-12}
$O^+(^2D^o) + N_2(X\,^1\Sigma_g^+) \rightarrow O(^3P) + N_2^+(X\,^2\Sigma_g^+)$	8.0×10^{-10}
$O^+(^2D^o) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + O_2^+(X\,^2\Pi_g)$	7.0×10^{-10}
$O^+(^2D^o) + NO(X\,^2\Pi) \rightarrow O(^3P) + NO^+(X\,^1\Sigma^+)$	1.2×10^{-9}
$O^+(^2D^o) + e_t^- \rightarrow O^+(^4S^o) + e_t^-$	$6.6 \times 10^{-8}(T_e/300)^{-0.5}$
$O^+(^2D^o) \rightarrow O^+(^4S^o) + h\nu(3726.0, 3728.8\ \text{Å})^\dagger$	$1.0409 \times 10^{-4}\ \text{s}^{-1\ddagger}$

† Multiple wavelengths are due to fine structure splittings (see Figure 9).
‡ Kramida et al. [40]

Table B.17. Sources for $O^+(^2P^o)$.

Reaction	Rate
$O(^3P) + e^- \rightarrow O^+(^2P^o) + e^- + e^-$	Equation (270)

Table B.18. Sinks for $O^+(^2P^o)$.

Reaction	Rate
$O^+(^2P^o) + O(^3P) \rightarrow O(^3P) + O^+(^2D^o)$	5.2×10^{-11}
$O^+(^2P^o) + N_2(X\,^1\Sigma_g^+) \rightarrow O(^3P) + N_2^+(X\,^2\Sigma_g^+)$	4.8×10^{-10}
$O^+(^2P^o) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + O_2^+(X\,^2\Pi_g)$	4.8×10^{-10}
$O^+(^2P^o) + e_t^- \rightarrow O^+(^4S^o) + e_t^-$	$3.31 \times 10^{-8}(T_e/300)^{-0.5}$
$O^+(^2P^o) + e_t^- \rightarrow O^+(^2D^o) + e_t^-$	$1.39 \times 10^{-7}(T_e/300)^{-0.5}$
$O^+(^2P^o) \rightarrow O^+(^4S^o) + h\nu(2470.2, 2470.3\ \text{Å})^\dagger$	$3.67 \times 10^{-2}\ \text{s}^{-1\ddagger}$
$O^+(^2P^o) \rightarrow O^+(^2D^o) + h\nu(7318.9, 7320.0\ \text{Å})^\dagger$	$7.5485 \times 10^{-2}\ \text{s}^{-1\ddagger}$
$O^+(^2P^o) \rightarrow O^+(^2D^o) + h\nu(7329.7, 7330.7\ \text{Å})^\dagger$	$7.006 \times 10^{-2}\ \text{s}^{-1\ddagger}$

† Multiple wavelengths are due to fine structure splittings (see Figure 9).
‡ Kramida et al. [40]

Table B.19. Sources for $O^-(^2P^o)$.

Reaction	Rate
$O(^3P) + e_t^- \rightarrow O^-(^2P^o)$	2.5×10^{-15}

Table B.20. Sinks for $O^-(^2P^o)$.

Reaction	Rate
$O^-(^2P^o) + O(^3P) \rightarrow O_2(X\,^3\Sigma_g^-) + e_t^-$	1.4×10^{-10}
$O^-(^2P^o) + O^+(^4S^o) \rightarrow O(^3P) + O^{*\dagger}$	1.0×10^{-7}

† O* denotes $O(^5S^o)$, $O(^3S^o)$, $O(^5P)$, or $O(^3P)$.

Table B.21. Sources for $N_2(A\,^3\Sigma_u^+)$.

Reaction	Rate
$N_2(X\,^1\Sigma_g^+) + e^- \rightarrow N_2(A\,^3\Sigma_u^+) + e^-$	Equation (270)

Table B.22. Sinks for $N_2(A\,^3\Sigma_u^+)$.

Reaction	Rate
$N_2(A\,^3\Sigma_u^+) + O(^3P) \rightarrow O(^3P) + N_2(X\,^1\Sigma_g^+)$	1.125×10^{-11}
$N_2(A\,^3\Sigma_u^+) + O(^3P) \rightarrow O(^1S) + N_2(X\,^1\Sigma_g^+)$	3.375×10^{-11}
$N_2(A\,^3\Sigma_u^+) + O_2(X\,^3\Sigma_g^-)$ $\rightarrow N_2(X\,^1\Sigma_g^+) + O_2(X\,^3\Sigma_g^-)$	4.0×10^{-12}
$N_2(A\,^3\Sigma_u^+) + NO(X\,^2\Pi)$ $\rightarrow N_2(X\,^1\Sigma_g^+) + NO(X\,^2\Pi)$	8.9×10^{-11}
$N_2(A\,^3\Sigma_u^+) + N(^4S^o) \rightarrow N_2(X\,^1\Sigma_g^+) + N(^2D^o)$	4.0×10^{-12}
$N_2(A\,^3\Sigma_u^+) + N(^4S^o) \rightarrow N_2(X\,^1\Sigma_g^+) + N(^2P^o)$	3.6×10^{-11}
$N_2(A\,^3\Sigma_u^+) \rightarrow N_2(X\,^1\Sigma_g^+) + h\nu$(Vegard-Kaplan)†	4.663×10^{-1} s$^{-1\ddagger}$

† The Vegard-Kaplan group consists of any transition between $N_2(A\,^3\Sigma_u^+, v')$ and $N_2(X\,^1\Sigma_g^+, v'')$ (see Figure 12).

‡ Gilmore et al. [24]

Table B.23. Sources for $N_2^+(X\,^3\Sigma_u^+)$.

Reaction	Rate
$N_2(X\,^1\Sigma_g^+) + e^- \rightarrow N_2^+(X\,^3\Sigma_u^+) + e^- + e^-$	Equation (270)
$N_2(X\,^1\Sigma_g^+) + O^+(^2D^o) \rightarrow N_2^+(X\,^3\Sigma_u^+) + O(^3P)$	8.0×10^{-10}
$N_2(X\,^1\Sigma_g^+) + O^+(^2P^o) \rightarrow N_2^+(X\,^3\Sigma_u^+) + O(^3P)$	4.8×10^{-10}
$NO(X\,^2\Pi) + N^+(^3P) \rightarrow N_2^+(X\,^3\Sigma_u^+) + O(^3P)$	7.95×10^{-11}

Table B.24. Sinks for $N_2^+(X\,^3\Sigma_u^+)$.

Reaction	Rate	
$N_2^+(X\,^3\Sigma_u^+) + e_t^- \rightarrow N(^4S^o) + N(^2D^o)$	$1.584 \times 10^{-7}(T_e/300)^{-0.39}$	
$N_2^+(X\,^3\Sigma_u^+) + e_t^- \rightarrow N(^2D^o) + N(^2D^o)$	$2.16 \times 10^{-8}(T_e/300)^{-0.39}$	
$N_2^+(X\,^3\Sigma_u^+) + O(^3P) \rightarrow N_2(X\,^1\Sigma_g^+) + O^+(^4S^o)$	$9.8 \times 10^{-12}(T_i/300)^{-0.23}$,	$T_i \leqslant 1500$ K
	$3.64 \times 10^{-12}(T_i/300)^{0.41}$,	$T_i > 1500$ K
$N_2^+(X\,^3\Sigma_u^+) + O(^3P) \rightarrow N(^2D^o) + NO^+(X\,^1\Sigma^+)$	$1.4 \times 10^{-10}(T_i/300)^{-0.44}$	
	$-9.8 \times 10^{-12}(T_i/300)^{-0.23}$,	$T_i \leqslant 1500$ K
	$5.2 \times 10^{-11}(T_i/300)^{0.2}$	
	$-3.64 \times 10^{-12}(T_i/300)^{0.41}$,	$T_i > 1500$ K
$N_2^+(X\,^3\Sigma_u^+) + O_2(X\,^3\Sigma_g^-)$		
$\rightarrow N_2(X\,^1\Sigma_g^+) + O_2^+(X\,^2\Pi_g)$	$5.0 \times 10^{-11}(T_i/300)^{-0.8}$	
$N_2^+(X\,^3\Sigma_u^+) + NO(X\,^2\Pi)$		
$\rightarrow N_2(X\,^1\Sigma_g^+) + NO^+(X\,^1\Sigma^+)$	3.3×10^{-10}	

Table B.25. Sources for $O_2^+(X\,^2\Pi_g)$.

Reaction	Rate
$O_2(X\,^3\Sigma_g^-) + e^- \rightarrow O_2^+(X\,^2\Pi_g) + e^- + e^-$	Equation (270)
$O_2(X\,^3\Sigma_g^-) + N_2^+(X\,^3\Sigma_u^+)$	
$\rightarrow O_2^+(X\,^2\Pi_g) + N_2(X\,^1\Sigma_g^+)$	$5.0 \times 10^{-11}(T_i/300)^{-0.8}$
$O_2(X\,^3\Sigma_g^-) + O^+(^4S^o) \rightarrow O_2^+(X\,^2\Pi_g) + O(^3P)$	$1.7 \times 10^{-11}(T_n/300)^{-0.77}$
	$+8.54 \times 10^{-11} e^{-3464/T_n}$
$O_2(X\,^3\Sigma_g^-) + O^+(^2D^o) \rightarrow O_2^+(X\,^2\Pi_g) + O(^3P)$	7.0×10^{-10}
$O_2(X\,^3\Sigma_g^-) + O^+(^2P^o) \rightarrow O_2^+(X\,^2\Pi_g) + O(^3P)$	4.8×10^{-10}
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow O_2^+(X\,^2\Pi_g) + N(^4S^o)$	2.1777×10^{-10}
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow O_2^+(X\,^2\Pi_g) + N(^2D^o)$	9.333×10^{-11}

Table B.26. Sinks for $O_2^+(X\,^2\Pi_g)$.

Reaction	Rate	
$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^3P) + O(^1D)$	$1.2207 \times 10^{-7}(T_e/300)^{-0.7}$,	$T_e \leqslant 1200$ K†
	$1.0078 \times 10^{-7}(T_e/300)^{-0.56}$,	$T_e > 1200$ K†
$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^1D) + O(^1D)$	$5.655 \times 10^{-8}(T_e/300)^{-0.7}$,	$T_e \leqslant 1200$ K†
	$4.669 \times 10^{-8}(T_e/300)^{-0.56}$,	$T_e > 1200$ K†
$O_2^+(X\,^2\Pi_g) + e_t^- \rightarrow O(^1D) + O(^1S)$	$1.638 \times 10^{-8}(T_e/300)^{-0.7}$,	$T_e \leqslant 1200$ K†
	$1.3524 \times 10^{-8}(T_e/300)^{-0.56}$,	$T_e > 1200$ K†
$O_2^+(X\,^2\Pi_g) + NO(X\,^2\Pi)$		
$\rightarrow O_2(X\,^3\Sigma_g^-) + NO^+(X\,^1\Sigma^+)$	4.4×10^{-10}	
$O_2^+(X\,^2\Pi_g) + N(^4S^o) \rightarrow O(^3P) + NO^+(X\,^1\Sigma^+)$	1.2×10^{-10}	

† Guberman [27]

Table B.27. Sources for $N(^4S^o)$.

Reaction	Rate
$N_2(X\,^1\Sigma_g^+) + e^- \rightarrow N(^4S^o) + N(^2D^o) + e^-$	Equation (270)
$O(^3P) + N(^2D^o) \rightarrow N(^4S^o) + O(^3P)$	1.06×10^{-12}
$O(^3P) + N^+(^3P) \rightarrow N(^4S^o) + O^+(^4S^o)$	2.2×10^{-12}
$N_2(X\,^1\Sigma_g^+) + O^+(^4S^o) \rightarrow N(^4S^o) + NO^+(X\,^1\Sigma^+)$	$1.2 \times 10^{-12}(T_n/300)^{-0.5}(1 - e^{-3394/T_n})$ $+8.0 \times 10^{-11} e^{-6788/T_n}$
$N_2(X\,^1\Sigma_g^+) + N(^2D^o) \rightarrow N(^4S^o) + N_2(X\,^1\Sigma_g^+)$	$1.0 \times 10^{-13} e^{-510/T_n}$
$N_2(X\,^1\Sigma_g^+) + N(^2P^o) \rightarrow N(^4S^o) + N_2(X\,^1\Sigma_g^+)$	2.0×10^{-18}
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow N(^4S^o) + O_2^+(X\,^2\Pi_g)$	2.1777×10^{-10}
$NO(X\,^2\Pi) + N^+(^3P) \rightarrow N(^4S^o) + NO^+(X\,^1\Sigma^+)$	4.505×10^{-10}
$N(^2D^o) + e_t^- \rightarrow N(^4S^o) + e_t^-$	$3.8 \times 10^{-12} T_e^{0.81}$
$N(^2P^o) + e_t^- \rightarrow N(^4S^o) + e_t^-$	$1.6 \times 10^{-12} T_e^{0.85}$
$N_2^+(X\,^2\Sigma_g^+) + e_t^- \rightarrow N(^4S^o) + N(^2D^o)$	$1.584 \times 10^{-7}(T_e/300)^{-0.39}$
$NO^+(X\,^1\Sigma^+) + e_t^- \rightarrow N(^4S^o) + O(^3P)$	$1.008 \times 10^{-7}(T_e/300)^{-0.75}$
$N(^2D^o) \rightarrow N(^4S^o) + h\nu(5197.9, 5200.3\ \text{Å})^\dagger$	$1.3951 \times 10^{-5}\ \text{s}^{-1}$‡
$N(^2P^o) \rightarrow N(^4S^o) + h\nu(3466.5, 3466.6\ \text{Å})^\dagger$	$4.55 \times 10^{-3}\ \text{s}^{-1}$‡

† Multiple wavelengths are due to fine structure splittings (see Figure 10).
‡ Kramida et al. [40]

Table B.28. Sinks for $N(^4S^o)$.

Reaction	Rate	
$N(^4S^o) + e^- \rightarrow N(^2D^o) + e^-$	Equation (270)	
$N(^4S^o) + e^- \rightarrow N(^2P^o) + e^-$	Equation (270)	
$N(^4S^o) + e^- \rightarrow N^+(^3P) + e^- + e^-$	Equation (270)	
$N(^4S^o) + O(^3P) \rightarrow NO(X\,^2\Pi)$	$3.33 \times 10^{-16} T_n^{-0.5}(1 - 0.567 T_n^{-0.5})$	
$N(^4S^o) + N_2(A\,^3\Sigma_u^+) \rightarrow N_2(X\,^1\Sigma_g^+) + N(^2D^o)$	4.0×10^{-12}	
$N(^4S^o) + N_2(A\,^3\Sigma_u^+) \rightarrow N_2(X\,^1\Sigma_g^+) + N(^2P^o)$	3.6×10^{-11}	
$N(^4S^o) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + NO(X\,^2\Pi)$	$1.5 \times 10^{-11} e^{-3573/T_n}$	
$N(^4S^o) + NO(X\,^2\Pi) \rightarrow O(^3P) + N_2(X\,^1\Sigma_g^+)$	$2.2 \times 10^{-11} e^{160/T_n}$,	$T_n \leqslant 400$ K
	3.3×10^{-11},	$T_n > 400$ K
$N(^4S^o) + O_2^+(X\,^2\Pi_g) \rightarrow O(^3P) + NO^+(X\,^1\Sigma^+)$	1.2×10^{-10}	

Table B.29. Sources for $N(^2D^o)$.

Reaction	Rate
$N(^4S^o) + e^- \rightarrow N(^2D^o) + e^-$	Equation (270)
$N_2(X\,^1\Sigma_g^+) + e^- \rightarrow N(^2D^o) + N(^4S^o) + e^-$	Equation (270)
$N_2^+(X\,^3\Sigma_u^+) + e_t^- \rightarrow N(^2D^o) + N(^4S^o)$	$1.584 \times 10^{-7}(T_e/300)^{-0.39}$
$N_2^+(X\,^3\Sigma_u^+) + e_t^- \rightarrow N(^2D^o) + N(^2D^o)$	$2.16 \times 10^{-8}(T_e/300)^{-0.39}$
$NO^+(X\,^1\Sigma^+) + e_t^- \rightarrow N(^2D^o) + O(^3P)$	$3.192 \times 10^{-7}(T_e/300)^{-0.75}$
$O(^3P) + N(^2P^o) \rightarrow N(^2D^o) + O(^3P)$	1.7×10^{-11}
$O(^3P) + N_2^+(X\,^3\Sigma_u^+) \rightarrow N(^2D^o) + NO^+(X\,^1\Sigma^+)$	$1.4 \times 10^{-10}(T_i/300)^{-0.44}$
	$-9.8 \times 10^{-12}(T_i/300)^{-0.23}$, $T_i \leqslant 1500$ K
	$5.2 \times 10^{-11}(T_i/300)^{0.2}$
	$-3.64 \times 10^{-12}(T_i/300)^{0.41}$, $T_i > 1500$ K
$N_2(A\,^3\Sigma_u^+) + N(^4S^o) \rightarrow N(^2D^o) + N_2(X\,^1\Sigma_g^+)$	4.0×10^{-12}
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow N(^2D^o) + O_2^+(X\,^2\Pi_g)$	9.333×10^{-11}
$NO(X\,^2\Pi) + N(^2P^o) \rightarrow N(^2D^o) + NO(X\,^2\Pi)$	3.0×10^{-11}
$N(^4S^o) + N(^2P^o) \rightarrow N(^2D^o) + N(^4S^o)$	6.0×10^{-13}
$N(^2P^o) + e_t^- \rightarrow N(^2D^o) + e_t^-$	9.5×10^{-9}
$N(^2P^o) \rightarrow N(^2D^o) + h\nu(10397.7, 10398.2$ Å$)^\dagger$	4.776×10^{-2} s^{-1}‡
$N(^2P^o) \rightarrow N(^2D^o) + h\nu(10407.2, 10407.6$ Å$)^\dagger$	4.027×10^{-2} s^{-1}‡

† Multiple wavelengths are due to fine structure splittings (see Figure 10).
‡ Kramida et al. [40]

Table B.30. Sinks for $N(^2D^o)$.

Reaction	Rate
$N(^2D^o) + O(^3P) \rightarrow O(^3P) + N(^4S^o)$	1.06×10^{-12}
$N(^2D^o) + O(^3P) \rightarrow NO^+(X\,^1\Sigma^+) + e_t^-$	$2.5 \times 10^{-18} T_n^{0.5}(2205 + T_n)e^{-4410/T_n}$
$N(^2D^o) + N_2(X\,^1\Sigma_g^+) \rightarrow N_2(X\,^1\Sigma_g^+) + N(^4S^o)$	$1.0 \times 10^{-13} e^{-510/T_n}$
$N(^2D^o) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + NO(X\,^2\Pi)$	5.9×10^{-12}
$N(^2D^o) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^1D) + NO(X\,^2\Pi)$	6.0×10^{-13}
$N(^2D^o) + NO(X\,^2\Pi) \rightarrow O(^3P) + N_2(X\,^1\Sigma_g^+)$	6.7×10^{-11}
$N(^2D^o) + O^+(^4S^o) \rightarrow O(^3P) + N^+(^3P)$	1.3×10^{-10}
$N(^2D^o) + e_t^- \rightarrow N(^4S^o) + e_t^-$	$3.8 \times 10^{-12} T_e^{0.81}$
$N(^2D^o) \rightarrow N(^4S^o) + h\nu(5197.9, 5200.3$ Å$)^\dagger$	1.3951×10^{-5} s^{-1}‡

† Multiple wavelengths are due to fine structure splittings (see Figure 10).
‡ Kramida et al. [40]

Table B.31. Sources for $N(^2P^o)$.

Reaction	Rate
$N(^4S^o) + e^- \rightarrow N(^2P^o) + e^-$	Equation (270)
$N_2(A\,^3\Sigma_u^+) + N(^4S^o) \rightarrow N(^2P^o) + N_2(X\,^1\Sigma_g^+)$	3.6×10^{-11}

Table B.32. Sinks for $N(^2P^o)$.

Reaction	Rate
$N(^2P^o) + O(^3P) \rightarrow O(^3P) + N(^2D^o)$	1.7×10^{-11}
$N(^2P^o) + N_2(X\,^1\Sigma_g^+) \rightarrow N_2(X\,^1\Sigma_g^+) + N(^4S^o)$	2.0×10^{-18}
$N(^2P^o) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^1D) + NO(X\,^2\Pi)$	2.2×10^{-12}
$N(^2P^o) + NO(X\,^2\Pi) \rightarrow NO(X\,^2\Pi) + N(^2D^o)$	3.0×10^{-11}
$N(^2P^o) + N(^4S^o) \rightarrow N(^4S^o) + N(^2D^o)$	6.0×10^{-13}
$N(^2P^o) + e_t^- \rightarrow N(^4S^o) + e_t^-$	$1.6 \times 10^{-12} T_e^{0.85}$
$N(^2P^o) + e_t^- \rightarrow N(^2D^o) + e_t^-$	9.5×10^{-9}
$N(^2P^o) \rightarrow N(^4S^o) + h\nu(3466.5, 3466.6\ \text{Å})^\dagger$	$4.55 \times 10^{-3}\ s^{-1\ddagger}$
$N(^2P^o) \rightarrow N(^2D^o) + h\nu(10397.7, 10398.2\ \text{Å})^\dagger$	$4.776 \times 10^{-2}\ s^{-1\ddagger}$
$N(^2P^o) \rightarrow N(^2D^o) + h\nu(10407.2, 10407.6\ \text{Å})^\dagger$	$4.027 \times 10^{-2}\ s^{-1\ddagger}$

† Multiple wavelengths are due to fine structure splittings (see Figure 10).
‡ Kramida et al. [40]

Table B.33. Sources for $N^+(^3P)$.

Reaction	Rate
$N(^4S^o) + e^- \rightarrow N^+(^3P) + e^- + e^-$	Equation (270)
$N(^2D^o) + O^+(^4S^o) \rightarrow N^+(^3P) + O(^3P)$	1.3×10^{-13}

Table B.34. Sinks for $N^+(^3P)$.

Reaction	Rate
$N^+(^3P) + O(^3P) \rightarrow N(^4S^o) + O^+(^4S^o)$	2.2×10^{-12}
$N^+(^3P) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^3P) + NO^+(X\,^1\Sigma^+)$	7.869×10^{-11}
$N^+(^3P) + O_2(X\,^3\Sigma_g^-) \rightarrow O(^1D) + NO^+(X\,^1\Sigma^+)$	1.8361×10^{-10}
$N^+(^3P) + O_2(X\,^3\Sigma_g^-) \rightarrow N(^4S^o) + O_2^+(X\,^2\Pi_g)$	2.1777×10^{-10}
$N^+(^3P) + O_2(X\,^3\Sigma_g^-) \rightarrow N(^2D^o) + O_2^+(X\,^2\Pi_g)$	9.333×10^{-11}
$N^+(^3P) + O_2(X\,^3\Sigma_g^-) \rightarrow NO(X\,^2\Pi) + O^+(^4S^o)$	3.66×10^{-11}
$N^+(^3P) + NO(X\,^2\Pi) \rightarrow O(^3P) + N_2^+(X\,^2\Sigma_g^+)$	7.95×10^{-11}
$N^+(^3P) + NO(X\,^2\Pi) \rightarrow N(^4S^o) + NO^+(X\,^1\Sigma^+)$	4.505×10^{-10}

Table B.35. Sources for $NO(X\,^2\Pi)$.

Reaction	Rate
$O(^3P) + N(^4S^o) \rightarrow NO(X\,^2\Pi)$	$3.33 \times 10^{-16} T_n^{-0.5}(1 - 0.567 T_n^{-0.5})$
$O_2(X\,^3\Sigma_g^-) + N(^4S^o) \rightarrow NO(X\,^2\Pi) + O(^3P)$	$1.5 \times 10^{-11} e^{-3573/T_n}$
$O_2(X\,^3\Sigma_g^-) + N(^2D^o) \rightarrow NO(X\,^2\Pi) + O(^3P)$	5.9×10^{-12}
$O_2(X\,^3\Sigma_g^-) + N(^2D^o) \rightarrow NO(X\,^2\Pi) + O(^1D)$	6.0×10^{-13}
$O_2(X\,^3\Sigma_g^-) + N(^2P^o) \rightarrow NO(X\,^2\Pi) + O(^1D)$	2.2×10^{-12}
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow NO(X\,^2\Pi) + O^+(^4S^o)$	3.66×10^{-11}

Table B.36. Sinks for $NO(X\,^2\Pi)$.

Reaction	Rate	
$NO(X\,^2\Pi) + e^- \rightarrow NO^+(X\,^1\Sigma^+) + e^- + e^-$	Equation (270)	
$NO(X\,^2\Pi) + N(^4S^o) \rightarrow O(^3P) + N_2(X\,^1\Sigma_g^+)$	$2.2 \times 10^{-11} e^{160/T_n}$,	$T_n \leqslant 400$ K
	3.3×10^{-11},	$T_n > 400$ K
$NO(X\,^2\Pi) + N(^2D^o) \rightarrow O(^3P) + N_2(X\,^1\Sigma_g^+)$	6.7×10^{-11}	
$NO(X\,^2\Pi) + O^+(^4S^o) \rightarrow O(^3P) + NO^+(X\,^1\Sigma^+)$	8.0×10^{-13}	
$NO(X\,^2\Pi) + O^+(^2D^o) \rightarrow O(^3P) + NO^+(X\,^1\Sigma^+)$	1.2×10^{-9}	
$NO(X\,^2\Pi) + N_2^+(X\,^3\Sigma_u^+)$ $\rightarrow N_2(X\,^1\Sigma_g^+) + NO^+(X\,^1\Sigma^+)$	3.3×10^{-10}	
$NO(X\,^2\Pi) + O_2^+(X\,^2\Pi_g)$ $\rightarrow O_2(X\,^3\Sigma_g^-) + NO^+(X\,^1\Sigma^+)$	4.4×10^{-10}	
$NO(X\,^2\Pi) + N^+(^3P) \rightarrow O(^3P) + N_2^+(X\,^2\Sigma_g^+)$	7.95×10^{-11}	
$NO(X\,^2\Pi) + N^+(^3P) \rightarrow N(^4S^o) + NO^+(X\,^1\Sigma^+)$	4.505×10^{-10}	

Table B.37. Sources for $NO^+(X\,^1\Sigma^+)$.

Reaction	Rate
$NO(X\,^2\Pi) + e^- \rightarrow NO^+(X\,^1\Sigma^+) + e^- + e^-$	Equation (270)
$O(^3P) + N(^2D^o) \rightarrow NO^+(X\,^1\Sigma^+) + e_t^-$	$2.5 \times 10^{-18} T_n^{0.5}(2205 + T_n)e^{-4410/T_n}$
$O(^3P) + N_2^+(X\,^3\Sigma_u^+) \rightarrow NO^+(X\,^1\Sigma^+) + N(^2D^o)$	$1.4 \times 10^{-10}(T_i/300)^{-0.44}$
	$-9.8 \times 10^{-12}(T_i/300)^{-0.23}, \quad T_i \leqslant 1500$ K
	$5.2 \times 10^{-11}(T_i/300)^{0.2}$
	$-3.64 \times 10^{-12}(T_i/300)^{0.41}, \quad T_i > 1500$ K
$N_2(X\,^1\Sigma_g^+) + O^+(^4S^o) \rightarrow NO^+(X\,^1\Sigma^+) + N(^4S^o)$	$1.2 \times 10^{-12}(T_n/300)^{-0.5}(1 - e^{-3394/T_n})$
	$+8.0 \times 10^{-11} e^{-6788/T_n}$
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow NO^+(X\,^1\Sigma^+) + O(^3P)$	7.869×10^{-11}
$O_2(X\,^3\Sigma_g^-) + N^+(^3P) \rightarrow NO^+(X\,^1\Sigma^+) + O(^1D)$	1.8361×10^{-10}
$NO(X\,^2\Pi) + O^+(^4S^o) \rightarrow NO^+(X\,^1\Sigma^+) + O(^3P)$	8.0×10^{-13}
$NO(X\,^2\Pi) + O^+(^2D^o) \rightarrow NO^+(X\,^1\Sigma^+) + O(^3P)$	1.2×10^{-9}
$NO(X\,^2\Pi) + N_2^+(X\,^2\Sigma_g^+)$	
$\rightarrow NO^+(X\,^1\Sigma^+) + N_2(X\,^1\Sigma_g^+)$	3.3×10^{-10}
$NO(X\,^2\Pi) + O_2^+(X\,^2\Pi_g)$	
$\rightarrow NO^+(X\,^1\Sigma^+) + O_2(X\,^3\Sigma_g^-)$	4.4×10^{-10}
$NO(X\,^2\Pi) + N^+(^3P) \rightarrow NO^+(X\,^1\Sigma^+) + N(^4S^o)$	4.505×10^{-10}
$N(^4S^o) + O_2^+(X\,^2\Pi_g) \rightarrow NO^+(X\,^1\Sigma^+) + O(^3P)$	1.2×10^{-10}

Table B.38. Sinks for $NO^+(X\,^1\Sigma^+)$.

Reaction	Rate
$NO^+(X\,^1\Sigma^+) + e_t^- \rightarrow O(^3P) + N(^4S^o)$	$1.008 \times 10^{-7}(T_e/300)^{-0.75}$
$NO^+(X\,^1\Sigma^+) + e_t^- \rightarrow O(^3P) + N(^2D^o)$	$3.192 \times 10^{-7}(T_e/300)^{-0.75}$

DISTRIBUTION:

1	MS 0406	Fawn R. Turner, 5785
1	MS 0406	William C. Sailor, 5785
1	MS 0406	Mark C. Woods, 5785
1	MS 0359	D. Chavez, LDRD Office, 1911
1	MS 0899	Technical Library, 9536 (electronic copy)


Sandia National Laboratories