A numerical model for heating by nanoflares in 3D MHD simulations of the solar atmosphere

Lars Frogner

Thesis submitted for the degree of
Master of Science in Astronomy
60 credits

Institute of Theoretical Astrophysics
Faculty of Mathematics and Natural Sciences

University of Oslo

June 2018

Abstract

Solar flares play a key role in the dynamics of the solar atmosphere. They occur with a wide range of energies, from the largest X-class flares representing some of the most violent releases of energy in the solar system, to the tiny nanoflares whose faintness imposes a considerable observational challenge for current telescopes. While unimpressive in isolation, nanoflares are believed to occur in great numbers, and their collective heating has been proposed as a possible explanation for the high temperature of the corona. Detailed 1D numerical models have provided great insight into the behaviour of individual flares, but they are unsuited for examining how flares collectively influence the atmosphere.

We develop a numerical model for the generation and evolution of accelerated electron beams associated with small flares in the solar atmosphere. This is integrated into the 3D radiative magnetohydrodynamics code Bifrost. The model tackles four primary tasks: detecting electron acceleration sites, determining the resulting electron energies, tracing the trajectories of the accelerated electron beams and computing the amount of heat they deposit. The latter two tasks are the focus of this thesis.

A realistic simulation of the solar atmosphere is run with the electron beam physics included. Regions of strong beam heating are produced in the lower transition region, at locations where magnetic coronal loops are anchored in the lower atmosphere. The heat input shifts the transition region downwards locally by approximately 10 km, which is expected to lead to a slightly enhanced emission in transition region spectral lines. A modest increase in pressure accelerates the plasma at the heating sites upwards by a few kilometres per second.

The relatively small response of the plasma to the presence of electron beams is a consequence of the abnormally cool and dense corona of the initial atmosphere. This leads to fewer high-energy electrons being generated and more of the beam energy being deposited in the corona. A larger simulation box is likely required for obtaining an atmosphere capable of producing stronger flare events.

Acknowledgements

I would like to thank my advisor Boris Vilhelm Gudiksen for being an invaluable source of knowledge and motivation during the course of this project. Thank you also to Helle Bakke for being a great collaborator and for our many fruitful academic and not-so-academic discussions. I also owe a thanks to the other people at ITA who have aided me in this project, as well as to my parents for always being supportive. Finally, I am very grateful to my partner Jenny for her vast supply of encouragement, support and patience, and for always pretending to be interested when I explain things from my thesis.

Contents

Chapter 1

Introduction

Our host star, the Sun, is a radiant ball of plasma making up the bulk of the matter in the solar system. It is kept stable by a balance between the inward gravitational pull and the outward push due to the rise in pressure with depth. This balance is sustained by nuclear fusion of hydrogen to helium in the Sun's core. Outside the core, in a layer known as the *radiative region*, the energy produced in the core is transported outwards primarily in the form of radiation. Farther out, in the *convective region*, the plasma is convectively unstable, meaning that heated cells of plasma become buoyant and start rising towards the surface. This creates circular movements of hot rising plasma and cold sinking plasma, which is the main mechanism of energy transport between the top of the radiative region and the solar surface.

On top of the convective region lies the solar atmosphere. This is where all the radiation we can observe from the Sun is emitted. The atmosphere is a very diverse place, with plasma in a range of different temperatures and densities being heated, cooled and accelerated by interactions with the Sun's electromagnetic field. The bottom of the atmosphere is where most of the visible light escapes from the Sun, and is known as the *photosphere*. Above lies the *chromosphere*, where the plasma becomes less dense and undergoes large changes in temperature. The hot upper chromosphere transitions rapidly into the extremely hot and tenuous *corona*.

The Sun rotates, but separate parts of the Sun rotate in different ways. The core and radiative region rotate together as a solid body, while the convective region exhibits differential rotation: it rotates faster near the solar equator and slower near the poles. Complex plasma flows in the convective region due to solar rotation and convection are believed to be the source of the Sun's magnetic field, via a mechanism known as the *solar dynamo*. On large scales, the solar magnetic field resembles that of a bar magnet, with the bar aligned with the Sun's axis of rotation. However, the smaller scale magnetic field is highly non-uniform and dynamic, and is the source of most of the activity in the solar atmosphere.

Regions of high activity and strong magnetic fields tend to produce energetic phenomena like *solar flares*. Flares are typically caused by loops of concentrated magnetic field rising up from the solar interior and interacting with the pre-existing magnetic field

in the corona. This releases magnetic energy in a process known as *magnetic reconnection*, which heats the plasma and accelerates charged particles like electrons and ions into very high energies. Beams of high-energy particles travel downwards through the corona along magnetic field lines, eventually colliding with the denser chromospheric plasma and releasing large amounts of energy.

The different phenomena believed to govern solar flares have been extensively studied, and detailed numerical models have been created for many of them. Examples include 2D simulations of particle acceleration during magnetic reconnection (e.g. Borissov et al. (2017)) and 1D simulations of the interactions between a beam of charged particles and the solar atmospheric plasma (e.g. Allred et al. (2015)). Such models are critical tools for understanding the physics behind individual flare events. However, there is also value in trying to understand how large numbers of flares of different sizes collectively influence the atmosphere. In particular, very small but frequent flares called *microflares* and *nanoflares* have been proposed as an important mechanism for heating the chromosphere and corona (Parker, 1988). Recent observations and modelling by Testa et al. (2014) provides support for this theory. Confirming this observationally remains a great challenge, because nanoflares are too faint to be directly observable at all relevant wavelengths with today's instruments. Numerical simulations are thus a necessary tool for gaining an increased understanding of these processes.

The primary purpose of this thesis is to develop a numerical model for the heating due to electron beams generated by reconnection in a realistic solar atmospheric simulation, and use it to study the collective heating effect of numerous small flares on the solar atmosphere. This flare model is implemented into the Bifrost framework, a 3D radiative magnetohydrodynamics (MHD) code with support for optional modules for including additional physics. By using the state of the atmosphere to derive the locations and properties of the beams, we minimise the number of free parameters in the model. There are four primary tasks that the flare model has to solve: determining where in the atmosphere reconnection is happening, finding the energies of accelerated electrons, tracing the magnetic field lines that the electrons will follow and finally simulating the evolution of the electrons as they move through the atmosphere. This thesis is mainly concerned with the latter two tasks. The former tasks are the topic of Helle Bakke's master's thesis (Bakke, 2018), which should be read in conjunction with this one. My aim is to use the flare model to answer the question of how the electron beams associated with nanoflares will influence the solar atmospheric plasma.

This thesis is structured as follows: Chapter 2 contains background theory relevant for understanding the work done in the project, particularly related to the solar atmosphere (Section 2.1) and solar flares (Section 2.2). The details of how the flare model was developed and implemented are then described throughout Chapter 3. Section 3.1 presents the Bifrost code and how the flare model is integrated into it. The tasks of detecting reconnection sites and computing electron energies are briefly discussed in Section 3.2. The subject of tracing magnetic field lines is the topic of Section 3.3, while Section 3.4 discusses how the processes governing the evolution of the electron beam are simulated. Chapter 4 then presents the results of the simulations, starting with a look at the initial atmosphere in Section 4.1, followed by the results of some tests for

deciding how many beams to simulate in Section 3.5. Section 4.2 then briefly presents the detected electron acceleration sites. The beam heating produced during a single time step is discussed in Section 4.3, and the effect of beam heating on the atmosphere is explored in Section 4.4. Finally, the thesis is summarised in Chapter 5, which also contains some concluding remarks. Some supplementary material is included in the appendices; a derivation of the interpolation algorithm used when tracing field lines can be found in Appendix A, and a description of how the simulation code is parallelised is given in Appendix B.

Chapter 2

Theory

2.1 The solar atmosphere

The solar atmosphere is the layer of plasma making up the outermost part of the Sun. It begins with the photosphere, the thin layer from which most of the Sun's visible light is emitted. Atop the photosphere lies the more tenuous chromosphere, where plasma is heated to higher temperatures, eventually reaching millions of kelvins on top of the *transition region*, a thin region where the plasma undergoes an extreme rise in temperature and fall in density. The transition region marks the beginning of the outermost layer of the atmosphere, the corona, where plasma is confined to move along the field lines of the Sun's magnetic field, which powers most of the activity in the solar atmosphere.

While it can be tempting from this description to picture the atmosphere as series of uniform layers on top of each other, this view is far from the truth. Figure 2.1 shows a snapshot from a 2D numerical simulation of a small, relatively calm region in the atmosphere, and we can see that there are significant variations with horizontal position as well as with height. Indeed, the real solar atmosphere is highly inhomogeneous and dynamic, with plasma being heated, cooled and accelerated at very different heights depending on the local conditions. Much of this inhomogeneity is due to the magnetic field. Generated by large-scale motions of plasma in the solar interior, the magnetic field emerges at the surface as loops and strands rising through the atmosphere. Tensions in the magnetic field are released through heating and acceleration of the atmospheric plasma, driving energetic events like solar flares.

The following sections discuss the photosphere, chromosphere and corona in more detail. Several of the important features of the atmosphere are apparent in Figure 2.1, so it can be useful to refer to the figure as a visual aid throughout the discussion.

2.1.1 The photosphere

The photosphere is the deepest part of the solar atmosphere. It sits on top of the convective region, the outer layer of the solar interior where most of the energy is transported outwards by convection. The photosphere begins where the plasma in the
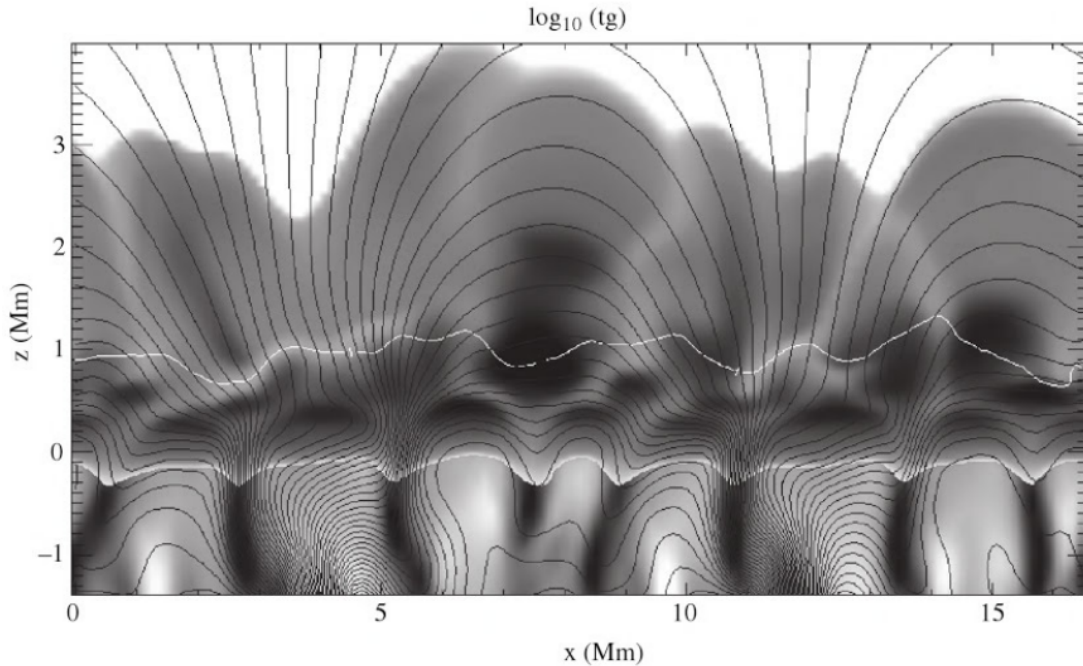
Figure 2.1: A 2D numerical simulation of the quiet solar atmosphere. Below the lower white curve (indicating the top of the photosphere, near $z = 0$ Mm), bright areas correspond to upward velocity and dark areas correspond to downward velocity. Above it, brighter areas indicate higher temperature. The white area in the top of the figure is the corona. Below the central white curve (near $z = 1$ Mm), the plasma β is larger than unity, while above the curve it is smaller. The black lines are magnetic field lines. From Carlsson and Hansteen (2005).

upper layers of the convective region becomes thin enough that most of the visible photons can escape the Sun without any further scattering. Hence, the majority of the radiation that we can observe from the Sun comes from the photosphere. In more precise terms, the photosphere is usually considered to begin where the *optical depth* is close to unity for visible light[1]. The intensity emitted at an optical depth τ will have been reduced by a factor of $e^{-\tau}$ by the time it leaves the solar atmosphere (ignoring the additional contribution to intensity from emission at greater heights).

The photosphere represents a transition from *local thermodynamic equilibrium* (LTE) to *non-local thermodynamic equilibrium* (non-LTE). Most of the plasma below the photosphere is in LTE, meaning that the radiation field everywhere only depends on the local conditions and is described by *Planck's law* (radiation described by Planck's law is known as *blackbody radiation*). This locality of the radiation field below the atmosphere is a result of the relatively high density. Photons emitted at some position will immediately interact with nearby plasma particles and have their energy converted into

---

[1]The one-dimensional VAL model of Vernazza et al. (1981) defines the bottom of the photosphere as the location where $\tau = 1$ for light with a wavelength of 500 nm.

heat, so they are unable to directly contribute to the radiation field farther away. However, as the density drops drastically throughout the atmosphere, this situation changes. Radiation emitted at one place can travel significant distances before interacting with the plasma, making the radiation field at each point in the atmosphere coupled to the conditions everywhere else. Hence, the plasma is in non-LTE. This global coupling significantly complicates any attempt of realistically modelling the solar atmosphere.

The visual appearance of the photosphere is dominated by small, short-lived bright areas separated by darker lanes. These features are called *granules*, and are typically a few thousand kilometres in diameter. They exist as a result of cells of hot plasma from the interior being convected out to the photosphere, where they appear as bright regions. The emerging plasma is pushed sideways by the underlying column of ascending gas, while at the same time radiating away much of its heat, cooling down and becoming denser. Eventually, the plasma becomes denser than its surroundings, and sinks back into the interior, completing the circular convection pattern. The lower part of Figure 2.1 shows convection cells below the photosphere giving rise to granules.

On larger scales, *sunspots* are often apparent. These are extended regions much darker than their surroundings, usually consisting of a central dark *umbra* surrounded by a somewhat brighter *penumbra*. While not actually dark by terrestrial standards, the sunspots have a significantly lower flux than the average photospheric plasma due to their relatively low temperature of around 3800 K, compared to the typical photospheric temperature of around 5800 K. The sunspots are cool because they are regions where the transport of energy from the upper interior to the surface is less effective than normal. This happens when convection is inhibited by the presence of a strong magnetic field directed perpendicularly to the surface. The magnetic forces prevent the plasma from moving across the magnetic field lines, shutting off the sideways motion required for maintaining the circular convection pattern.

The magnetic field lines emerging in the photosphere from the solar interior are mainly concentrated in bundles called *magnetic flux tubes*. The field lines spread out with height, making the magnetic field weaker higher up. When the plasma is highly conductive (which is the case for most of the solar atmosphere), the motion of the plasma and the magnetic field lines is always coupled (the magnetic field is said to be *frozen in* to the plasma). Plasma is prohibited from moving perpendicularly to the magnetic field direction, but it might pull the magnetic field lines along with it. Whether it is the plasma or the magnetic field that dictates the coupled motion is indicated by the *plasma β*, defined as the ratio of gas pressure to magnetic pressure. If $\beta > 1$, the gas pressure forces are stronger than the magnetic forces, and the magnetic field lines have to follow along with the motion of the plasma. This is the case in most of the photosphere, where the plasma is dense enough that it is relatively unaffected by the magnetic field. The convective motion of the plasma gives the magnetic field lines a tendency to come together in the lanes between granular cells, where the plasma flow converges (this is very apparent in Figure 2.1). However, the constant shifting and shuffling by convective motion makes the photospheric magnetic field quite complex and highly non-uniform.

2.1.2   The chromosphere

Because a large amount of energy is radiated away from the photospheric plasma, the temperature decreases on average as we move higher up from the bottom of the photosphere. The density and pressure also decrease, since the weight of the overlying atmosphere becomes smaller and easier to support. After a few hundred kilometres, the photosphere transitions into the chromosphere. In the lower part of the chromosphere, we reach a temperature minimum (of around 4300 K on average), above which the temperature starts to increase. Throughout the rest of the chromosphere the temperature increases with height, culminating in the thin transition region where the temperature increases from around $30\,000$ K to more than 10^6 K in just a few tens to hundreds of kilometres. This marks the interface between the chromosphere and the hot, tenuous corona.

We do not have a complete picture of how the chromosphere and corona are heated, but several different physical processes are believed to play a role (see Jess et al. (2015) for a more comprehensive review of these). They mainly fall into two categories: magnetic reconnection and *MHD waves*. Magnetic reconnection is a change in the topology of the magnetic field; separate magnetic field lines merging to form new field lines. When magnetic flux tubes get shuffled around by the convective motion in the photosphere, tensions are created in the magnetic field, and these are released through reconnection. During a reconnection event, magnetic energy is converted into heat and acceleration of the nearby plasma, which can lead to solar flares (the topic of Section 2.2). Local heating from reconnection within the chromosphere, as well as heating from flares originating higher up in the atmosphere, are thought to play a major role in the heating of the chromosphere. The other main heating mechanism, MHD waves, also has its origin in the churning motions on top of the convective region. The waves generated by these motions propagate outwards in the atmosphere and eventually dissipate their energy as heat.

In addition to the transition from low to high temperature, the chromosphere also contains the transition from high to low plasma β. As mentioned earlier, gas pressure controls the dynamics in most of the photosphere, so the plasma β there is high. Above the photosphere, the magnetic field tends to decrease in magnitude with height. However, the gas pressure decreases even faster, so the magnetic pressure eventually exceeds the gas pressure and magnetic forces start to dominate. The plasma is then confined to move along magnetic field lines.

The appearance of the upper chromosphere is dominated by hot plasma jets called *spicules*. There are typically a few hundred thousand spicules present throughout the solar chromosphere at any given time, and they reach up several thousand kilometres through the atmosphere. They are relatively short-lived, with typical lifetimes of a few minutes. The origin of spicules is still not well understood, but one possible creation mechanism is the acceleration of gas inside magnetic flux tubes by shocks powered by global oscillations of the solar surface (De Pontieu et al., 2004).

In regions of high magnetic activity, *prominences* may form in the chromosphere. They are loop-shaped concentrations of chromospheric gas extending high into the much

hotter and thinner corona, and can be observed as bright arcs protruding out from the solar limb[2]. When viewed against the more luminous solar disk rather than at the limb, they appear as dark bands across the surface, and are referred to as *filaments*. We do not yet have the whole picture of how prominences and filaments are created.

2.1.3 The corona

The corona is the upper part of the solar atmosphere. It begins where the chromospheric plasma reaches temperatures of several hundred thousand to millions of kelvins, in the transition region. As the temperature rises, the density also drops by many orders of magnitude, so the coronal plasma is extremely sparse. Although it can be tempting to imagine the transition region as a uniform, static layer lying at a given height above the photosphere, this is far from the truth. It is really the set of individual locations at different heights throughout the atmosphere where plasma happens to transition between chromospheric and coronal temperatures. Figure 2.1 shows how the transition region (the bottom of the white area in the top of the figure) occurs at very different heights depending on the local conditions.

The presence of the sharp temperature rise of the transition region does not reflect a sudden increase in the efficiency of the heating processes described in the previous subsection. In fact, the heat input in the chromosphere is substantially larger than that in the transition region. It is rather the result of an abrupt inability of the chromospheric plasma to radiate away its energy effectively. In the chromosphere, where the temperature is a few tens of thousands of kelvins, helium is only partially ionised, meaning that one of its two electrons is still bound to the helium nucleus. This electron can be excited into higher energy levels by collisions with nearby atoms, effectively absorbing heat from the plasma. When the electron eventually falls down to a lower energy level, a photon containing the excess energy is emitted, and may leave the plasma, transporting the energy away. However, as soon as the temperature is high enough that helium becomes fully ionised, this cooling process is no longer possible, and the plasma needs to be at a much higher temperature before radiative cooling again is able to balance the heating. Hence, the transition region forms wherever the chromospheric plasma reaches the helium ionisation temperature.

The corona predominately consists of *coronal holes* and *coronal loops*. Coronal holes are large regions where the coronal plasma has a relatively low temperature and density, so they emit less radiation than the hotter components of the corona. They are mainly located around the polar regions of the Sun. The magnetic field lines in coronal holes are mostly open, meaning that they extend out to infinity rather than connecting two points on the surface. Plasma flows along the open field lines away from the Sun, giving rise to the *solar wind*, the steady stream of charged particles permeating the solar system.

Coronal loops make up most of the corona. They consist of higher concentrations of hot plasma situated along arc-shaped magnetic field lines. These field lines connect magnetic regions of opposite polarity in the photosphere, referred to as the loop's *foot*

---

[2]The solar limb is the edge of the visible solar disk.

points. Plasma can be injected from the photosphere or chromosphere into a coronal loop foot point, before rising through the loop, being heated through the chromosphere and transition region to coronal temperatures. This process is called *chromospheric evaporation.* The reverse process, *chromospheric condensation*, happens when plasma is drained from the loop through a foot point and cools down through the transition region and chromosphere. Coronal loops come in a large range of sizes, with the longest ones having lengths of several hundred thousand kilometres.

2.2   Solar flares

A solar flare is a sudden brightening on the Sun's surface, with the main release of energy typically lasting in the order of minutes. The emitted radiation spans the entire electromagnetic spectrum, and is often accompanied by high-energy charged particles like protons and electrons. Flares occur with a wide range of energies. The largest flares release up to 10^{32} ergs of energy, while the smaller ones typically release around 10^{28} ergs.

The origin of flares is thought to be the release of magnetic energy through magnetic reconnection. Plasma is heated directly at the reconnection site, but the reconnection process also causes an acceleration of charged particles to very high energies. The high-energy particles travel along magnetic field lines through the atmosphere, until they reach a sufficiently dense region of plasma (typically the bottom of the transition region), where their energy is converted to heat through collisions. Heat is also conducted from the hot reconnection site down to the cooler chromosphere. The heated chromospheric plasma expands and moves upwards into the corona, in the process of chromospheric evaporation mentioned in the previous section. So a flare is often accompanied by heating at several different locations in the atmosphere, as well as bulk motions of heated plasma.

Flares are classified according to their peak emission of soft X-rays[3] as measured by the GOES satellite in orbit around the Earth. The weakest flares are given the class A, followed by B, C, M and finally X for the strongest flares. The strength of a flare within a class is denoted by a number between 1 and 9 (or possibly more than 9 for X-class flares).

Closely related to ordinary flares are the phenomena called microflares and nano-flares (Cargill, 2013). These are the type of flares that we aim to model in this thesis. Statistical studies of flares with a wide range of different energies indicate that these events could be governed by similar physical processes as ordinary flares, with the distinction mainly being one of scale (Christe et al., 2008; Crosby et al., 1993; Feldman et al., 1996). This view is supported by Testa et al. (2014), who studied IRIS (Interface Region Imaging Spectrograph) observations of nanoflares in the form of rapidly varying ultraviolet brightenings at coronal loop foot points. They found the observations to be consistent with results from 1D simulations of beams of high-energy electrons trav-

---

[3]The qualifiers "soft" and "hard" are typically used to denote ionising radiation in the lower and higher end of the energy spectrum, respectively.

elling along a coronal loop. However, the energy of a typical nanoflare is only in the order of 10^{24} ergs. This is too small to be directly observable by current instruments at all the wavelengths (particularly X-rays) relevant for probing their underlying physical processes, and as a result, the role of nanoflares in the solar atmosphere is still not well understood.

The purpose of the following discussion is to provide some background useful for understanding the flare model developed in the thesis by giving a basic grasp of the concept of a solar flare, working from the assumption that nanoflares behave like downscaled versions of ordinary flares. Hence, the following sections discuss solar flares in more detail, starting with the physical processes related to flare creation and propagation of high-energy particles, followed by an overview of their observable spectral, spatial and temporal characteristics. However, the reader should note that flares are complex phenomena that come in a large variety of types. This discussion is by no means a comprehensive treatment of the topic; it rather tries to touch upon some general ideas and features. I recommend the article of Shibata and Magara (2011) for a more detailed overview of solar flares.

2.2.1 Physical origins

It is now widely accepted that magnetic reconnection is a likely mechanism for powering solar flares. The process is hard to observe directly, and our understanding of the details of magnetic reconnection in three dimensions is still not complete. Nevertheless, we have numerous observations of phenomena suggestive of magnetic reconnection during the birth of a flare (see e.g. Shibata (1999)).

Magnetic reconnection is the annihilation of magnetic energy by the transition of the field to a configuration with lower energy. The magnetic energy is converted into thermal and kinetic energy. Reconnection involves a topological change of the field in which separate magnetic field lines become connected, hence the name. Two domains of oppositely directed magnetic fields are brought together by an inflow of plasma towards a *current sheet*, a thin region of enhanced electrical current (Figure 2.2). The magnetic
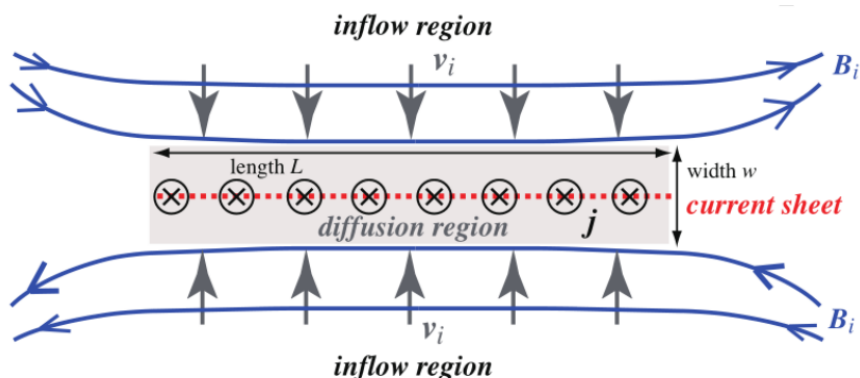

Figure 2.2: Simple model of magnetic reconnection in a current sheet. The encircled crosses denote magnetic null points. From Shibata and Magara (2011).

field vanishes at the interface between the two magnetic domains. A point where the magnetic field is zero is called a *magnetic null point*, and the set of null points making up the interface between the two magnetic domains is called a *separatrix surface*. Near the separatrix surface, there will naturally be a large gradient in the magnetic field. As the oppositely directed magnetic field lines are pushed together by the inflowing plasma, the gradient tends towards infinity. A larger magnetic gradient requires a larger electrical current in order to sustain it, but the finite resistivity in the current sheet prohibits the current from growing without limits. Instead, when the gradient becomes sufficiently large, the frozen-in condition mentioned in Section 2.1.1 breaks down near the separatrix surface, and *magnetic diffusion* sets in. Magnetic diffusion lets the two magnetic domains meet and cancel magnetic flux at the separatrix surface.

The electrical current driven by the reconnection process produces heat due to the resistivity in the current sheet, in a process known as *Ohmic heating* or *Joule heating*. At either end of the current sheet, the magnetic field lines tend to be strongly bent, and this produces a force called *magnetic tension*, which tries to straighten out the field lines. The magnetic tension accelerates plasma outwards along the magnetic field lines. These outflows of plasma (called *reconnection jets*) reduce the pressure inside the reconnection region, so more plasma will be pulled in, sustaining the reconnection process.

Part of the energy released by magnetic reconnection is responsible for accelerating nearby charged particles to very high speeds. Some observations suggest that nearly 50% of the released magnetic energy goes into accelerated particles (Emslie et al., 2005, 2004), although this number might vary significantly. The exact mechanism behind the acceleration process is not properly understood, but several candidates have been proposed. They can be broadly divided into three groups: acceleration by direct current (DC) electric fields, acceleration by shocks and stochastic acceleration by waves. A proper explanation of these processes would require an unreasonable amount of additional background and is thus outside the scope of this discussion. The interested reader can instead refer to e.g. Miller et al. (1997) or Zharkova et al. (2011) for detailed descriptions of particle acceleration mechanisms in flares.

Observations of hard X-rays from flares (see Section 2.2.3) suggest that the distribution of accelerated electrons resembles a power law in energy:

$$N(E \geq E_{\min}) \propto E^{-\delta}, \tag{2.1}$$

where N is the number of accelerated electrons with a given energy E, $E_{\min}$ is the lowest energy of any electron in the distribution and δ is a parameter describing how rapidly the number of electrons decreases with energy. Values for δ inferred from observations typically range from 3 to 7 (e.g. Leach (1984)). We will return to this distribution many times throughout the thesis.

There are many ways in which separate magnetic field domains can be forced together to initiate magnetic reconnection in the solar atmosphere. They typically involve magnetic flux tubes emerging from the solar interior by convection and expanding up into the pre-existing coronal magnetic field (Figure 2.3).
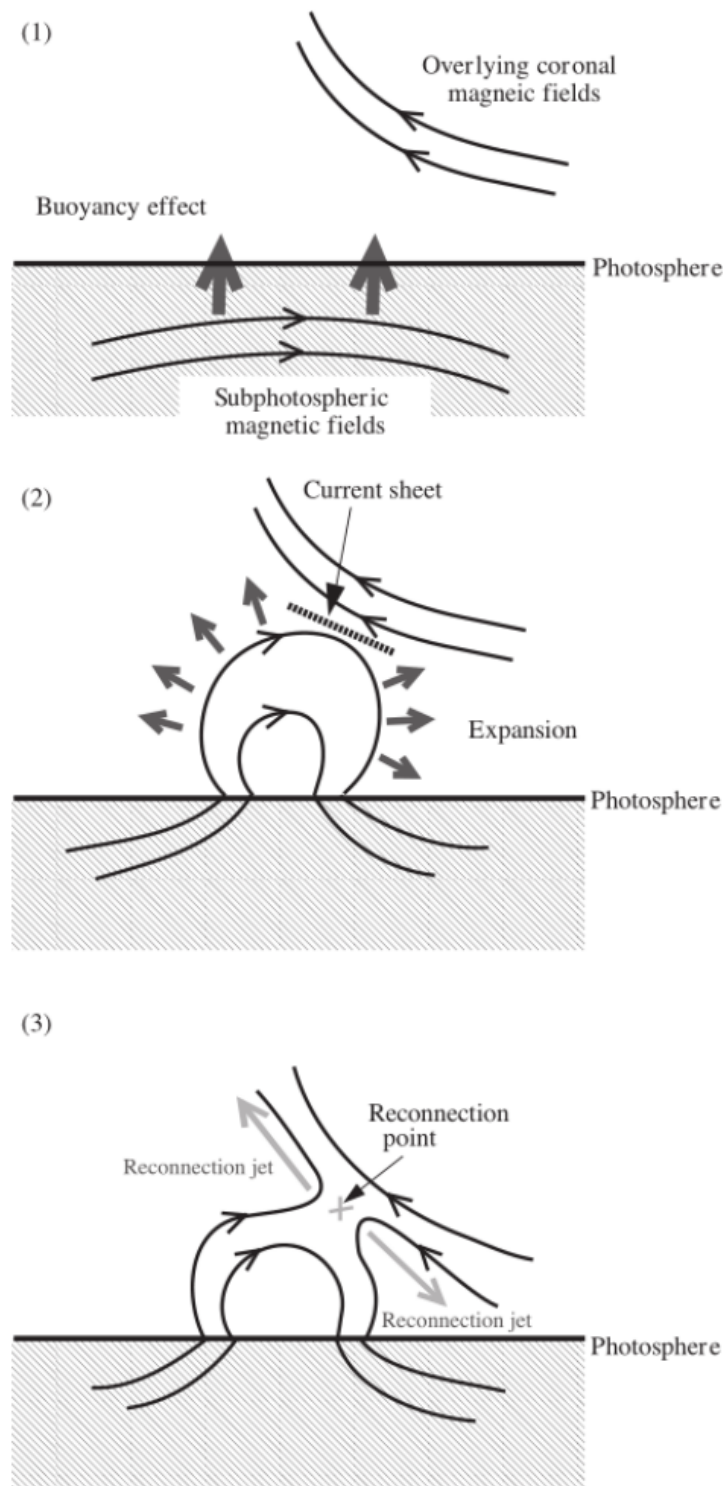
Figure 2.3: Magnetic field lines from the solar interior expanding into the atmosphere and reconnecting with overlying field lines. From Shibata and Magara (2011).

2.2.2 Transport of high-energy particles

The energy released by magnetic reconnection is transported away from the reconnection site by many different means, including heat conduction, high-energy particles, mass flow, waves and radiation. This discussion focusses on the transport of high-energy particles, partly because this is a complicated process which is central for understanding flares, but mainly because it is the basis for the numerical simulation code developed for this thesis.

Although high-energy ions (predominately protons) are believed to be produced at the acceleration sites of flares, this thesis is only concerned with accelerated electrons. The main reason for this is that electrons are typically accelerated to speeds greatly exceeding the average electron speed in the thermal plasma (the accelerated electrons are therefore referred to as *non-thermal*). This allows us to view the ambient plasma as a *cold target* with respect to accelerated electrons, meaning that the thermal speed of the ambient particles is negligible compared to the speed of the non-thermal electrons. This simplifies the theoretical description of the interaction between the accelerated particles and the ambient plasma. Because ions are thousands of times more massive than electrons, they will typically not be able to achieve non-thermal speeds with respect to the ambient electron gas, meaning that the cold target approximation fails and a more complicated description is required. Ions are indeed believed to be important for the energetics of many flares (Emslie et al., 2012), so a more accurate treatment of the beam heating process should include the effect of accelerated ions.

As the accelerated electrons travel through the atmosphere, they are affected by a number of different processes caused by the interactions between particles, the electromagnetic field, ambient plasma, radiation and waves. The following sections present the most important of these.

The Lorentz force

In general, a particle with electric charge q and velocity $\mathbf{v}$ will experience a force

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \tag{2.2}$$

due to the electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ at its location. This is called the *Lorentz force*. In a plasma, electric fields are generated everywhere by charged particles like electrons and ions. However, as the total number of positive and negative charges typically are very similar, plasmas tend to be electrically neutral on large scales. Since plasmas are highly conductive, any significant build-up of charge will quickly lead to electrical currents between oppositely charged regions, restoring neutrality. This cancellation of charges results in a damping of the average electric field, and is referred to as *Debye screening*. It is effective on scales exceeding the *Debye length*. On smaller scales, the charge is no longer uniformly distributed and hence strong electric fields may be present.

A consequence of the small electric fields in a plasma is that the magnetic term in equation (2.2) typically dominates. As the non-thermal electrons move through the plasma, the magnetic field exerts a force $\mathbf{F}$ on the electrons in the direction perpendicular

to their velocity $\mathbf{v}$ and the magnetic field $\mathbf{B}$. This confines the electrons to move along the magnetic field lines, as any transversal motion is transformed to a gyration around the direction of the magnetic field. The radius of this helical motion, or *gyroradius*, increases with the magnitude of the transversal speed.

Since the magnetic force always points perpendicularly to the direction of motion, it is unable change the kinetic energy of the electrons. But a stronger magnetic field will cause a larger fraction of the electron energy to go into the gyrating motion. Hence, if the magnetic field strength increases with depth, the electrons will have to gyrate more and more rapidly, meaning that the component of the velocity parallel to the magnetic field direction will have to decrease (or even switch direction). This effect is sometimes called *magnetic mirroring*.

Any electric field $\mathbf{E}$ that might be present can introduce two different modifications to the behaviour caused by the magnetic field alone. Firstly, the electric field component parallel to the magnetic field will lead to an acceleration of the electrons along the field lines[4]. Secondly, the electric field component perpendicular to the magnetic field will result in a drift of the centre of the helical motion away from the field line. This is sometimes referred to as $\mathbf{E} \times \mathbf{B}$ *drift*. We do not include these effects in our flare model, since the electric fields typically are very small. The minor differences in electron distributions and trajectories that would be obtained are negligible compared to the other uncertainties in the model.

Coulomb collisions

The non-thermal electrons interact with other particles (electrons, protons and heavier elements) due to their respective electric fields. The law describing the mutual forces between electrically charged particles is *Coulomb's law*, so the resulting interactions are called *Coulomb collisions*. The collisions will change the velocities the non-thermal electrons, and the amount of change depends on several factors, including the relative velocities of the colliding particles and the number density of the other particle species, as well as the scattering cross section for the interaction. The combined effect of Coulomb collisions on the distribution of electrons in the beam can be found by integrating the individual velocity changes over all collisions (Rosenbluth et al., 1957). This yields the main component of equation (3.44) in Section 3.4.1, which is used to simulate the electron beam. The actual derivation involves an integral of the scattering cross section over all scattering angles. This integral turns out to diverge in the limit of very small scattering angles due to the long-range nature of the Coulomb force between the particles. However, as mentioned previously, Debye screening causes the average electric field in a plasma to become negligible at scales exceeding the Debye length, so the long-range electrical forces will in reality vanish. The Debye length for the plasma can be written as

$$\lambda_{\mathrm{D}} = \sqrt{\frac{k_{\mathrm{B}}T}{4\pi n_{\mathrm{e}}e^2}}, \tag{2.3}$$

---

[4]This is what happens in the DC acceleration process mentioned in Section 2.2.1.

where k_B is the Boltzmann constant, T is the temperature, n_e is the number density of plasma electrons and e is the elementary charge. The divergence of the scattering integral is avoided by introducing a lower limit $\theta_{\min}$ for the scattering angle, taken to be $\theta_{\min} \approx 1/\Lambda$, where $\Lambda \equiv \lambda_D/r_c$. Here, r_c is the closest separation of the particles during the collision and is from energy considerations given (classically) by $r_c = \frac{1}{2} m_e v^2/e^2$, where m_e is the electron mass and v is the speed of the electron relative to the other particle (which is assumed here to be much heavier than the electron).

The quantity Λ indicates to which extent Coulomb collisions influence the electron velocities, and is central to computing the evolution of the electron beam. It is more commonly expressed in the form $\ln \Lambda$, and referred to as the *Coulomb logarithm*. Using the expressions for λ_D and r_c we get

$$\ln \Lambda \approx \frac{1}{2} \ln \left(\frac{e^2}{3\pi n_e m_e v^2} \right). \tag{2.4}$$

So the Coulomb logarithm depends on the number density of electrons in the plasma and the speed of the electron. However, using the Debye length as the screening distance for the electric force is not always appropriate[5], so the value of the Coulomb logarithm is subject to some uncertainty. A more suitable expression for the Coulomb logarithm for solar flare conditions, which is also valid for relativistic electrons, was derived by Ginzburg and Syrovatskii (1964). This form can be written as

$$\ln \Lambda = \frac{1}{2} \ln \left(\frac{(2\pi m_e c/h)^3}{\pi \alpha} \cdot \frac{(E(E+2))^2}{n_e} \right), \tag{2.5}$$

where c is the speed of light, h is the Planck constant, α is the fine structure constant and E is the total kinetic energy of the electron in units of the rest energy $m_e c^2$.

Note that the term "collision" for describing the interaction between a beam electron and the ambient plasma is somewhat misleading. At any instant the beam electron experiences a force from the combined electric field of the surrounding particles (within the Debye length), so the notion of a chain of separate binary collisions determining the motion of the beam electron is incorrect. However, it turns out that considering the average effect of a large number of collisions yields the same result as the more physically correct approach of considering the effect of the combined electric force (Leach, 1984).

The effect of Coulomb collisions is to transfer energy from the non-thermal electrons to the thermal plasma, as well as randomising the electron energies and directions. This is the dominating process affecting the evolution of the electron beam. Because the collisions transfer energy from the beam to the thermal plasma, the plasma is heated. And since the ambient electrons are by far the least massive particle species in the plasma, they are heated the fastest. The heavier species are then heated at a slower rate both from the collisions with the non-thermal electrons and from collisions with the ambient electrons that have already been heated.

In addition to heating the ambient plasma, direct collisions with non-thermal electrons can also excite or even ionise atoms in the plasma. This is mainly relevant in the

---

[5]In some cases the electron's mean free path or gyroradius could be a better choice.

chromosphere, where there is a significant fraction of neutral hydrogen and other partially ionised atoms. When free electrons are eventually recaptured by ions, in a process known as *recombination*, they emit radiation. Radiation emitted by recombination and de-excitations in the lower atmosphere can make up a significant fraction of the total radiative output in many flares.

Reverse currents

During a solar flare, a very large number of electrons are accelerated into the electron beam from a small region. The large flux of negative charge leaving the acceleration region gives rise to a charge separation at the reconnection site. In response to this charge separation, an electric field is generated, forcing ambient electrons to stream back towards the acceleration region. This *reverse current* thus counteracts the depletion of electrons at the acceleration site, ensuring that global charge neutrality is maintained. The reverse current will deposit energy in the plasma through Joule heating. The electric field driving the reverse current will also slow down the non-thermal electrons. If the non-thermal electrons have very high energies, the ambient electrons making up the reverse current will have to move very fast to ensure charge neutrality (unless the density is very high), and this can cause the reverse current to become unstable and significantly influence the evolution of the electron beam (Emslie, 1980). On the other hand, if the beam energy is relatively low, the reverse current plays a much less significant role.

Synchrotron emission

When the electrons gyrate around the magnetic field lines, they emit radiation due to their continuous acceleration. This radiation removes (a small amount of) energy from the electron beam. When the electrons are non-relativistic, they emit *cyclotron radiation* (typically microwaves if the electron energies are not too high) in the directions perpendicular to the magnetic field. Relativistic electrons emit *synchrotron radiation* (up to X-ray energies for high-energy electrons), which is beamed towards the direction of the magnetic field. The energy loss of the electrons to synchrotron radiation is typically smaller than 1% of the energy lost to Coulomb collisions (Kane, 1973).

Bremsstrahlung emission

When the beam electrons interact with ambient particles through Coulomb collisions, a small fraction of their energy loss is not transferred to the other particles, but is instead emitted in the form of radiation. This radiation is termed *bremsstrahlung*. The fraction of the energy converted to bremsstrahlung is usually in the order of 10^{-4} (Brown, 1971). Most of the bremsstrahlung produced by a typical flare consists of soft and hard X-rays.

Inverse Compton scattering

High-energy electrons moving through the plasma can interact with the photon gas, transferring energy to the photons. This effect is called *inverse Compton scattering*.

2.2.3 Spectral characteristics

As mentioned in the previous section, the non-thermal flare electrons emit soft and hard X-rays in the form of bremsstrahlung as they collide with the ambient plasma particles. The spectrum of emitted hard X-rays can typically be approximated as a power law[6] (Jeffrey, 2014). For the strongest flares, high-energy accelerated ions can collide with other particles and cause nuclear reactions that produce gamma-rays. The emission of hard X-rays and gamma-rays is an important diagnostic tool for flares, because such radiation can travel relatively unaffected through the upper solar atmosphere, giving a direct view of the conditions at the collision sites.

Chromospheric evaporation resulting from the energy release of the flare gives hot, rising chromospheric plasma emitting thermal radiation in the extreme ultraviolet, from which we can also observe blue-shifted versions of various chromospheric emission lines (e.g. Antonucci et al. (1985)). The hot plasma eventually fills the coronal loop associated with the flare. This loop is sometimes observable from soft X-rays emitted as thermal bremsstrahlung by the electrons in the plasma. Microwaves will also be emitted in the flaring loop, in the form of synchrotron radiation from the gyration of non-thermal electrons around the magnetic field lines. Enhanced emission of visible light, especially from hydrogen recombination as well as in the $H\alpha$ spectral line[7], is often observed deeper in the atmosphere below the loop foot points (see for instance Zarro et al. (1988)). This light tends to be red-shifted, indicating that the emitting plasma is moving downwards in the atmosphere.

2.2.4 Temporal behaviour

The temporal evolution of solar flares is usually divided into three phases; the *initial phase*, the *explosive phase* and the *decay phase*[8] (Syrovatskii, 1972). The initial phase can last for a few seconds to minutes, the impulsive phase is typically in the order of seconds, while the decay phase can last for several hours.

During the initial phase, chromospheric plasma at the loop foot points is quickly heated by the electron beam and the energy conducted from the hot reconnection site. The plasma reaches a thermal equilibrium at around 10^4 K, where radiative cooling by neutral hydrogen is quite efficient. However, collisions with non-thermal electrons gradually ionise the hydrogen atoms, eventually disabling the cooling. A new phase of

---

[6]Because of how the hard X-ray spectrum and the non-thermal electron distribution are related, the electron distribution can consequently also be approximated as a power law (but with a different power law index), as we saw in Section 2.2.1.

[7]$H\alpha$ radiation has a wavelength of 656.28 nm, and is emitted when the electron of a hydrogen atom is de-excited from the third to the second lowest energy level.

[8]These are also commonly referred to as the *preflare*, *impulsive* and *gradual* phases.

rapid heating follows, ending in a new equilibrium at 10^5 K, when radiative cooling by various metal ions becomes effective. There is some bulk motion of plasma due to the heating, but relatively little compared to what is to come in the explosive phase.

The explosive phase begins when the chromospheric plasma can no longer radiate away the incoming beam heating efficiently. When this happens, nearly all the energy deposited by non-thermal electrons will contribute directly to heating and acceleration of the plasma. Large amounts of plasma are heated to more than 10^6 K, and the resulting overpressure creates waves and accelerate hot plasma upwards into the flaring loop. Some plasma also moves downwards, and reaches lower temperatures as the radiative cooling becomes more effective due to the increased density in the wave.

After the explosive heating event, the remainder of the lifetime of the flare constitutes the decay phase. The release of energy typically continues into the decay phase (Svestka, 1989). Increased emission of soft X-rays from the coronal loop is common, and there may also be an increase of brightness in Hα at the loop foot points[9]. The remainder of the decay phase involves the gradual decrease of emission and bulk motion as the heated plasma cools and settles into equilibrium.

All three of the phases described above are not necessarily apparent for all flares. Very strong flares can reach the explosive phase so quickly after the onset of the flare that the initial phase is hardly detectable. On the other hand, very small flare events, in particular microflares and nanoflares, will not have enough energy to initiate the explosive event at all. For more detailed discussions of the physics driving the flare evolution, see e.g. Abbett and Hawley (1999) or Allred et al. (2005).

2.2.5   Spatial structure

As described in previous sections, chromospheric evaporation at the foot points during a flare fills the coronal loop with hot plasma. As a result, the full loop can be observed as a source of thermal soft X-rays. The soft X-ray emission is usually the strongest near the apex of the loop. The loop foot points themselves appear as strong sources of hard X-rays due to non-thermal bremsstrahlung from beam electrons. A somewhat weaker source of hard X-rays can sometimes be observed near the top of the loop (Figure 2.4). Discovered by Masuda et al. (1994), this feature is called a *loop-top source*, and is present in a small minority of flares. Initially it was not clear whether its hard X-ray emission was caused by non-thermal bremsstrahlung, like for the foot point sources, or if it was actually thermal radiation from extremely hot plasma. Further analysis by Alexander and Metcalf (1997) showed that the loop-top emission of the Masuda flare was non-thermal in nature during the impulsive phase, but exhibited a 5-minute decay consisting of thermal emission.

The hard X-ray foot point sources tend to be elliptical in shape, with the semi-major axis oriented along the magnetic field direction. Kontar et al. (2010) showed that the length of the semi-major axis decreases with X-ray energy, meaning that harder X-rays tend to be emitted from a narrower depth interval. This depth interval lies deeper in the

---

[9]This period with increasing emission is sometimes separately referred to as the *rise phase* of *flash phase*.
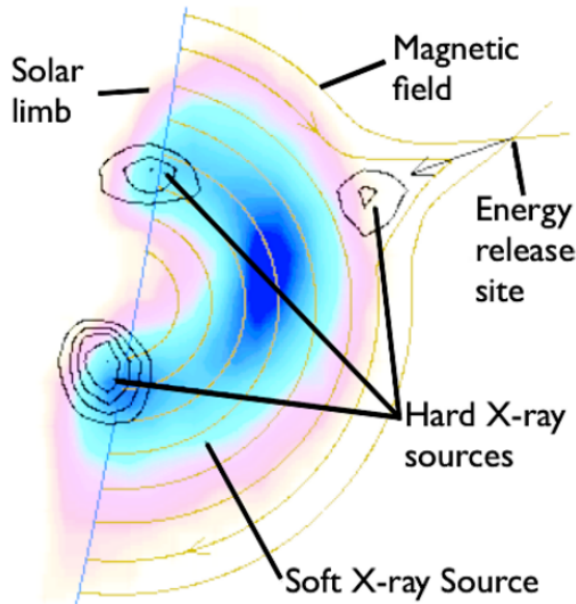
Figure 2.4: Soft and hard X-ray image of a flare (the Masuda flare of 13th January 1992) seen at the solar limb. The coronal loop is visible in soft X-rays, while hard X-rays are emitted from the two foot points as well as above the top of the loop. From Jeffrey (2014).

chromosphere for higher energies, consistent with the larger penetration depth expected from the high-energy electrons producing the X-rays. The perpendicular size of the foot point source was also shown to decrease with X-ray energy. This is attributed to the convergence of the magnetic field lines as one moves deeper down in the chromosphere.

Some flares exhibit very little foot point X-ray emission, with most of the emission coming from within the flare loop, in a so-called *thick target coronal source*. These flares have unusually high densities inside the loop, causing the non-thermal electrons to deposit their energy directly in the corona rather than propagating down to the chromosphere. Kontar et al. (2011) studied how the extent of thick target coronal sources varies with X-ray energy. In contrast to foot point sources, they found that the coronal sources increase in length with higher energy. This is consistent with the expected behaviour of non-thermal electrons colliding with a dense but relatively uniform region of plasma (the density in foot point sources, however, is far from uniform, hence the different dependence of source length on energy). More interestingly, they also found that the perpendicular extent of coronal sources increases with energy. As discussed earlier, the non-thermal electrons are prohibited from moving across the magnetic field lines (except for the possibility of a tiny $\mathbf{E} \times \mathbf{B}$ drift), so we would expect the perpendicular source extent to remain the same for all energies. The explanation proposed by Kontar et al. (2011) is the presence of *magnetic turbulence* within the flaring loop. Simply put, fluctuations in the small-scale magnetic field can introduce some perpendicular transport of charged particles with respect to the mean magnetic field.

Chapter 3

Methods

3.1 Simulating the solar atmosphere

3.1.1 The 3D MHD code Bifrost

The simulation of the flare process takes place inside a 3D box representing a region of the solar atmosphere, extending from the lower photosphere to the upper corona. The time evolution of the plasma, magnetic field and radiation inside this box is simulated using the Bifrost code (Gudiksen et al., 2011). Bifrost solves the MHD equations, coupled with the equations of radiative transfer, on a 3D Cartesian grid. The MHD equations is a set of eight partial differential equations, describing the time evolution of the mass density ρ, momentum $\rho\mathbf{u}$, magnetic field $\mathbf{B}$ and internal energy per unit volume e. They are given by

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho\mathbf{u} \tag{3.1}$$

$$\frac{\partial \rho\mathbf{u}}{\partial t} = -\nabla \cdot (\rho\mathbf{u}\mathbf{u} - \tau) - \nabla P + \mathbf{J} \times \mathbf{B} + \rho\mathbf{g} \tag{3.2}$$

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E} \tag{3.3}$$

$$\frac{\partial e}{\partial t} = -\nabla \cdot e\mathbf{u} - P\nabla \cdot \mathbf{u} + Q, \tag{3.4}$$

where τ is the viscous stress tensor, P is the gas pressure, $\mathbf{J}$ is the electric current density, $\mathbf{g}$ is the gravitational acceleration, $\mathbf{E}$ is the electric field and Q is a sum of contributions to the rate of heating of the plasma. In addition, we have the following two equations relating the secondary variables $\mathbf{J}$ and $\mathbf{E}$ to the primary variables $\mathbf{B}$ and $\mathbf{u}$:

$$\mu\mathbf{J} = \nabla \times \mathbf{B} \tag{3.5}$$

$$\mathbf{E} = \eta\mathbf{J} - \mathbf{u} \times \mathbf{B}. \tag{3.6}$$

Here μ is the permeability of vacuum and η is the magnetic diffusivity. Finally, an equation of state (EOS) relating P to ρ and e is needed to close the system of equations.

It should be chosen based on the required level of detail for the microphysics. Bifrost provides several options for specifying the EOS (see Gudiksen et al. (2011) for details).

Several different methods for computing radiative transfer can be used in a Bifrost simulation. They all provide the radiative losses for the source term Q in the energy equation (3.4). The different methods generally have different domains of applicability. For instance, the corona can be assumed to be very optically thin for most wavelengths, and in this case a relatively simple expression for the radiative losses can be used. The chromosphere, on the other hand, is a more challenging environment for computing radiative transfer, since it is generally neither optically thin nor thick. The chromospheric plasma is in non-LTE (Section 2.1.1), meaning that the local conditions are influenced significantly by the global radiation field. Here, Bifrost approximates the total radiative losses from the contributions of a hand-full of spectral lines that dominate the chromospheric radiative losses. The expression used for these contributions is given in Gudiksen et al. (2011). Bifrost also has a more detailed method for radiative transfer that solves the full radiative transfer equation under some simplifying assumptions. This is described in detail by Hayek et al. (2010).

The Bifrost code can integrate equations (3.1) – (3.4) forward in time using one of two possible explicit time stepping schemes; either a third-order Runge–Kutta scheme or a third-order Hyman scheme (Hyman, 1979). In order to keep the solution scheme stable, spatial gradients that are too strong must be smoothed out by adding artificial diffusion terms to the right-hand sides of the governing equations. Refer to Gudiksen et al. (2011) for more details on how these diffusion terms are computed. Spatial derivatives are computed using a sixth-order finite difference approximation. The resulting values for the derivatives will be shifted by half a grid cell relative to the quantities being differentiated. In principle, this would make it necessary to interpolate the derivatives so that they are defined at the same locations as the primary variables. However, by strategically defining the different variables to be slightly offset from the cell centra, the number of interpolations required for obtaining the variables at the correct locations can be significantly reduced. As a result, each primary variable in Bifrost is specified either in the cell centre, in the middle of a cell face or at a cell corner. Such a grid, where the different variables are not defined at the same locations, is often referred to as a *staggered grid*. When interpolations are required, they are computed by evaluating an interpolant fitted to the six closest surrounding grid points.

The Bifrost code is highly modular, making it easy to change which implementation to use for the various parts of the simulation process. The choice of EOS and time stepping scheme are examples of this, as well as the option to use various different boundary conditions and methods for radiative transfer. The modular nature of the code also makes it possible to implement additional physical phenomena separately from the main code. For instance, there is a module that implements the Spitzer model for conductivity, where the conductivity is computed based on the frequency of collisions between electrons and ions. There are also modules for computing the number densities of hydrogen and helium without having to assume LTE (which is what Bifrost does by default). Another example is the implementation of the generalised Ohm's law, where two additional terms are added to equation (3.6) (and thus to the induction equation

([3.3](#)) in order to better take into account that the plasma in general will not be fully ionised.

3.1.2 Integrating a flare simulation into Bifrost

With Bifrost providing the appropriate environment for flares to take place, a flare simulation code can be implemented as an additional module similar to the ones just mentioned. The code has the following main tasks:

1. Identify sites of reconnection. This is where electrons will be accelerated to non-thermal velocities.

2. Compute the distribution of accelerated electrons leaving each reconnection site.

3. Trace the magnetic field line passing through the acceleration region. This represents the path that the accelerated electrons will follow through the atmosphere.

4. Compute the thermal energy deposited by the electron beam in the plasma at each point along the field line. This heating is added to the source term Q in equation ([3.4](#)).

Task [1](#) is discussed in Section [3.2.1](#), task [2](#) is the subject of Section [3.2.2](#), and tasks [3](#) and [4](#) are described in Sections [3.3](#) and [3.4](#) respectively. As the transport and energy deposition of beam electrons is the primary focus of this thesis, the first two tasks, concerning the generation of electron beams, are only briefly outlined here. This aspect of the model was treated by Helle Bakke, whose master's thesis ([Bakke, 2018](#)) contains detailed discussions of all topics relevant to the beam generation part of the model.

3.2 Generating electron beams

3.2.1 Finding reconnection sites

The first step in generating the electron beams is to identify sites in the atmosphere where the magnetic field is reconnecting. In order to do this we follow [Biskamp](#) ([2005](#)) and compute the quantity

$$K(x, y, z) \equiv \|\mathbf{B}(x, y, z) \times (\nabla \times \mathbf{S}(x, y, z))\|, \tag{3.7}$$

at each point in the atmosphere. The vector $\mathbf{S}$ is defined as

$$\mathbf{S} \equiv E_\parallel \frac{\mathbf{B}}{\|\mathbf{B}\|}, \tag{3.8}$$

where $E_\parallel$ is the component of the electric field that is parallel to the magnetic field:

$$E_\parallel = \frac{\mathbf{E} \cdot \mathbf{B}}{\|\mathbf{B}\|}. \tag{3.9}$$

So we can express $\mathbf{S}$ as

$$\mathbf{S} = \left(\frac{\mathbf{E} \cdot \mathbf{B}}{\mathbf{B} \cdot \mathbf{B}} \right) \mathbf{B}. \tag{3.10}$$

Biskamp (2005) shows that there is reconnection wherever $K > 0$. In the numerical simulation there are hardly going to be any points where K evaluates exactly to zero, so for practical purposes it is better to use the criterion $K > K_{\text{lim}}$, where K_{lim} is chosen as small as possible while still not providing a larger number of reconnection sites than what is computationally manageable. However, one problem with this approach is that it produces a distribution of reconnection sites that strongly favours points low down in the atmosphere. Many of these points are not relevant with respect to energy transport by electron beams, because they lie in high-density regions where the beams are going to deposit all of their energy as soon as they are created. To resolve this, we can apply a horizontal normalisation to K, for example by dividing each K by the average of K over all points at the same height. Using discrete coordinates i, j and k for x, y and z, we can express this as

$$\tilde{K}_{i,j,k} \equiv \frac{K_{i,j,k}}{(N_x N_y)^{-1} \sum_i \sum_j K_{i,j,k}}, \tag{3.11}$$

where N_x and N_y are the total number of points in the x- and y-direction respectively. This expression was implemented in the code due to its computational efficiency, but other normalisation methods could also have been used. All points where $\tilde{K}_{i,j,k}$ exceeds some threshold $\tilde{K}_{\text{lim}}$ are then considered to be relevant reconnection sites.

3.2.2   Determining beam parameters

When the sites of electron acceleration have been located, the next step is to determine the properties of the resulting electron beams. We saw in Section 2.2.1 that the distribution of non-thermal electrons in a beam takes the form of a power law parametrised by the minimum energy E_{min} and the power law index δ. Additional properties of the beam are its total energy E_{tot}, as well as its mean energy $E_{\text{mean}} = E_{\text{tot}}/N_{\text{tot}}$, where N_{tot} is the total number of beam electrons. When the electron distribution is normalised to satisfy the properties of a probability distribution (i.e. that it must integrate to unity), it has the expression

$$f_{\text{PL}}(E) = (\delta - 1) E_{\text{min}}{}^{\delta - 1} E^{-\delta}. \tag{3.12}$$

Note that in order for this distribution to be valid, δ is required to be larger than 2. We can express the mean energy E_{mean} in terms of E_{min} and δ by solving the integral for the expected value of E:

$$E_{\text{mean}} = \int_{E_{\text{min}}}^{\infty} E f_{\text{PL}}(E) \, \mathrm{d}E. \tag{3.13}$$

Inserting equation (3.12) and evaluating the integral yields

$$E_{\text{mean}} = \left(\frac{\delta - 1}{\delta - 2} \right) E_{\text{min}}. \tag{3.14}$$

In order to determine E_{tot}, we assume that 50% of the reconnection energy is injected into the electron beam (see Section 2.2.1). By default, Bifrost deposits all the reconnection energy in the form of Joule heating Q_{Joule}. We therefore remove half the Joule heating at the reconnection site, and instead put it into the beam. The total beam energy can then be found as $E_{\text{tot}} = 0.5 \cdot Q_{\text{Joule}} \Delta t_{\text{rec}}$, where Δt_{rec} is the duration of the reconnection event. In practice, we never represent Δt_{rec} explicitly, but instead initialise and evolve new beams for each time step with durations corresponding to the duration Δt of the time step.

For the power law index δ, we simply set a constant value for all beams. In all the simulations performed for this thesis I used a value of $\delta = 4$, selected somewhat arbitrarily from the range mentioned in Section 2.2.1. A more precise determination of δ for individual beams would probably require a model of the acceleration process, which is outside the scope of this project.

All that remains is then to determine the minimum energy E_{min}. It represents a (rather artificial) limit separating the non-thermal electrons of the beam from the thermal electrons of the ambient plasma. The thermal electrons can be described collectively by a probability distribution known as the *Maxwell–Boltzmann distribution*. It has the form

$$f_{\text{MB}}(E) = \frac{2}{\sqrt{\pi}} \beta^{3/2} E^{1/2} e^{-\beta E}, \tag{3.15}$$

where $\beta \equiv 1/k_{\text{B}} T$. It is reasonable to approximate the full distribution of both thermal and non-thermal electrons at the reconnection site as a weighted sum of the thermal and non-thermal distributions:

$$f(E) = c_{\text{MB}} f_{\text{MB}}(E) + c_{\text{PL}} f_{\text{PL}}(E). \tag{3.16}$$

Because all probability distributions must integrate to unity, the two weights are subject to the constraint $c_{\text{MB}} + c_{\text{PL}} = 1$. We can impose another constraint by considering what the average energy of beam electrons would be in the framework of the combined distribution. Let us call this energy E^*_{mean}, to keep it separate from the average beam electron energy E_{mean} for the pure power-law distribution. Just like we did for the pure power-law distribution, we can find an analytical relation between E^*_{mean} and E_{min} by computing the integral

$$E^*_{\text{mean}} = \int_{E_{\text{min}}}^{\infty} E f(E) \, dE \approx \int_{E_{\text{min}}}^{\infty} E c_{\text{PL}} f_{\text{PL}}(E) \, dE, \tag{3.17}$$

where the approximate expression follows from the fact that the thermal component of $f(E)$ should become small for energies exceeding E_{min}. Solving the second integral yields the relation

$$c_{\text{PL}} = \left(\frac{\delta - 2}{\delta - 1} \right) \frac{E^*_{\text{mean}}}{E_{\text{min}}}. \tag{3.18}$$

The value of E^*_{mean} can be estimated from the conditions at the reconnection site, by assuming that N_{tot} corresponds roughly to the total number of electrons at the site. By inserting that estimate into equation (3.18) we can then express the coefficients c_{PL}

and c_{MB} purely in terms known quantities in addition to the unknown $E_{\min}$. Finally, we can estimate $E_{\min}$ as the intersection between the thermal component $c_{\mathrm{MB}}f_{\mathrm{MB}}(E)$ and the non-thermal component $c_{\mathrm{PL}}f_{\mathrm{PL}}(E)$ of the full electron distribution, i.e. finding $E_{\min}$ such that

$$c_{\mathrm{MB}}f_{\mathrm{MB}}(E_{\min}) = c_{\mathrm{PL}}f_{\mathrm{PL}}(E_{\min}). \tag{3.19}$$

In the code we solve this root-finding problem numerically using the Newton–Raphson method (see e.g. Press et al. (2007)).

Note that the value we compute for $E_{\min}$ will be closely tied to the local temperature. A high temperature leads to a broad Maxwell–Boltzmann distribution, so the intersection with the power law distribution will occur at a relatively high energy. The opposite is true for low temperatures: the thermal distribution will be narrow, and the intersection will occur at a low energy.

3.3 Tracing magnetic field lines

3.3.1 Interpolating the magnetic field

To accurately follow the magnetic field lines, we will have to trace the field direction in steps that are much smaller than the width of the grid cells. The magnetic field components are only known at the cell faces, so an interpolation must be performed to approximate them in the exact locations where they are needed.

Let us first consider how we can interpolate in one dimension. Given a set of discrete grid points x_i and corresponding function values $f(x_i)$, we can fit a polynomial P of order N to the function values of f at the points $x_k, x_{k+1}, \ldots, x_{k+N}$ and use $P(x)$ as an approximation to $f(x)$, given that $x_k \leq x \leq x_{k+N}$. There are a number of algorithms for evaluating the interpolating polynomial $P(x)$. In my implementation I have used Neville's algorithm, which is described in Appendix A.

To extend the interpolation scheme to two dimensions, i.e. approximating $f(x, y)$ when we also have a set of grid points y_i in the y-direction, we first apply the 1D scheme to find $P(x, y_i)$ for $y_k, y_{k+1}, \ldots, y_{k+N}$ (note that the value of k in general is different for x and y). We then use the 1D scheme again to interpolate $P(x, y_i)$ in the y-direction, and we get $P(x, y)$. For three dimensions we apply the same concept again. We have an additional set of grid points z_i in the z-direction, and find $P(x, y, z_i)$ as just described for the points $z_k, z_{k+1}, \ldots, z_{k+N}$. These function values are then interpolated in the z-direction to obtain $P(x, y, z)$.

The index k of the first grid point to use in the interpolation must be chosen so that $x \in [x_k, x_{k+N}]$ (otherwise the result would be an extrapolation rather than an interpolation). Assuming that this condition is met, there are still $N+1$ possible choices for k, representing how much the interpolating polynomial will be shifted relative to x. In other words, the choice of k will affect how much the interpolated value is influenced by the function values lying behind and ahead of x. If we let c denote an index such

that $x_c \leq x < x_{c+1}$, then a useful way of expressing k is

$$k = c + b + 1 - \left\lfloor \frac{N+1}{2} \right\rfloor, \tag{3.20}$$

where b is an integer describing the bias of the interpolation in the forward direction. The last term (note the integer division) makes sure that for a bias of zero, x lies as close as possible to the centre of the interpolation range. A positive bias will give a more forward-weighted interpolation, while a negative bias will give a more backward-weighted interpolation. In the implemented code, I use a bias of zero everywhere except near the non-periodic boundaries of the simulation box, where some of the neighbouring grid cells do not exist.

3.3.2 Stepping along the magnetic field

The problem of tracing a magnetic field line can be stated in terms of the following set of ordinary differential equations:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = \frac{\mathbf{B}(\mathbf{x})}{\|\mathbf{B}(\mathbf{x})\|}, \tag{3.21}$$

where $\mathbf{x} = [x, y, z]$ is the position of a point on the field line, s is the distance along the field line and $\mathbf{B}(\mathbf{x}) = [B_x(x, y, z), B_y(x, y, z), B_z(x, y, z)]$ is the magnetic field at the position $\mathbf{x}$. Equation (3.21) simply states that the tangent of the field line points in the direction of the magnetic field. A number of different methods have been devised for solving such equations numerically, all with different strengths and weaknesses. The schemes I have chosen for my implementation belong to the Runge–Kutta class of methods. The idea behind them is to sample the direction $\mathrm{d}\mathbf{x}/\mathrm{d}s$ at several points close to the current position, and use a weighted average of these directions as the actual direction to step in. Let $\mathbf{f}(\mathbf{x}) \equiv \mathrm{d}\mathbf{x}/\mathrm{d}s$[1] denote the direction at point $\mathbf{x}$, and let $\mathbf{x}_n$ be the current position for step n of the solution procedure. If we sample m different directions, the different samples are computed as

$$\begin{aligned}
\mathbf{k}_{n,1} &= \mathbf{f}(\mathbf{x}_n) \\
\mathbf{k}_{n,2} &= \mathbf{f}(\mathbf{x}_n + \Delta s_n(a_{21}\mathbf{k}_{n,1})) \\
\mathbf{k}_{n,3} &= \mathbf{f}(\mathbf{x}_n + \Delta s_n(a_{31}\mathbf{k}_{n,1} + a_{32}\mathbf{k}_{n,2})) \\
&\vdots \\
\mathbf{k}_{n,m} &= \mathbf{f}(\mathbf{x}_n + \Delta s_n(a_{m1}\mathbf{k}_{n,1} + a_{m2}\mathbf{k}_{n,2} + \ldots + a_{mm-1}\mathbf{k}_{n,m-1})),
\end{aligned} \tag{3.22}$$

where Δs_n is the size of the step we want to take and a_{ij} are weights that affect where the different directions will be sampled. We then average these direction samples with weights b_i to obtain an effective direction $\mathbf{d}_n$, given by

$$\mathbf{d}_n = b_1\mathbf{k}_{n,1} + b_2\mathbf{k}_{n,2} + \ldots + b_m\mathbf{k}_{n,m}, \tag{3.23}$$

---

[1] For a general set of ordinary differential equations, $\mathbf{f}$ can also be a function of s, in which case equation (3.22) also includes a set of coefficients c_i that describe how to vary s for the different $\mathbf{k}$s.

so that we can compute the next position as

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \Delta s_n \mathbf{d}_n. \tag{3.24}$$

The point $\mathbf{x}_{n+1}$ will not correspond precisely to an exact point $\tilde{\mathbf{x}}_{n+1}$ on the same field line, since the effective direction found from equation (3.23) is only an approximation to the true effective direction $\tilde{\mathbf{d}}_n$, defined by

$$\tilde{\mathbf{x}}_{n+1} = \mathbf{x}_n + \Delta s_n \tilde{\mathbf{d}}_n. \tag{3.25}$$

In fact, for some integer p we have

$$\tilde{\mathbf{d}}_n = \mathbf{d}_n + \mathcal{O}(\Delta s_n{}^p) \tag{3.26}$$

and thus from equation (3.25)

$$\tilde{\mathbf{x}}_{n+1} = \mathbf{x}_{n+1} + \mathcal{O}(\Delta s_n{}^{p+1}). \tag{3.27}$$

The exponent p is called the *order* of the method, and a higher order means that the error term will be smaller. It depends on the weights a_{ij} and b_i as well as the number of samples m that we use. In general a larger number of samples will allow us to better probe the curvature of $\mathbf{f}$, so higher order methods have to use more samples.

As I will discuss in Section 3.3.6, the solar magnetic field lines are often relatively smooth in the corona, but can become irregular and turbulent on the lower atmosphere where plasma motions move the field around. It is therefore necessary for the field line tracing algorithm to be able to effectively follow both smooth and irregular regions of a field line. Since the irregular regions will require significantly smaller step sizes than the smooth regions, we need a way of determining a suitable step size prior to each step. We can achieve this by comparing the error that we make in each step to a target error, and increase or decrease the step size based on how far below or above the target error we are.

Before we go into details on how to adjust the step size, we will consider how we can estimate the error made during a step. One way of doing this it to compare the results obtained by performing the step with two schemes of a different order. Indeed, a special class of methods called *adaptive* Runge–Kutta schemes have the advantage that a lower-order approximation to the effective direction $\tilde{\mathbf{d}}_n$ can be computed without sampling any additional directions. This lower-order approximation is given by

$$\mathbf{d}_n^* = b_1^* \mathbf{k}_{n,1} + b_2^* \mathbf{k}_{n,2} + \ldots + b_m^* \mathbf{k}_{n,m} + b_{m+1}^* \mathbf{k}_{n,m+1}, \tag{3.28}$$

where

$$\mathbf{k}_{n,m+1} \equiv \mathbf{f}(\mathbf{x}_{n+1}). \tag{3.29}$$

The corresponding approximation to the next position is then

$$\mathbf{x}_{n+1}^* = \mathbf{x}_n + \Delta s_n \mathbf{d}_n^*, \tag{3.30}$$

and for comparison with equations (3.26) and (3.27) we have

$$\tilde{\mathbf{d}}_n = \mathbf{d}_n^* + \mathcal{O}(\Delta s_n{}^{p-1}) \tag{3.31}$$

and

$$\tilde{\mathbf{x}}_{n+1} = \mathbf{x}_{n+1}^* + \mathcal{O}(\Delta s_n{}^p). \tag{3.32}$$

The coefficients b_i^* are specified by the scheme along with a_{ij} and b_i. From equation (3.29) it looks like we do have to sample an additional direction after all, but in practice this amounts to no extra work, because the quantity $\mathbf{f}(\mathbf{x}_{n+1})$ corresponds to the first sample direction $\mathbf{k}_{n+1,1}$ of the next step. So all we have to do is to replace the first expression in equation (3.22) with

$$\mathbf{k}_{n,1} = \mathbf{k}_{n-1,m+1}. \tag{3.33}$$

We can now compute an approximation $\boldsymbol{\delta}_n$ to the error term in equation (3.32) by pretending that $\mathbf{x}_{n+1}$ rather than $\tilde{\mathbf{x}}_{n+1}$ is the true solution, i.e.

$$\begin{aligned}
\boldsymbol{\delta}_n &= \mathbf{x}_{n+1} - \mathbf{x}_{n+1}^* \\
&= \Delta s_n(\mathbf{d}_n - \mathbf{d}_n^*).
\end{aligned} \tag{3.34}$$

Inserting equations (3.23) and (3.28) we get

$$\boldsymbol{\delta}_n = \Delta s_n(e_1\mathbf{k}_{n,1} + e_2\mathbf{k}_{n,2} + \ldots + e_m\mathbf{k}_{n,m} + e_{m+1}\mathbf{k}_{n,m+1}), \tag{3.35}$$

where $e_i \equiv b_i - b_i^*$ (with $b_{m+1} = 0$).

3.3.3 Adaptive step size control

To decide how much to adjust the step size for the next step, the error estimate obtained from equation (3.35) must be compared with a target error. We can define the target error vector as

$$\boldsymbol{\epsilon}_n = \epsilon_{\text{abs}} + \epsilon_{\text{rel}}\mathbf{S}_n, \tag{3.36}$$

where ϵ_{abs} is the tolerance for absolute error and ϵ_{rel} is the tolerance for relative error. $\mathbf{S}_n$ represents the scale of the current solution. Following Press et al. (2007), we will compute this scale as

$$\mathbf{S}_n = \max(|\mathbf{x}_n|, |\mathbf{x}_{n+1}|), \tag{3.37}$$

where the maximum is taken for each coordinate separately. By taking the maximum of two different positions we avoid instabilities that could occur if any of them has components very close to zero. For controlling the step size it is useful to introduce a factor E_n specifying the solution error relative to the target error:

$$E_n = \sqrt{\frac{1}{3}\left[\left(\frac{\delta_{n,x}}{\epsilon_{n,x}}\right)^2 + \left(\frac{\delta_{n,y}}{\epsilon_{n,y}}\right)^2 + \left(\frac{\delta_{n,z}}{\epsilon_{n,z}}\right)^2\right]} \tag{3.38}$$

This is simply the root-mean-square value of the components of $\boldsymbol{\delta}_n/\boldsymbol{\epsilon}_n$. If $E_n \approx 1$, the step error is roughly equal to the target error, so the step size is considered optimal. If $E_n < 1$ we have an unnecessarily small error, and the step size should be increased. On the other hand, if $E_n > 1$ the step error is too large, and we should decrease the step size. Since the error for the step was too large, we should reject the new position computed by the step and try it again with the decreased step size.

How should we compute the new step size from the error E_n and the current step size? We start by determining how the error scales with step size. From equations (3.27) and (3.32) it can be seen that $\boldsymbol{\delta}_n$ must scale as $\Delta s_n{}^p$ (since the $\Delta s_n{}^{p+1}$ term from equation (3.27) is dominated by the $\Delta s_n{}^p$ term from equation (3.32)). Since equation (3.38) does not affect this scaling, E_n must also scale as $\Delta s_n{}^p$. So, we know that the current step size Δs_n produced an error of E_n, and the next step size Δs_{n+1} will produce some error E_{n+1}. From the scaling of these errors with step size, we must have that

$$\frac{E_{n+1}}{E_n} \propto \left(\frac{\Delta s_{n+1}}{\Delta s_n}\right)^p. \tag{3.39}$$

We want the next step size Δs_{n+1} to produce an error of $E_{n+1} = 1$. Equation (3.39) then gives us the corresponding relation between the two step sizes:

$$\Delta s_{n+1} = \sigma E_n{}^{-1/p}\Delta s_n. \tag{3.40}$$

I have here included the proportionality constant σ. It is a safety factor that should be set to a value slightly below one (e.g. 0.9). The purpose of the safety factor is to bias the solution towards smaller errors to prevent oscillation of the error around $E_n = 1$ (which would lead to a lot of rejected steps). It is also a good idea to impose a lower and upper bound on the scale factor $\Delta s_{n+1}/\Delta s_n$, to avoid very abrupt changes in step size. In addition, we should not let the step size increase if the previous step was rejected, since this might lead to oscillation around $E_n = 1$.

Like other explicit integration schemes, the Runge–Kutta scheme presented in Section 3.3.2 can become unstable if the step size becomes too large. This can lead to large oscillations in the step size as the step size controller given by equation (3.40) tries to keep the error below the target error. Such oscillations are unfavourable because they result in many steps being rejected, which is computationally costly. To prevent this, it is possible to modify equation (3.40) in such a way that it tends to smooth out the step size sequence. Applying concepts from control theory, Gustafsson (1994) has suggested an alternative step size controller of the following form:

$$\Delta s_{n+1} = \sigma E_n{}^{-\alpha}E_{n-1}{}^{\beta}\Delta s_n. \tag{3.41}$$

It uses both the current and previous error for correcting the step size. Here, both α and β should be positive and proportional to $1/p$, but the actual values giving the best results will depend on the scheme and must be estimated by trial and error. Gustafsson found that a good place to start is $\alpha = 0.3/p$ and $\beta = 0.4/p$. This controller has the advantage that it is nearly as easy to implement as equation (3.40) (we just have to store the previous error in addition to the current one), while still providing the option

to stabilise the step size sequence. We can see that equation (3.41) reduces to equation (3.40) if we set $\alpha = 1/p$ and $\beta = 0$.

3.3.4 Interpolating output quantities

A consequence of using an adaptive step size when solving a differential equation is that we have no direct control over exactly where the solution is obtained, since the step size is continually modified throughout the solution process. The set of points where the traced field line is specified will not necessarily be very practical to use when computing the heating from the electron beam. It is better to be able to have the required output values be regularly spaced along the field line, or to be able to specify a set of distances along the field line where the output values should be produced. Suppose that we want to find the solution $\mathbf{x}_{\text{out}}$ at the distance s_{out} lying between two "naturally occurring" distances s_n and s_{n+1} where we know the solutions $\mathbf{x}_n$ and $\mathbf{x}_{n+1}$. This is achieved by performing an interpolation between s_n and s_{n+1}. Ideally, we would like to make due without having to perform any additional evaluations of $\mathbf{f}(\mathbf{x})$, since these evaluations are relatively expensive. If we only knew the values $\mathbf{x}_n$ and $\mathbf{x}_{n+1}$ at these distances, the best we could hope to do would then be a linear interpolation, i.e. only first-order accuracy. Fortunately, we also know the derivatives $\mathbf{f}(\mathbf{x}_n)$ and $\mathbf{f}(\mathbf{x}_{n+1})$, since they were used in the stepping scheme. This actually allows us to perform a third-order interpolation[2]. Indeed, even higher-order interpolations are possible in some schemes (e.g. the Dormand–Prince scheme mentioned in Section 3.3.5) by also making use of the intermediate derivatives $\mathbf{k}_{n,j}$ from equation (3.22). Note that only the positions $\mathbf{x}$ at the output location can be obtained in this way. Any other quantities that are also needed for the heating computations, like the density, must be interpolated in the ordinary way as described in Section 3.3.1.

3.3.5 Implemented schemes

I chose to implement two different adaptive Runge–Kutta schemes; a third-order method developed by P. Bogacki and L. F. Shampine, and a fifth-order method developed by J. R. Dormand and P. J. Prince. The corresponding weights a_{ij}, b_i and b_i^* can be found in Bogacki and Shampine (1989), and Dormand and Prince (1980) respectively. For the Bogacki–Shampine scheme I use a third-order Hermite interpolation for the output positions, while for the Dormand–Prince scheme I use a fourth-order interpolation described in Shampine (1986).

3.3.6 Terminating the tracing process

In order to ensure the stability and efficiency of the tracing algorithm, it is important to consider when and how the tracing procedure should be terminated. For instance, since a field line cannot be traced beyond the borders of the simulation box, the tracing

---

[2]Such an interpolation, where the interpolating polynomial matches both the solution values and their derivatives, is known as an *Hermite interpolation*.

should be stopped immediately if a non-periodic border is reached. Typically, this will be the bottom of the box, since the field lines originate deep below the solar atmosphere, but sometimes they will also extend through the top of the box. In Bifrost, periodic boundary conditions are often applied to the borders in x and y, so that the field lines will simply pass through one side and emerge from the opposite side.

The magnetic field lines tend to be relatively smooth in the solar corona. As described in Section 2.1.1, this is because the dynamics of the plasma typically is dominated by magnetic forces rather than gas pressure forces in this region, and the magnetic forces tend to smooth out the magnetic field as much as possible. Farther down, in the lower chromosphere, photosphere and beyond, the situation is the other way around. The plasma is much denser and the gas pressure forces are much stronger than the magnetic forces. The magnetic field lines are then frozen in to the plasma, and get pulled along as the plasma moves. Convection drives the overall mass flow, and as the
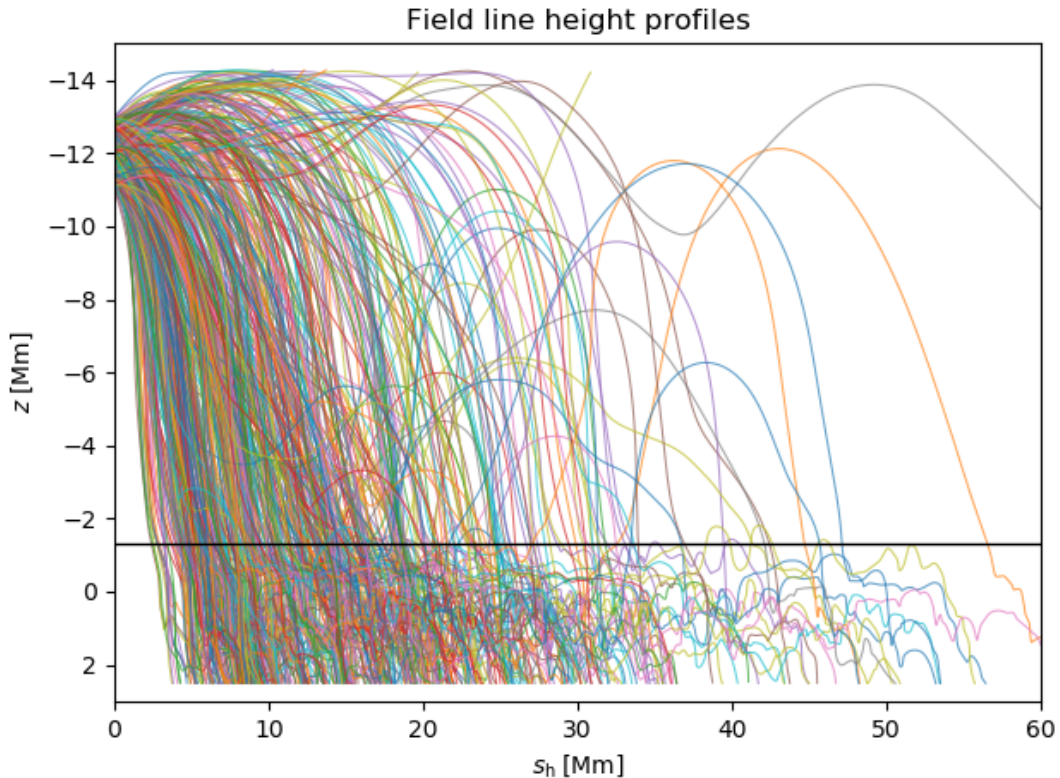
Figure 3.1: Height profiles for 100 field lines from a Bifrost simulation. The field lines were traced from random starting positions in the upper corona. The quantity along the x-axis is the integrated horizontal distance from the starting position, i.e. $s_\mathrm{h} \equiv \int \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$. z is defined to be zero at the top of the photosphere, and increases with depth. The horizontal line indicates the depth at which the plasma β tends to be smallest (see discussion in the main text).

plasma gets shuffled around by convective motions the magnetic field structure becomes turbulent and irregular.

Figure 3.1 gives an example of how the magnetic field lines change their behaviour with depth. As the figure shows, the field lines often "crawl" around a significant distance close to the bottom of the box before vanishing through the bottom. For the purpose of calculating the energy deposition from an electron beam along the field line, including the entire irregular part serves no useful purpose, because the relatively high density below the transition region will ensure that all the energy is deposited over a relatively small distance. The irregular parts of the field lines are relatively expensive to trace, since they require small step sizes. Field line tracing actually turns out to be the most computationally expensive part of the beam transport code, so we can save a significant amount of execution time by terminating the tracing procedure before the field line actually hits the bottom of the box.

A reasonable indicator of where the magnetic field will start to become irregular is the plasma β (briefly mentioned in Section 2.1.1), defined as

$$\beta \equiv \frac{P_\mathrm{g}}{P_B}, \tag{3.42}$$

where P_g is the gas pressure and $P_B = \|\mathbf{B}\|^2 / 8\pi$ (in cgs units) is the magnetic pressure. The plasma β thus measures to which extent the dynamics of the plasma is dominated by gas pressure forces rather than magnetic forces. Figure 3.2 shows how β varies with depth for a selection of different field lines. High in the corona, β tends to be in the order of unity. As the field lines descend towards the lower atmosphere, β decreases gradually, and reaches a minimum in the order of 10^{-2} in the lower chromosphere at a height of about 1–2 Mm above the photosphere. Farther down, β increases rapidly as the gas pressure forces begin to dominate. The height below which a field line is likely to become turbulent can thus be found by monitoring where the plasma β exceeds a certain value (probably of order unity). However, since β can sometimes reach relatively high values in the corona, this should only apply below the height of 1–2 Mm where β tends to be minimal.

As mentioned above, the high density of the lower atmosphere should prevent the electron beam from moving very far at large depths. Figure 3.3 shows the mass densities at the points where the field lines reach the β limit (i.e. the points where they enter the rectangle in Figure 3.2). Some field lines are actually terminated relatively high up at somewhat low densities. This means that the termination point might not be deep enough for all the electron beam energy to be deposited before the field line is terminated. A reasonable way to solve this issue is to extend the field line artificially. The extension should be as simple as possible in order to maintain efficiency, but it should also preserve the way the beam energy is deposited with depth. The simplest solution is to extend the field lines straight down until they reach the bottom of the simulation box. Although we know that the field lines also move horizontally in the lower atmosphere, there is no a priori reason for the field lines to prefer one horizontal direction over others, so on average we can expect the bulk of the energy deposition to occur directly below the points where the field lines enter the lower atmosphere.
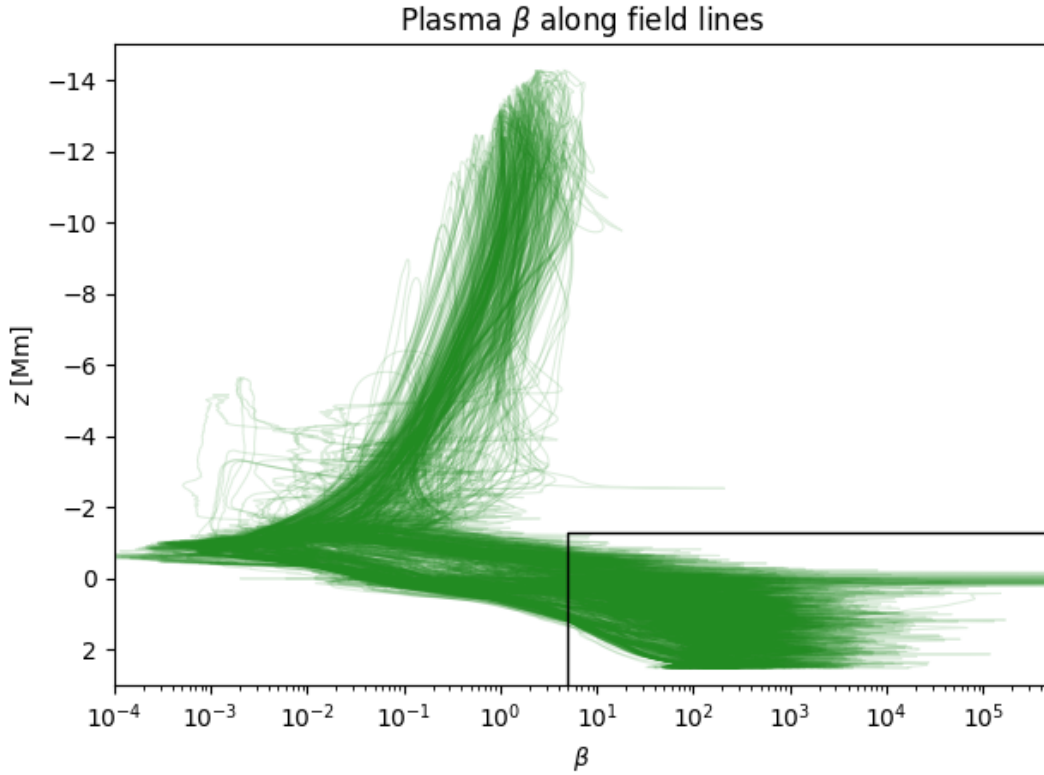
Figure 3.2: Plasma β along 300 field lines from the same Bifrost simulation as for Figure 3.1. The starting positions were sampled from the same region as in Figure 3.1. The horizontal line segment indicates the approximate depth where most of the β values are small. The vertical line segment indicates the upper limit for β that was used for deciding when the field lines should be terminated.

However, we still cannot completely neglect the horizontal components of the field lines. Since the field lines on average move through the lowest parts of the simulation box at an angle with respect to the vertical direction, the electrons will have to move through more plasma to get to a certain depth than if they were travelling straight down. So we must include the effect of this extra distance in order to avoid overestimating the penetration depth of the electron beam. By measuring distances from the termination points to the bottom for a large number of field lines, I found that the total distance on average is about twice as large as the vertical distance. So to obtain the correct distribution of energy deposition with depth when we extend the field lines straight down, we must effectively double the rate of collisions between the non-thermal electrons and the plasma particles. To avoid a discontinuity in the energy deposition at the point where the vertical field line extension begins, the transition from the field line to the vertical extension must be done smoothly. The simplest way of doing this is to have a small region after the termination point where the direction is found by a
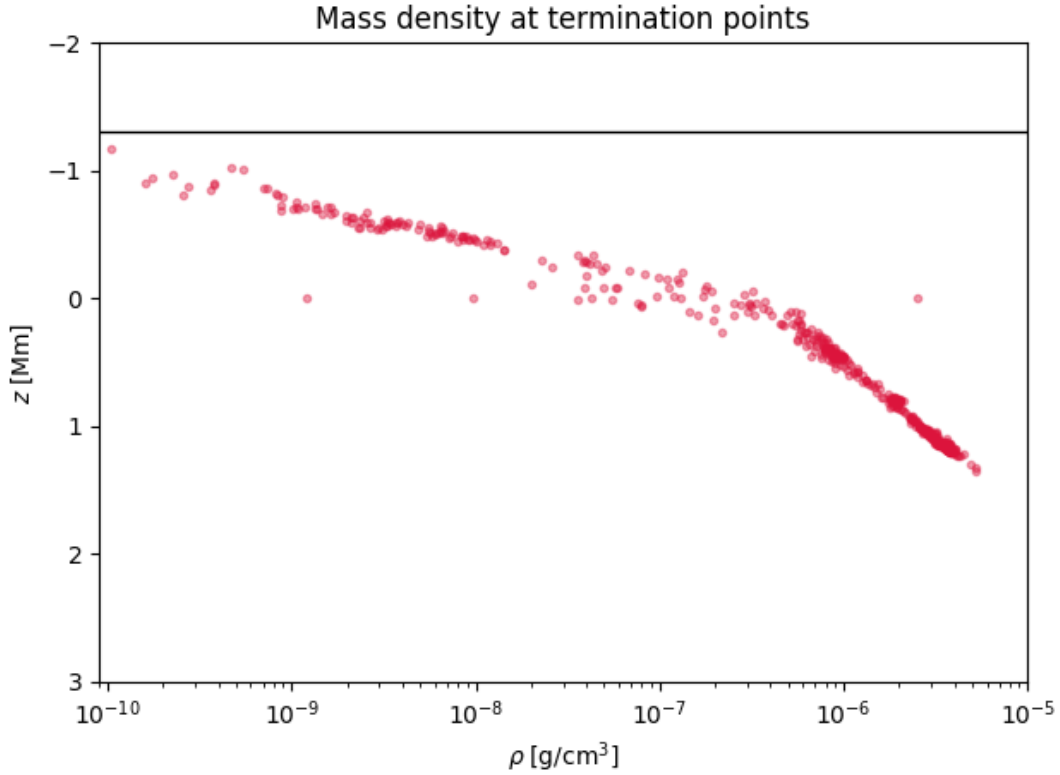
Figure 3.3: Mass densities at the points where the field lines in Figure 3.2 enter the rectangle and are terminated. The horizontal line corresponds to the same horizontal line as in Figures 3.1 and 3.2.

linear interpolation between the field line direction and the vertical direction. By also interpolating the distance correction factor in this region we ensure that the transition occurs smoothly.

We can see the effect of the artificial field line extensions in Figure 3.4. The two panels show the column masses along the field lines from the starting positions to the bottom, for both the full-length field lines and when they have been terminated and extended as described above. The column masses indicate how much the electron beam will have been obstructed at a given depth. Although the full-length field lines do have larger variations in their final column masses, the extended field lines still have the same average behaviour. Keeping in mind that the electron beams will never penetrate far beyond the transition region, the approximations discussed in this section appear to be quite reasonable.

Because our purpose for tracing field lines is to simulate electron beams moving along them, we can apply one more criterion for terminating the tracing process. The electrons lose energy as they move along the field line, and eventually the beam will contain so little energy that there is no point in transporting it further. So we can check
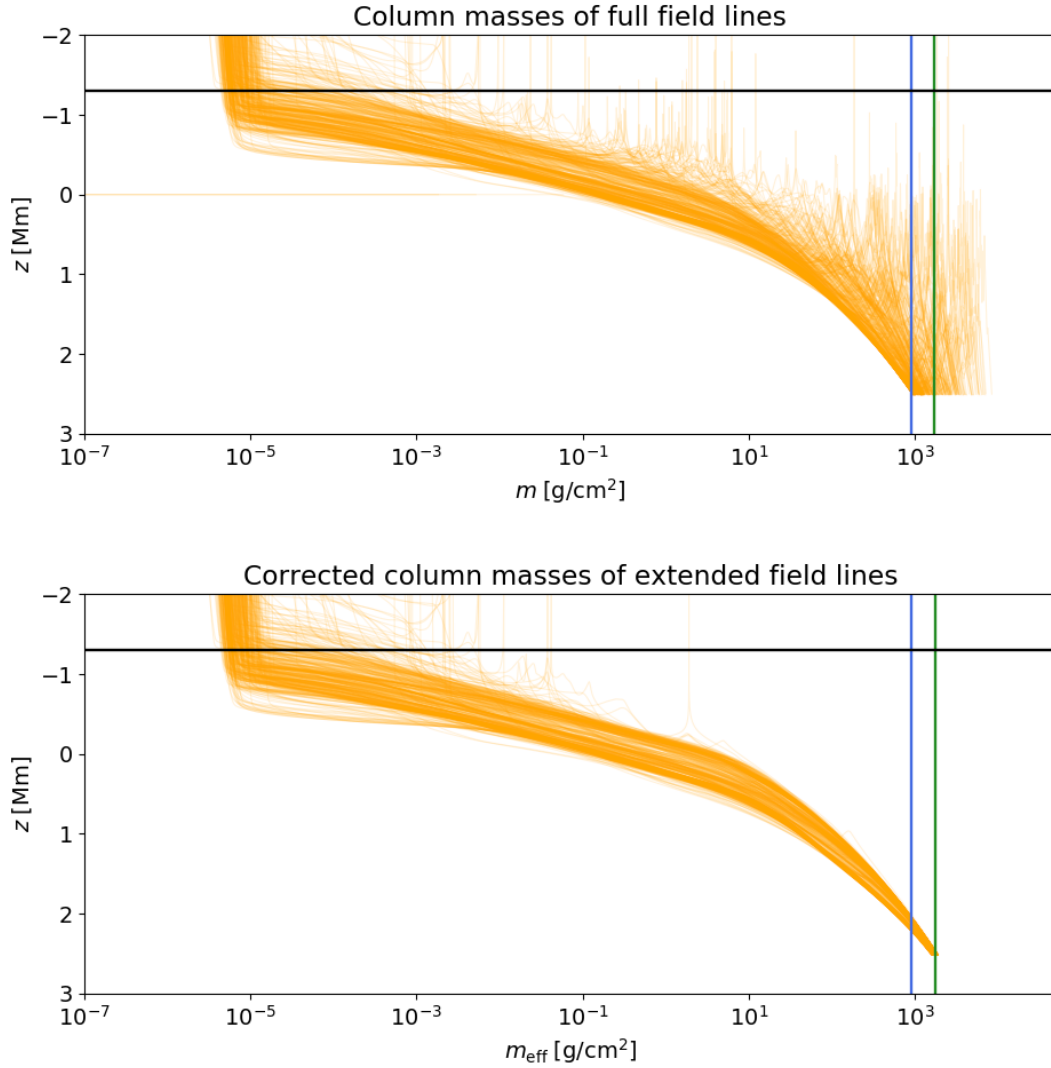
Figure 3.4: Column masses, i.e. mass density integrated from the starting positions along the length of the field lines, for the same field lines as in the previous figures. In the upper panel the field lines were not terminated until they hit the bottom of the box. In the lower panel they were terminated as described in the main text, and extended vertically downwards, with the distance correction factor included in the computation of column mass. The vertical lines indicate average column masses at the bottom of the simulation box. The blue line is the same in both panels and indicates what would be obtained for the extended field lines without the distance correction factor. The green lines show what is actually obtained for the field lines in the respective panels, and they make it clear that the average final column mass for the extended field lines (with the correction applied) is nearly the same as for the full field lines. The horizontal line is the same as in the previous figures.

the remaining energy of the electron beam at each depth, and when it drops below a certain value we stop the tracing procedure and simply dump the little energy that remains at the current position. Of course, this only works if the tracing and the beam transport calculations are performed in tandem, so that is what we do in our code.

3.3.7 Deciding which direction to trace field lines

During a reconnection event, electrons are accelerated to non-thermal energies by the electric field. The direction the electrons are allowed to propagate in must be either parallel or anti-parallel with the magnetic field direction (the only difference between the two cases is that the electrons will gyrate in opposite directions around the field line). So if we consider the direction of the electric field with respect to the magnetic field, we can decide whether the electron beam will travel in the forward or the backward direction along the field line. In case there are significant variations in the electric field direction around the reconnection site, it is more stable to consider the average direction in a small region around the site rather than at a single point. If the overall electric field around the reconnection site is roughly parallel to the magnetic field direction, we assume that the majority of the accelerated electrons will be forced to move backwards along the field line (because negatively charged particles are accelerated in the direction opposite to that of the electric field), and correspondingly in the opposite case. Finally, there are the cases where the electric field is more or less perpendicular to the magnetic field. In these cases the ability of the electric field to accelerate the electrons along the field line will be strongly reduced, so we would not expect to see significant beam energies in such a situation. In the code we therefore discard these cases from the set of relevant reconnection events. This is also in accordance with our criterion from Section 3.2.1 for detecting reconnection sites. Recall that the quantity $\mathbf{S}$ in the definition of K (equation (3.7)) is given by

$$\mathbf{S} = \left(\frac{\mathbf{E} \cdot \mathbf{B}}{\mathbf{B} \cdot \mathbf{B}}\right) \mathbf{B}. \tag{3.43}$$

If $\mathbf{E}$ and $\mathbf{B}$ are close to perpendicular, then $\mathbf{E} \cdot \mathbf{B} \approx 0$, so K will tend to be small and the criterion will probably not be satisfied.

3.4 Simulating the electron distribution

3.4.1 Transporting the electron distribution

The processes governing electron transport along a field line are discussed in Section 2.2.2. In order to make the problem of electron transport and energy deposition more manageable, we can make some simplifying assumptions:

- The typical time scale for the traversal of a field line by non-thermal electrons is in the order of a few tenths of a second, which is much smaller than the typical time scales for magnetohydrodynamic changes in the plasma. So we do not have to consider the time evolution of the electron beam: we can assume that the electrons

deposit their energy instantaneously and that their distribution has reached a steady state (Leach and Petrosian, 1981).

- The electron gyroradius (which is typically the order of 10 cm) is much smaller than MHD length scales, so the variation of the distribution perpendicularly to the field lines can be ignored. Only the one-dimensional distance along the field line has to be considered.

- Hydrogen dominates the atmospheric composition, and will thus be the element that has the strongest influence on the electron beam through Coulomb collisions. The $\sim 10\%$ fraction of Helium might add 20 to 30% to the effect of collisions (Leach and Petrosian, 1981), but this is not very significant considering the uncertainties in the parameters. So we can neglect collisions with other species than hydrogen. Interactions between the non-thermal electrons themselves will be so rare that they can be neglected. However, collisions with ambient electrons, protons and neutral hydrogen are all potentially important[3]. Energy exchange between beam electrons and ambient free electrons is the most effective process, because the low mass of electrons compared to protons and neutral hydrogen causes them to be heated much faster. The ambient electrons will then transfer their excess heat to the other plasma species over a larger time scale.

- The enhanced rates of excitation and ionisation in the plasma caused by the colliding non-thermal electrons are ignored. While ionisation from the electron beam is a central mechanism for initiating the explosive phase of flares (see Section 2.2.4), we do not expect to achieve high enough beam energies for this to happen in the simulations performed for this thesis.

- Reverse currents do not significantly affect the electron beam as long as the beam energy is small. So we neglect reverse currents. The same is true for inverse Compton scattering.

- The energy losses of the non-thermal electrons from the emission of synchrotron radiation and bremsstrahlung are typically very small compared to the losses from Coulomb collisions, so they can be ignored.

The probability distribution of the non-thermal electrons is governed by a type of partial differential equation from statistical mechanics known as a Fokker–Planck equation. From the discussions above, the contributing processes that we have to consider are Coulomb collisions and magnetic mirroring. The resulting Fokker–Planck equation is

---

[3]Although it does turn out that the electrons will be thermalised at relatively high ambient temperatures, so the fraction of neutrals will be very small.

given by Leach (1984) as

$$\mu\frac{\partial f(E,\mu,s)}{\partial s} = \frac{1-\mu^2}{2}\cdot\frac{\mathrm{d}\ln B}{\mathrm{d}s}\cdot\frac{\partial f(E,\mu,s)}{\partial\mu}+$$
$$\frac{2}{\beta}\cdot\frac{\partial}{\partial E}\left(\frac{C_1}{\beta}\cdot f(E,\mu,s)\right)+ \tag{3.44}$$
$$\frac{C_2}{\beta^4(E+1)^2}\cdot\frac{\partial}{\partial\mu}\left((1-\mu^2)\frac{\partial f(E,\mu,s)}{\partial\mu}\right).$$

The quantity μ is the *pitch angle*; the cosine of the angle between the electron velocity vector and the magnetic field direction. In this equation, β does not represent the plasma β, but rather the speed of the electron relative to the speed of light. It can be written as $\beta = \sqrt{1-(E+1)^{-2}}$. Like in Section 2.2.2, the energy E is in units of the electron rest energy. The first term on the right-hand side describes the influence of a gradient in the magnetic field strength on the electron pitch angles (magnetic mirroring). The second term describes the energy loss from Coulomb collisions, while the third term describes the diffusion of pitch angles from Coulomb collisions. The two collision coefficients are

$$C_1 = 2\pi r_0^2\left(n_{\mathrm{e}}\ln\Lambda + n_{\mathrm{H}}\ln\Lambda'_{\mathrm{H}}\right) \tag{3.45}$$
$$C_2 = 2\pi r_0^2\left(n_{\mathrm{e}}\ln\Lambda + n_{\mathrm{p}}\ln\Lambda + n_{\mathrm{H}}\ln\Lambda''_{\mathrm{H}}\right). \tag{3.46}$$

where n_{p} and n_{H} are the number densities of protons and neutral hydrogen atoms, respectively. $\ln\Lambda'_{\mathrm{H}}$ and $\ln\Lambda''_{\mathrm{H}}$ are the effective Coulomb logarithms for hydrogen, and are computed according to

$$(\Lambda'_{\mathrm{H}})^2 = E^2(E+2)/I_{\mathrm{H}}^2 \tag{3.47}$$
$$(\Lambda''_{\mathrm{H}})^2 = E(E+2)/2\alpha^2, \tag{3.48}$$

where I_{H} is the ionisation energy for hydrogen.

If we assume an electron distribution with no electrons moving backwards (i.e. $\mu < 0$), but which is otherwise isotropic, it can be written on the form

$$f(E,\mu,s) = \begin{cases} g(E,s) & \text{if } \mu \geq 0 \\ 0 & \text{if } \mu < 0 \end{cases}. \tag{3.49}$$

We then have

$$\frac{\partial f(E,\mu,s)}{\partial\mu} = 0. \tag{3.50}$$

Inserting equation (3.49) into equation (3.44) and using equation (3.50) we get

$$\mu\frac{\partial g(E,s)}{\partial s} = \frac{2}{\beta}\cdot\frac{\partial}{\partial E}\left(\frac{C_1}{\beta}\cdot g(E,s)\right). \tag{3.51}$$

We can integrate over all pitch angles from 0 to 1 on both sides to get rid of the μ factor on the left-hand side. This gives

$$\frac{\partial g(E,s)}{\partial s} = \frac{4}{\beta}\cdot\frac{\partial}{\partial E}\left(\frac{C_1}{\beta}\cdot g(E,s)\right). \tag{3.52}$$

A different simplification is to assume that all the electrons move straight forward, so $\mu = 1$. In that case, we have

$$f(E, \mu, s) = \begin{cases} g(E, s) & \text{if } \mu = 1 \\ 0 & \text{if } \mu \neq 1 \end{cases}, \tag{3.53}$$

and equation (3.44) becomes

$$\frac{\partial g(E, s)}{\partial s} = \frac{2}{\beta} \cdot \frac{\partial}{\partial E} \left(\frac{C_1}{\beta} \cdot g(E, s) \right). \tag{3.54}$$

Since the two simplified forms of the Fokker–Planck equation are almost identical, let us represent both cases with the following expression:

$$\frac{\partial g(E, s)}{\partial s} = \frac{\xi}{\beta} \cdot \frac{\partial}{\partial E} \left(\frac{C_1}{\beta} \cdot g(E, s) \right), \tag{3.55}$$

where $\xi = 4$ for the isotropic distribution and $\xi = 2$ for the peaked distribution. An accurate simulation of the electron beam would need to include the dependence on pitch angle. However, in this thesis I will simply assume a peaked distribution. This is partly because we do not treat the electron acceleration process in enough detail to be able to say anything about the pitch angle distribution, and partly in order to make the simulation of a large number of beams more computationally manageable.

From equation (3.45), the dependence of C_1 on energy E is logarithmic, and thus quite weak. We will therefore replace $C_1(E, s)$ with an energy-independent approximation $\tilde{C}_1(s) \equiv C_1(E_{\text{mean}}, s)$, where E_{mean} (from Section 3.2.2) is the average non-thermal electron energy at $s = 0$. In addition, we define $\phi(E, s) \equiv g(E, s)/\beta$. We then get

$$\frac{\partial \phi(E, s)}{\partial s} = \frac{\xi \tilde{C}_1}{\beta^2} \cdot \frac{\partial \phi(E, s)}{\partial E}. \tag{3.56}$$

This can be solved by introducing the new variables η and τ, given by

$$d\eta \equiv \beta^2 \, dE \tag{3.57}$$
$$d\tau \equiv \xi \tilde{C}_1 \, ds, \tag{3.58}$$

so that

$$\eta(E) = E^2/(E + 1) \tag{3.59}$$
$$\tau(s) = \xi \int_0^s \tilde{C}_1(s') \, ds'. \tag{3.60}$$

η plays the role of energy, while τ plays a role similar to that of optical depth in radiative transfer theory. Note that both of these quantities are dimensionless. The energy can be computed from η as

$$E(\eta) = \frac{1}{2} \left(\eta + \sqrt{\eta \cdot (\eta + 4)} \right), \tag{3.61}$$

or, in the non-relativistic limit ($\beta \ll 1$)

$$E(\eta) = \sqrt{\eta}. \tag{3.62}$$

In terms of η and τ, equation (3.56) can be written as

$$\frac{\partial \phi(\eta, \tau)}{\partial \tau} = \frac{\partial \phi(\eta, \tau)}{\partial \eta}, \tag{3.63}$$

which has the solution

$$\phi(\eta, \tau) = \phi_0(\eta + \tau), \tag{3.64}$$

or equivalently

$$g(\eta, \tau) = g_0(\eta + \tau), \tag{3.65}$$

where $g_0(\eta)$ is the electron distribution at the acceleration site ($\tau = 0$). In terms of energy E we have $g_0(E) \, \mathrm{d}E = g_0(\eta) \, \mathrm{d}\eta$, so (from equation (3.57)) $g_0(E) = \beta^2(E) g_0(\eta(E))$, or $g_0(\eta) = g_0(E(\eta))/\beta^2(E(\eta))$. We assume that g_0 is normalised like a probability distribution, so that

$$\int_{\eta_{\min}}^{\infty} g_0(\eta) \, \mathrm{d}\eta = \int_{E_{\min}}^{\infty} g_0(E) \, \mathrm{d}E = 1. \tag{3.66}$$

Recall from Section 3.2.2 that N_{tot} denoted the total number of thermal electrons that are accelerated into the beam at the acceleration site. If we instead only consider the (infinitesimal) number of electrons $n(\eta)$ with energy η, we have

$$n(\eta) = N_{\text{tot}} \cdot g_0(\eta). \tag{3.67}$$

This is at $\tau = 0$. More generally, the number of electrons with energy η remaining in the beam at a depth τ will be

$$n(\eta, \tau) = N_{\text{tot}} \cdot g_0(\eta + \tau). \tag{3.68}$$

Since each of these electrons has a kinetic energy $E(\eta)$, we can write the total energy contained in electrons with energy η at a depth τ as

$$\varepsilon(\eta, \tau) = N_{\text{tot}} \cdot E(\eta) \cdot g_0(\eta + \tau). \tag{3.69}$$

Note that the numerical values of equations (3.68) and (3.69) do not correspond directly to kinetic energies or numbers of electrons, but rather densities of these quantities with respect to energy η. To avoid confusion, I will therefore refer to equation (3.68) as a *number distribution* and equation (3.69) as an *energy distribution*. Integrating these distributions over an η-interval would yield the actual kinetic energy or number of beam electrons in the interval. Equations (3.65), (3.68) and (3.69) are plotted in Figure 3.5 when $g_0(E)$ is set to equal the power-law distribution in equation (3.12).

Since $g(\eta, \tau)$ is a probability distribution function, it does not tell us directly how each individual electron will move. However, we can still determine the average behaviour of an electron by considering the curves defined by $g(\eta, \tau) = \text{constant}$, which from
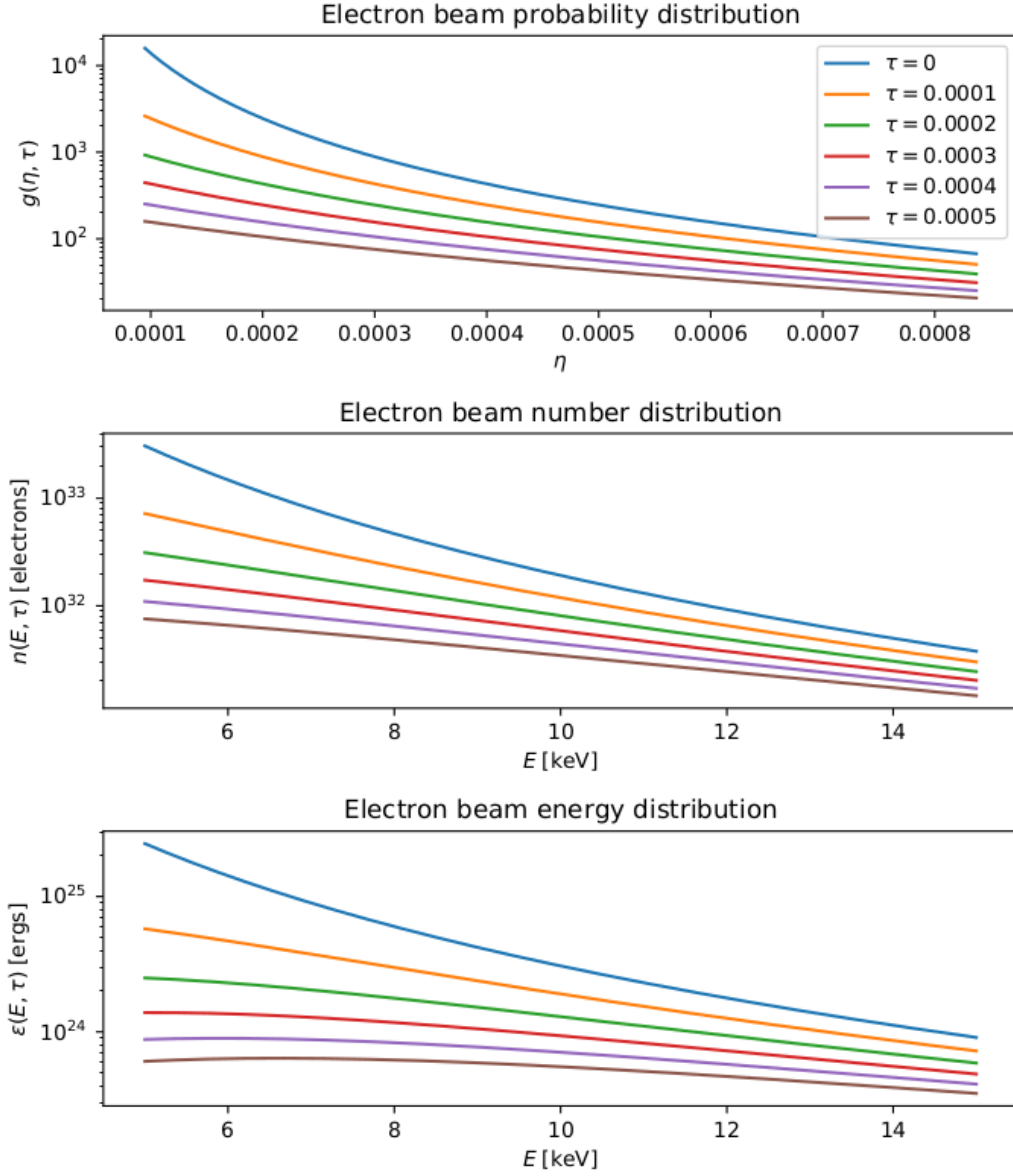
Figure 3.5: Variants of the electron distribution. The plots were computed using equation (3.12) as $g_0(E)$, with $\delta = 4$ and $E_{\min} = 5$ keV. The value used for N_{tot} is 10^{31} electrons, corresponding to a total beam energy similar to that of nanoflares. Upper panel: the probability distribution of the beam plotted as function of η for various equally spaced depths τ. The distribution is simply shifted to the left (lower energies) by a constant amount for each τ, in accordance with equation (3.65). Middle panel: the electron number distribution of the beam (equation (3.68)) as function of energy, for the same values of τ as in the upper panel. Lower panel: the energy distribution of the beam (equation (3.69)) as function of energy for the same values of τ as in the two other panels.

equation (3.65) corresponds to $\eta + \tau = $ constant. From equation (3.68), the number of electrons is conserved along these curves, so they describe a way the electrons can move that is consistent with their probability distribution. In other words, if all the electrons really moved according to $\eta + \tau = $ constant, we would find that their distribution was given by $g(\eta, \tau)$. Let $\eta_0 = \eta(E_0)$ represent the energy of an electron at $\tau = 0$. The kinetic energy of the electron at a different depth τ must then be given by

$$E = E(\eta_0 - \tau) \tag{3.70}$$

with $E(\eta)$ from equation (3.61) or equation (3.62). In the non-relativistic case, equation (3.70) becomes

$$E = E_0 \sqrt{1 - \tau/E_0{}^2}. \tag{3.71}$$

This is a commonly used expression for energy loss in a thick target (see e.g. Brown (1972), equation 15). The energy of an electron as function of distance s along a field line as computed from equation (3.71) is plotted in Figure 3.6 for a large number of electrons with initial energies sampled from equation (3.12). We can see that the
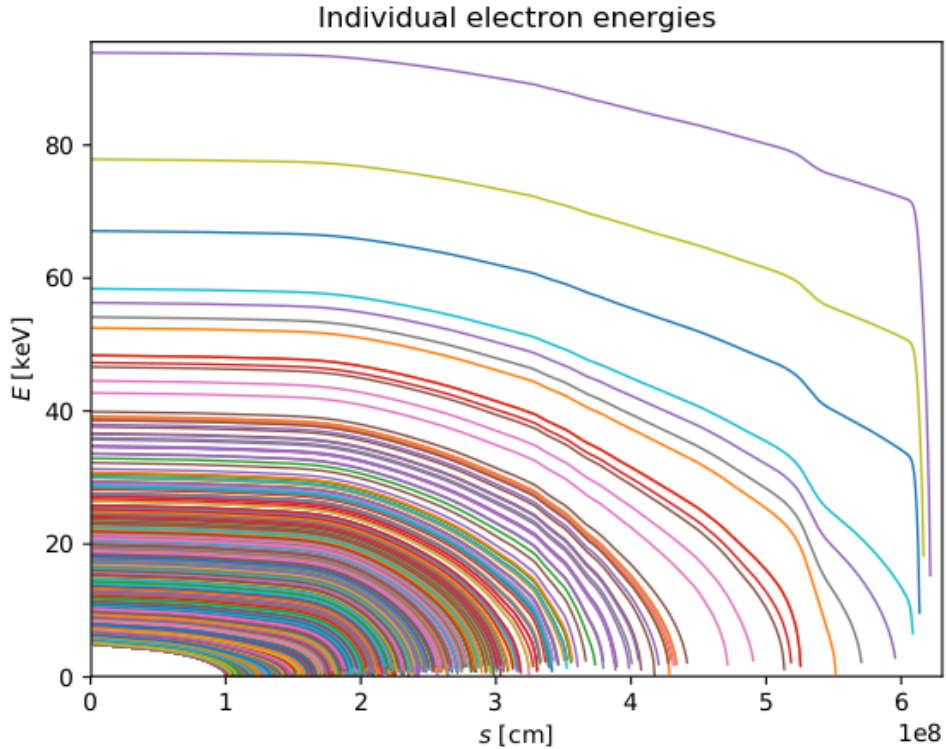
Figure 3.6: Energies of $10\,000$ individual electrons as functions of distance along a magnetic field line. The initial electron energies were sampled from the distribution in equation (3.12), with $E_{\min} = 5$ keV and $\delta = 4$.

electrons with low initial energies stop relatively quickly. From equation (3.71), the electrons have lost their energy when τ reaches $E_0{}^2$. The top part of the field line lies in the corona, where the density, and thus also the collisional coefficient $\tilde{C}_1$, is relatively constant. From the definition of τ, this means that $\tau \propto s$ in this region. So the stopping distance of an electron in the corona is proportional to the square of its initial energy. As the initial energies increase, the electrons penetrate correspondingly deeper. At around $s = 6.1$ Mm in the figure, the energies drop very steeply. This happens because the electrons have reached the transition region where the density increases abruptly, with a corresponding increase in the rate of collisions. The reason that not all the curves go all the way to zero is that the field line was not traced deep enough into the lower atmosphere for the highest-energy electrons to completely thermalise. It is clear that only a very small fraction of the electrons have high enough energies to avoid thermalisation before the end of the field line, so this will not influence the calculated energy deposition from the beam in any meaningful way.

3.4.2 Computing beam heating

In principle, we can integrate equation (3.69) over all energies η to find the total energy $\mathcal{E}(\tau)$ remaining in the beam when it has reached a given depth τ:

$$\begin{aligned} \mathcal{E}(\tau) &= \int_{\eta_{\min}}^{\infty} \varepsilon(\eta, \tau) \, \mathrm{d}\eta \\ &= N_{\text{tot}} \int_{\eta_{\min}}^{\infty} E(\eta) \cdot g_0(\eta + \tau) \, \mathrm{d}\eta. \end{aligned} \tag{3.72}$$

Then, to determine the loss of energy from the beam with depth, we can take the derivative of equation (3.72) with respect to τ, giving

$$\begin{aligned} \frac{\mathrm{d}\mathcal{E}(\tau)}{\mathrm{d}\tau} &= N_{\text{tot}} \cdot \frac{\mathrm{d}}{\mathrm{d}\tau} \left(\int_{\eta_{\min}}^{\infty} E(\eta) \cdot g_0(\eta + \tau) \, \mathrm{d}\eta \right) \\ &= N_{\text{tot}} \int_{\eta_{\min}}^{\infty} E(\eta) \cdot \frac{\partial g_0(\eta + \tau)}{\partial \eta} \, \mathrm{d}\eta. \end{aligned} \tag{3.73}$$

We can avoid the need for working with the (possibly complicated) derivative of g_0 by performing an integration by parts and using the fact that $g_0(\eta)$ has to vanish as $\eta \to \infty$. This gives the expression

$$\frac{\mathrm{d}\mathcal{E}(\tau)}{\mathrm{d}\tau} = -N_{\text{tot}} \cdot E_{\min} \cdot g_0(\eta_{\min} + \tau) - N_{\text{tot}} \int_{\eta_{\min}}^{\infty} \frac{\mathrm{d}E(\eta)}{\mathrm{d}\eta} \cdot g_0(\eta + \tau) \, \mathrm{d}\eta. \tag{3.74}$$

However, the problem with equation (3.74) is that it requires the numerical solution of an integral at every single depth. This quickly becomes too numerically expensive. We will therefore consider a simplified approach based on the fact that electrons with the same energy will be thermalised at the same depth. We start by using equation (3.69) to write the beam energy distribution at $\tau = 0$ as

$$\varepsilon(\eta_0) = N_{\text{tot}} \cdot E(\eta_0) \cdot g_0(\eta_0). \tag{3.75}$$

The total energy $\mathrm{d}\mathcal{E}_0$ contained in all the electrons with initial energies in the infinitesimal interval $\eta_0 \pm \mathrm{d}\eta_0/2$ is then

$$\mathrm{d}\mathcal{E}_0 = \varepsilon(\eta_0)\,\mathrm{d}\eta_0 = N_{\mathrm{tot}} \cdot E(\eta_0) \cdot g_0(\eta_0)\,\mathrm{d}\eta_0. \tag{3.76}$$

An electron is considered thermalised when its energy reaches E_{min}, which from equation (3.70) (when expressed in terms of η instead of E) happens at the depth

$$\tau_{\mathrm{therm}} = \eta_0 - \eta_{\mathrm{min}}. \tag{3.77}$$

So the electrons holding the initial energy $\mathrm{d}\mathcal{E}_0$ are all going to be thermalised in the depth interval $\tau_{\mathrm{therm}} \pm \mathrm{d}\tau_{\mathrm{therm}}/2$. Let us make the approximation that the electrons do not lose any energy until they reach τ_{therm}, at which point all of their energy is converted into thermal energy. In that case, all the energy $\mathrm{d}\mathcal{E}_0$ is going to be deposited in the interval $\tau_{\mathrm{therm}} \pm \mathrm{d}\tau_{\mathrm{therm}}/2$. We can then obtain the deposited energy as a function of thermalisation depth by substituting η_0 with $\tau_{\mathrm{therm}} + \eta_{\mathrm{min}}$ and $\mathrm{d}\eta_0$ with $\mathrm{d}\tau_{\mathrm{therm}}$ in equation (3.76):

$$\mathrm{d}Q = N_{\mathrm{tot}} \cdot E(\eta_{\mathrm{min}} + \tau_{\mathrm{therm}}) \cdot g_0(\eta_{\mathrm{min}} + \tau_{\mathrm{therm}})\,\mathrm{d}\tau_{\mathrm{therm}}. \tag{3.78}$$

I have here changed the name of the variable $\mathcal{E}_0$ to Q to reflect that equation (3.78) concerns energy that is being deposited from the beam. The energy remaining in the beam is then $\mathrm{d}\mathcal{E}(\tau) = -\,\mathrm{d}Q(\tau)$. The rate of change of beam energy with depth τ is then given by

$$\frac{\mathrm{d}\mathcal{E}(\tau)}{\mathrm{d}\tau} = -N_{\mathrm{tot}} \cdot E(\eta_{\mathrm{min}} + \tau) \cdot g_0(\eta_{\mathrm{min}} + \tau). \tag{3.79}$$

Notice that this expression resembles the first term in equation (3.74), the difference being that equation (3.79) uses the "shifted" energy $E(\eta_{\mathrm{min}} + \tau)$ rather than E_{min}.

So how well does this simplified approach work? Figure 3.7 shows the beam energy deposition with distance along a field line computed for both the full result in equation (3.74) and the simplified result in equation (3.79). As an additional point of comparison, both cases were computed with both the relativistic and non-relativistic version of the relation between E and η. We can see that the simplified expression results in somewhat stronger heating near the transition region compared to the full expression. This is due to the approximation that all the electrons with a given energy deposit all their energy at the thermalisation depth. In reality, as Figure 3.6 shows, they deposit their energy continuously above this depth, causing the heating to occur higher up than in the simplified case. Consequently, the simplified expression yields less heating near the reconnection site than the full expression, but this part of the figure is cropped out to better show the peak at the transition region.

The figure also shows that the non-relativistic expressions overestimate the heating compared to the relativistic expressions. This happens because η is mapped to a smaller E in the non-relativistic case, which causes the probability distribution $g_0(E)$ to be evaluated for a smaller E at a given depth τ. This results in a higher numerical value of g_0 at that depth, which translates to stronger heating.
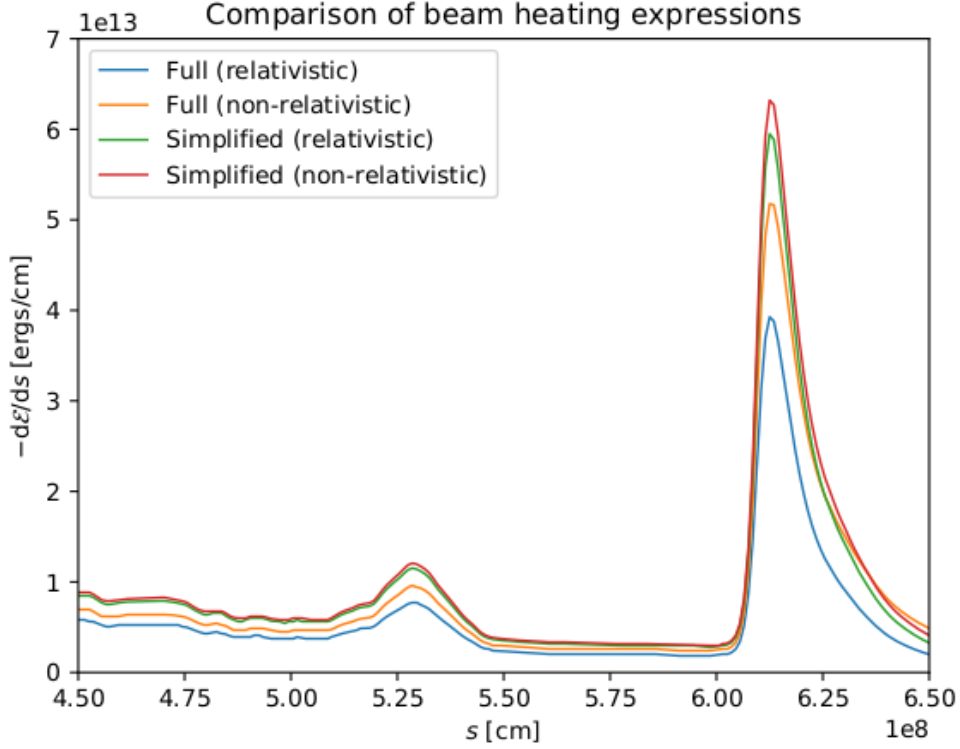
Figure 3.7: Energy deposition from a non-thermal electron beam as a function of distance along a field line. The field line and initial distribution are the same as for Figure 3.6. As with Figure 3.5, the value used for N_{tot} is 10^{31} electrons. The curves denoted as "full" were computed from equation (3.74), while the curves denoted as "simplified" were computed from equation (3.79). In addition, the "relativistic" curves used equation (3.61) for $E(\eta)$, while the "non-relativistic" curves used equation (3.62). Only the lower part of the beam trajectory (in the vicinity of the transition region) is shown in the plot.

The value computed from the most simple expression (simplified and non-relativistic) is about 60% higher at the transition region peak than that computed from the most accurate expression (full and relativistic). I find this error to be acceptable considering the major savings in computational complexity. As we will see in Section 3.5, there is a limit to the number of beams that we can hope to simulate, which will tend to make us underestimate the total amount of heating near the transition region. This will partially cancel out the extra transition region heating that we introduce by using the simplified expression.

When it comes to relativistic versus non-relativistic, the two cases for the simplified expression are sufficiently close that the extra computations for the relativistic case do not appear to be worth it (especially when considering the error we are already making

by using the simplified expression). For the electron distribution plotted in Figure 3.5, a small calculation shows that about 1% of the electrons will have initial speeds exceeding 30% of the speed of light, and these together account for roughly 5% of the total beam energy. The small fraction of beam energy inherent in relativistic electrons supports the choice of neglecting relativistic effects. Still, considering that the most relativistic electrons are also the ones that will be able to make it to the transition region and beyond, this assumption would probably need to be reconsidered if we were to simulate harder distributions or more energetic beams.

In the non-relativistic approximation, and using equation (3.12) for $g_0(E)$, equation (3.79) gives

$$\frac{\mathrm{d}\mathcal{E}(\tau)}{\mathrm{d}\tau} = -N_{\text{tot}} \left(\frac{\delta - 1}{2E_{\min}} \right) \left(1 + \frac{\tau}{E_{\min}{}^2} \right)^{-\delta/2}. \tag{3.80}$$

A closed-form expression for the remaining beam energy $\mathcal{E}(\tau)$ at depth τ can now be found by integrating equation (3.80) from infinity to τ (the lower integration limit is infinity because we know that $\mathcal{E}(\tau) \to 0$ when $\tau \to \infty$):

$$\mathcal{E}(\tau) = \int_{\infty}^{\tau} \frac{\mathrm{d}\mathcal{E}(\tau')}{\mathrm{d}\tau'}\, \mathrm{d}\tau' = N_{\text{tot}} \left(\frac{\delta - 1}{\delta - 2} \right) E_{\min} \left(1 + \frac{\tau}{E_{\min}{}^2} \right)^{1-\delta/2}. \tag{3.81}$$

Comparing with equation (3.14) we can recognise the two factors in the middle of the above expression as nothing more than E_{mean}. When multiplying this with the total number of beam electrons, we simply obtain the total beam energy E_{tot}, so equation (3.81) reduces to

$$\mathcal{E}(\tau) = E_{\text{tot}} \left(1 + \frac{\tau}{E_{\min}{}^2} \right)^{1-\delta/2}. \tag{3.82}$$

This result is useful for computing the total beam heating between two arbitrary depths τ and $\tau' > \tau$, which simply becomes

$$Q(\tau, \tau') = \mathcal{E}(\tau) - \mathcal{E}(\tau'). \tag{3.83}$$

Numerically, using equation (3.83) for calculating the beam heating is better than performing a numerical integration of equation (3.80), since equation (3.83) does not introduce truncation errors in the result.

3.4.3 Excluding local heating events

At some acceleration sites the gas density might be high enough that the electron beam will only be able to travel a very small distance before all of its energy has been deposited in the surrounding plasma. These events have the same effect as when the reconnection energy is converted directly into heat, so they do not represent a significant transport of energy from one place to another. For the purpose of computational efficiency, it is a good idea to exclude these sites from the transport simulation in the first place. Fortunately, this is easy to achieve with the help of equation (3.82). We want to obtain the distance δs at which the remaining beam energy drops below a small threshold

value E_{lim}. Over small distances, we can assume that the collision coefficient $\tilde{C}_1$ stays nearly uniform, so that the corresponding depth $\delta\tau$ can be written as $\delta\tau \approx \xi\tilde{C}_1\delta s$. The threshold energy can then be written as

$$E_{\text{lim}} = E_{\text{tot}} \left(1 + \frac{\xi\tilde{C}_1\delta s}{E_{\text{min}}^2} \right)^{1-\delta/2}. \tag{3.84}$$

Solving this equation for δs gives

$$\delta s = \frac{E_{\text{min}}^2}{\xi\tilde{C}_1} \left(\left(\frac{E_{\text{lim}}}{E_{\text{tot}}} \right)^{2/(2-\delta)} - 1 \right). \tag{3.85}$$

$\tilde{C}_1$ can be computed from the particle density at the acceleration site. Given the beam parameters E_{min}, δ, E_{tot} and ξ, as well as the energy threshold E_{lim}, we can check whether the depletion distance δs is smaller than a few grid cell widths. If it is, we can safely assume that no significant amount of energy will be transported away from the site via the electron beam.

The code also performs an initial, more coarse rejection of reconnection sites based solely on their depth and temperature. Sites lying very low in the atmosphere (e.g. below the lower chromosphere) will not contribute to energy transport by non-thermal electrons because the surrounding density would stop the beams immediately, so they can be excluded straight away. In addition, sites with low temperatures will generate electron beams with small values of E_{min} (see Section 3.2.2). These beams will consist mainly of relatively low-energy electrons that lose their energies very quickly and hence not contribute significantly to the energy transport. So low-temperature sites are also ignored.

3.5   Deciding how many beams to simulate

As discussed in Section 3.2.1, we consider electron beams to originate wherever the value of the reconnection heuristic value $\tilde{K}$ is higher than some threshold $\tilde{K}_{\text{lim}}$. Hence, the total number of electron beams we need to simulate depends on the value we set for $\tilde{K}_{\text{lim}}$. Because this is a free parameter that is significantly going to affect the total heating by electron beams, it is important to consider carefully which value to use. We want to include enough beams that we do not underestimate the total heating, while at the same time being able to complete the simulation in a reasonable amount of time.

Figure 3.8 shows the horizontal average of the total beam heating in the atmosphere as a function of depth for different values of $\tilde{K}_{\text{lim}}$. The curves are also annotated with the total number of beams generated for each threshold value. We will study a similar plot of the beam heating in more detail in Section 4.3.2. For now, we will focus on the peaks near $z = -1$ Mm (just below the transition region), where the beams deposit most of their energies. As one would naturally expect, the height of the peaks increase with the number of beams. However, this effect slows down for the largest beam numbers: although the number of beams increases by nearly $50\,000$ from $\tilde{K}_{\text{lim}} = 2$ to $\tilde{K}_{\text{lim}} = 1$,
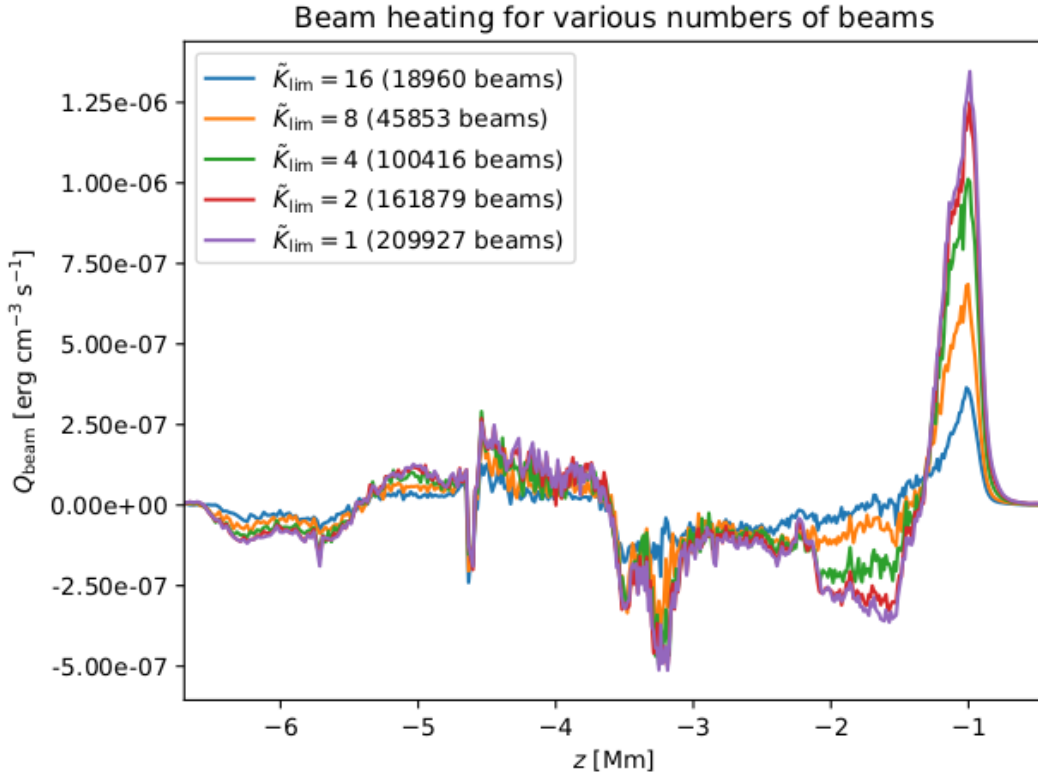
Figure 3.8: Total beam heating in the atmosphere as a function of depth, averaged over each horizontal layer. Each curve corresponds to a separate simulation with a different value of $\tilde{K}_{\mathrm{lim}}$, and correspondingly a different number of electron beams.

the corresponding heating curves are very similar. This suggests that there is little extra heating to be gained by going below 1–2 for the threshold value.

Another consideration is how the total number of beams affects the execution time of the code. This is shown in Figure 3.9, where the measured execution time for a single time step is plotted against the number of beams, for the same threshold values as in Figure 3.8. The execution time is affected not only by the number of beams but also by how they are distributed throughout the atmosphere. This will determine the average length of the field lines and also how the work load is distributed between different processes (see Appendix B). However, apart from a dip for $\tilde{K}_{\mathrm{lim}} = 4$, the execution time appears to scale quite linearly with the number of field lines, at least for the particular snapshot used in these runs. Compared to the MHD simulations, the execution times for the electron beam module are higher by a factor of 1 to 5. From $\tilde{K}_{\mathrm{lim}} = 2$ to $\tilde{K}_{\mathrm{lim}} = 1$, the execution time increases by roughly 30%. Considering the very minor increase in heating that is gained by using $\tilde{K}_{\mathrm{lim}} = 1$ instead of $\tilde{K}_{\mathrm{lim}} = 2$, this is in my opinion not worth the extra execution time. The results in the following sections are therefore produced with $\tilde{K}_{\mathrm{lim}} = 2$.
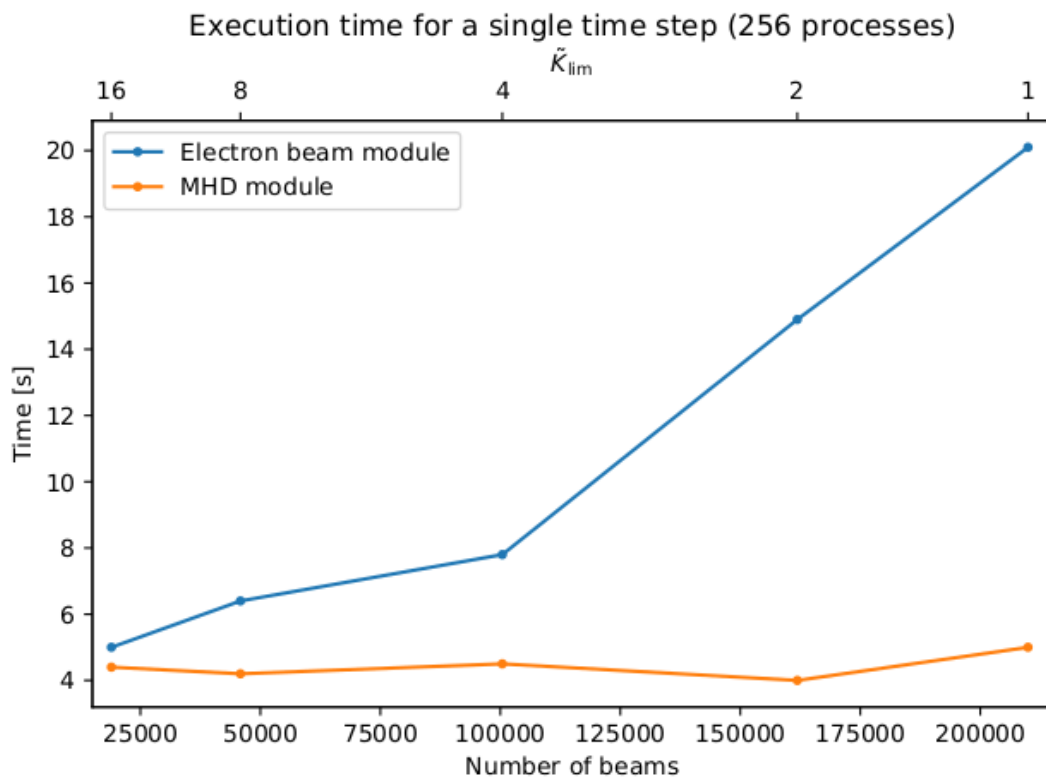
Figure 3.9: Measured execution times of the electron beam module for a single time step with various numbers of beams. The corresponding execution times for the module responsible for the MHD simulation are also included for reference. The axis on top indicates the corresponding values of $\tilde{K}_{\mathrm{lim}}$ that were used. Each run was performed with 256 MPI processes.

Chapter 4

Results

4.1 The initial atmosphere

In order to study the effect of the electron beams in a realistic solar atmospheric box, I re-ran a pre-existing Bifrost simulation of such an atmosphere with the electron beam physics included. This makes it easy to compare the results with and without electron beams. The box measures about 25 Mm in the horizontal directions, with depth ranging from $z = 2.5$ Mm at the bottom to $z = -14.3$ Mm at the top. The zero level of depth is made to correspond roughly to the top of the photosphere. There are 768 grid cells along each dimension, meaning that each grid cell spans about 32 km in each horizontal direction. The resolution is non-uniform in depth, as grid cells near the transition region are packed much tighter together in the z-direction than the cells at the far top and bottom. This is done to accommodate the requirement for higher resolution near the transition region where the atmosphere changes very rapidly with depth. The horizontal boundary conditions are periodic: everything exiting one side of the box will re-enter through the opposite side. The fluid motions in the bottom of the box are driven by a forced convection. In addition, magnetic flux is pushed up through the bottom boundary, causing reconnection as emerging magnetic flux tubes collide with the pre-existing coronal magnetic field.

Figure 4.1 shows the temperature, density and Joule heating through a vertical slice of the Bifrost atmosphere at the time when we first include the beam simulations. Note that I will use the term *snapshot* to refer to the state of the atmosphere at a particular point in time. The initial snapshot has a relatively cold corona, as can be seen in the top panel of the figure. There are some hot regions with temperatures of several million kelvins, but most of the corona has temperatures below 10^5 K. There are also cold spots with temperatures of just a few thousand kelvins. The coronal density is correspondingly high, as can be seen in the middle panel. It is also evident that the areas with the highest temperatures are typically also the ones with the lowest densities. The reason for the low coronal temperatures and high densities is likely that the magnetic field emerging through the bottom boundary has lifted cool, dense photospheric and chromospheric material into the corona. The arcing structure of this magnetic field is
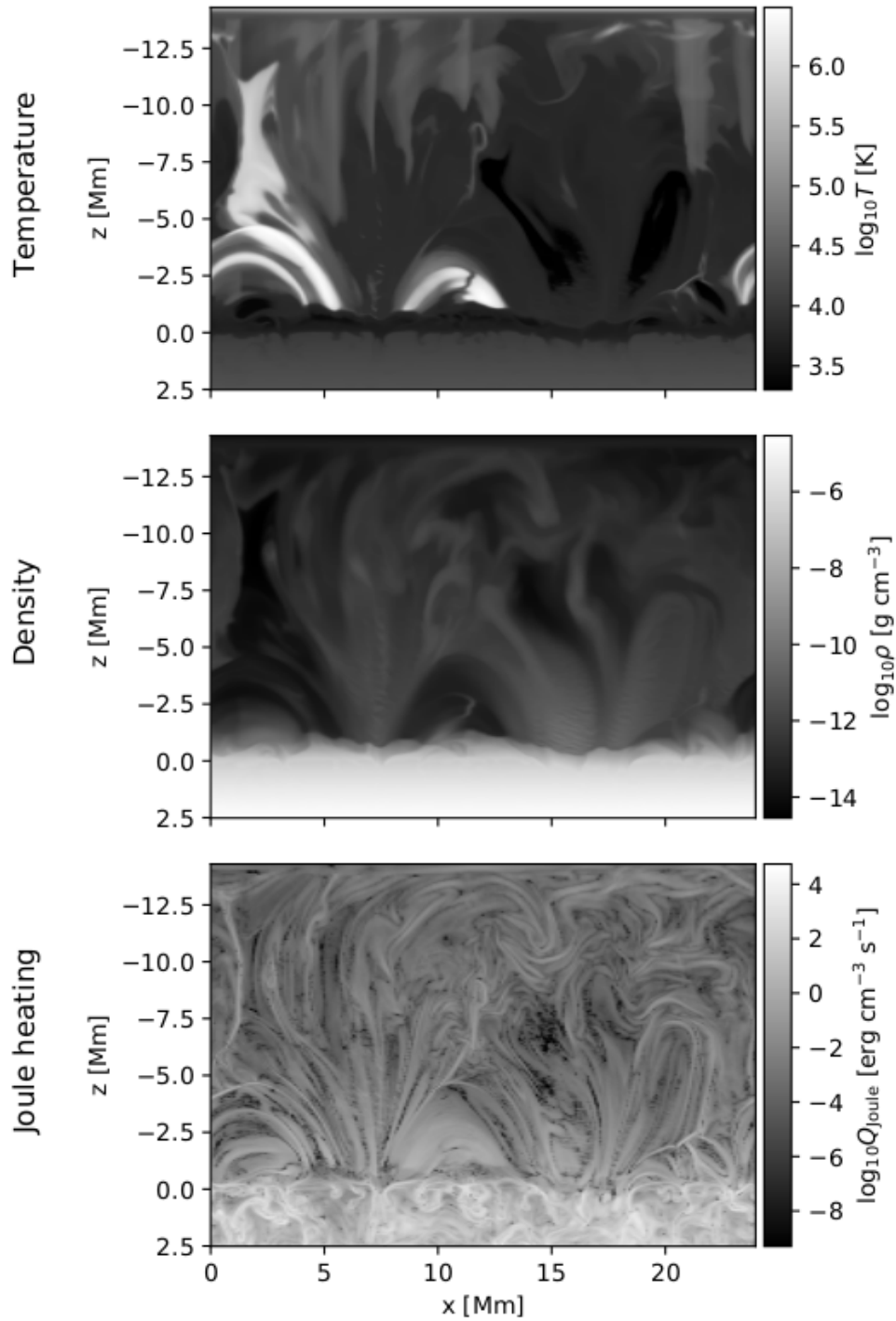
Figure 4.1: Slices in the xz-plane of the initial atmosphere, near the centre of the box at $y = 12.3$ Mm. The upper panel shows the gas temperature, the middle panel shows the mass density and the lower panel shows the Joule heating, all in logarithmic scales.

quite evident in all three plots. Due to reconnection of the emerging magnetic field lines, the corona is relatively active, as the plot of Joule heating in the bottom panel shows. Joule heating occurs where gradients in the magnetic field generate electric currents. The current flows along the coronal magnetic field lines, revealing them as thin arcs of concentrated Joule heating in the figure.

I investigated three different atmospheric simulations as candidates for the electron beam simulations. My reason for choosing this particular atmosphere was simply that, despite the cool, dense corona, it yielded the highest amount of beam heating in the transition region during the first time step. The other two atmospheres both had hotter and more tenuous coronas, which would let the electron beams retain a larger fraction of their energies before hitting the transition region. However, they also had significantly less Joule heating, meaning that the available energy for the electron beams would be much smaller in the first place.

4.2   Acceleration sites

Figure 4.2 shows a 3D view of the simulation box with the acceleration sites obtained using a threshold value of $\tilde{K}_{\mathrm{lim}} = 2$. The colours indicate the amount of energy e_{tot}
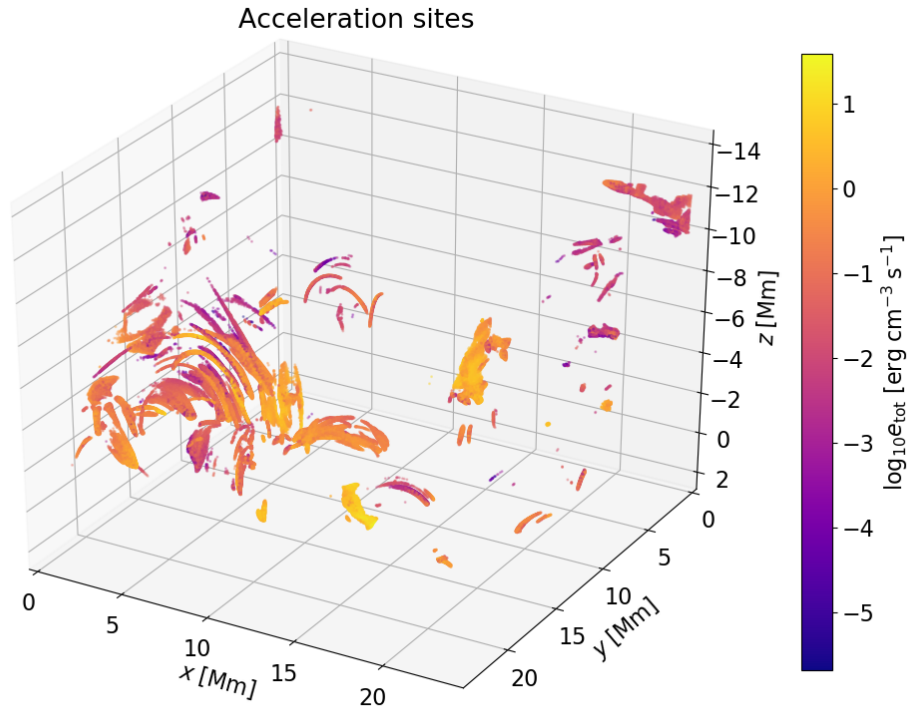
Figure 4.2: Points in a Bifrost simulation for which $\tilde{K} > 2$. There are roughly $160\,000$ acceleration sites in the figure. The colours represent the total beam energy.

injected into the electron beams per volume and time. This quantity is related to the total beam energy by $E_{\text{tot}} = e_{\text{tot}} V_{\text{rec}} \Delta t_{\text{rec}}$, where V_{rec} and Δt_{rec} are the volume of the reconnection region and duration of the reconnection event, respectively. The beam energies in the figure vary by a few orders of magnitude, but typical values of e_{tot} appear to be between 0.1 and 10 erg cm^{-3} s^{-1}. To get a better sense of the scale of these energies, consider a reconnection event in a volume extending 1 Mm in each direction and lasting 1 s. With $e_{\text{tot}} = 1$ erg cm^{-3} s^{-1}, the total beam energy would be 10^{24} erg. This makes it clear that we are dealing with very small flares, with energies comparable to those of the nanoflares mentioned in Section 2.2.

The acceleration sites in the figure tend to occur in elongated clusters, presumably following the shape of the current sheets forming at the intersections of different magnetic domains. There is a majority of points at lower heights, and these also tend to produce somewhat higher beam energies than the points higher up. This is likely due to the fact that the available magnetic energy decreases with height as the magnetic field spreads out, and at the same time becomes less turbulent and dynamic. The absence of points in the lowest part of the box is due to the restrictions that were imposed on the temperature and depth of the acceleration sites, as discussed in Section 3.4.3. In particular, points lying deeper than $z = -0.5$ Mm or with temperatures below 10^5 K were automatically rejected.

4.3 Beam heating during a single time step

4.3.1 Heating from a single electron beam

Before studying the effect of beam heating in a full Bifrost simulation evolving in time, it is informative to look at how electron beams transport energy during a single time step. We will first look at the effect of a single beam. An example of the heating along a field line due to an electron beam is plotted in Figure 4.3. The beam originates at an acceleration site slightly higher than 6 Mm above the photosphere. We can see that a great deal of heating occurs in the immediate vicinity of the acceleration site. This is because a large fraction of the beam electrons have a relatively low energy, meaning that they will be thermalised quickly. This local beam heating will partially compensate for the removed Joule heating near the acceleration site (recall that the energy allocated to the beam is removed from the already computed Joule heating at the site)[1]. After leaving the acceleration site, the beam electrons that still remain follow the magnetic field and descend downwards through the corona. Due to the relatively dense corona in this particular Bifrost snapshot, the amount of heating is quite high throughout the corona, but decreases with distance as an ever larger fraction of the beam electrons become thermalised. The sudden peak in heating near $z = -1$ Mm occurs when the electrons hit the transition region, where the plasma is a lot denser. Over the next few megametres, all the remaining electrons thermalise and the beam vanishes.

---

[1]Note that the presence of beams still ends up reducing the net heating near the acceleration sites, since more energy is injected into the beams than what they deposit locally.
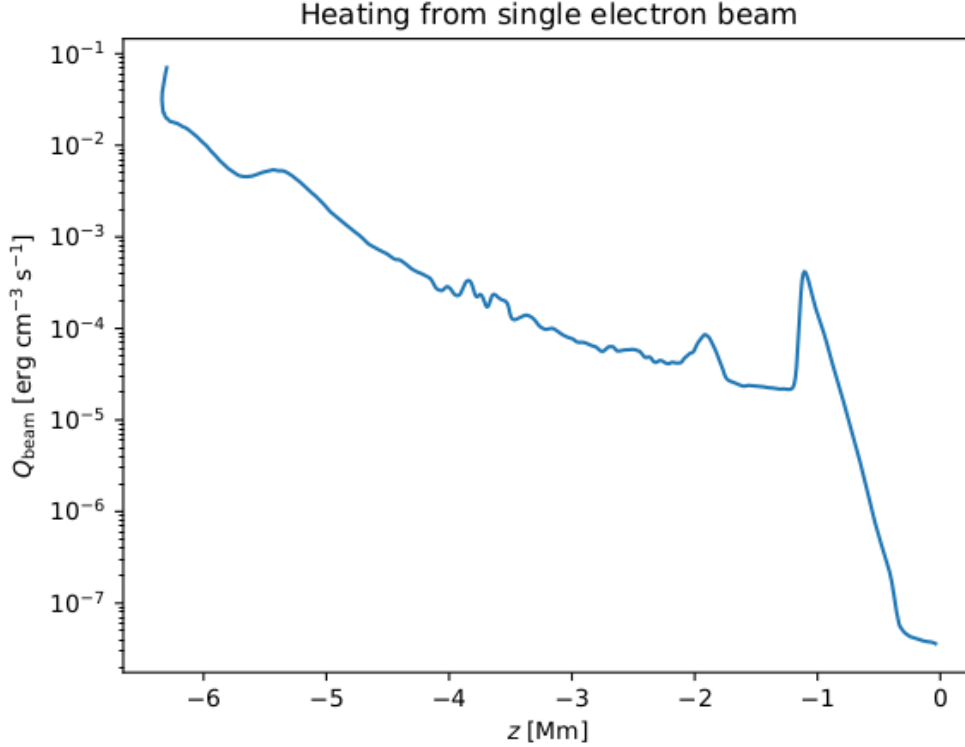
Figure 4.3: The heat deposited by a single electron beam as a function of vertical depth. The field line is the same as the one used in Figures 3.6 and 3.8. The total energy of the beam is about $e_{\text{tot}} = 3 \text{ erg cm}^{-3} \text{ s}^{-1}$.

4.3.2 Vertical distribution of beam heating

Having followed the path of a single electron beam, we now look at the collective effect of a large number of beams. The horizontal average of the total beam heating in the initial atmosphere after a single time step is plotted in Figure 4.4. Also shown for each depth are the horizontal averages of the initial energies of the beams, as well as their total deposited energy. Notice that these two curves are quite similar. This is due to the rapid thermalisation of low-energy beam electrons (as we saw in Figure 4.3), causing a large fraction of the total beam energy to be deposited very close to the acceleration sites. The net beam heating curve, which is the difference between the two, thus corresponds to energy transport via the electrons that travel farther through the atmosphere before stopping.

There is a clear peak in beam heating near $z = -1$ Mm, just below the transition region. This is where the density becomes high enough that all the non-thermal electrons collide with ambient plasma and deposit their energy. Many of these electrons appear to come from a layer with a width of a few megametres just above the transition region, where we see a net reduction of heating. There is also a similar layer of reduced heating
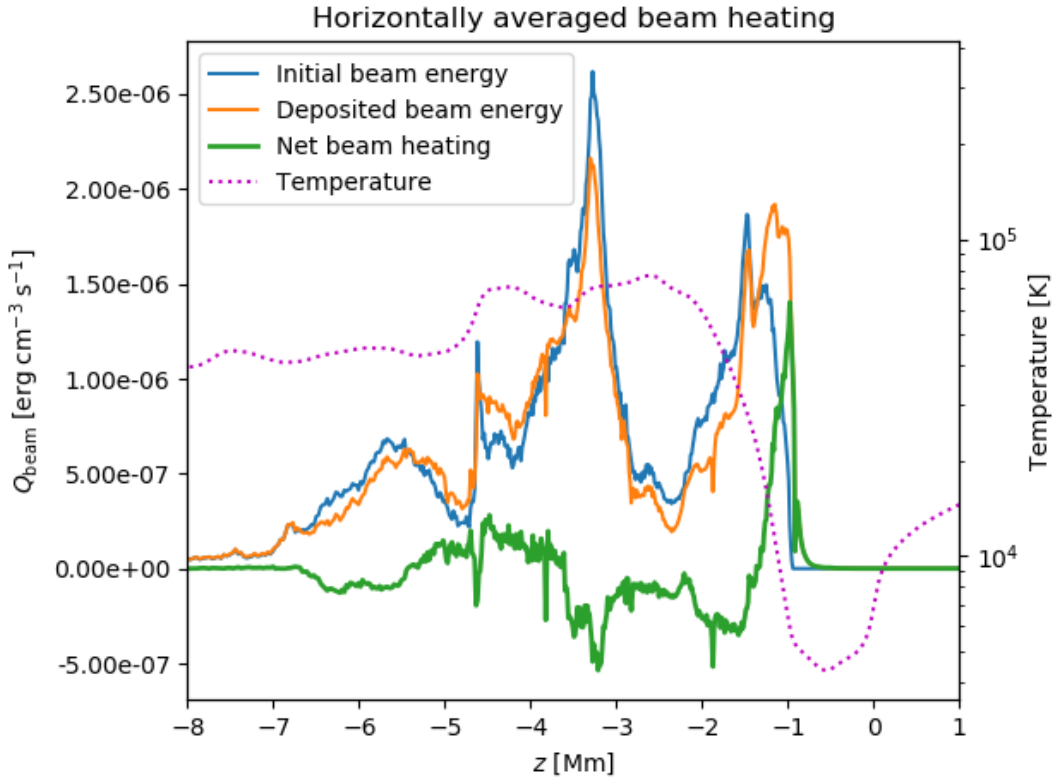
Figure 4.4: Horizontally averaged beam heating as a function of depth. The initial beam energies and total energy deposited by the beams (also horizontally averaged) are shown as well. The net beam heating is the deposited energy minus the initial beam energy. In order to indicate the location of the transition region, the horizontally averaged temperature is also included. A threshold of $\tilde{K}_{\text{lim}} = 2$ was used for the acceleration sites.

somewhat higher up in the corona near $z = -6$ Mm, contributing to the transition region heating via electrons with initial energies high enough to make it through most of the atmosphere.

Notice that the curve of initial beam energies goes to zero near $z = -1$ Mm. This is above the artificial depth limit of $z = -0.5$ Mm that was used to exclude acceleration sites that are thought not to contribute to the heating, so we know that our particular choice of the value for the limit does not affect the results. We also had a corresponding lower limit of 10^5 K for the temperature at the acceleration sites. This is what is causing the beam energies to go to zero where they do. I found that decreasing the temperature limit increased the number of acceleration sites by a factor of several, but it did not affect the net beam heating in any significant way. The reason for this is that low temperatures will lead to very small values of E_{min} (see Section 3.2.2), so the corresponding beams will deposit all their energy almost immediately.

Average beam heating in the range $z = [-1.2, -0.9]$ Mm

Average Joule heating in the range $z = [-1.2, -0.9]$ Mm
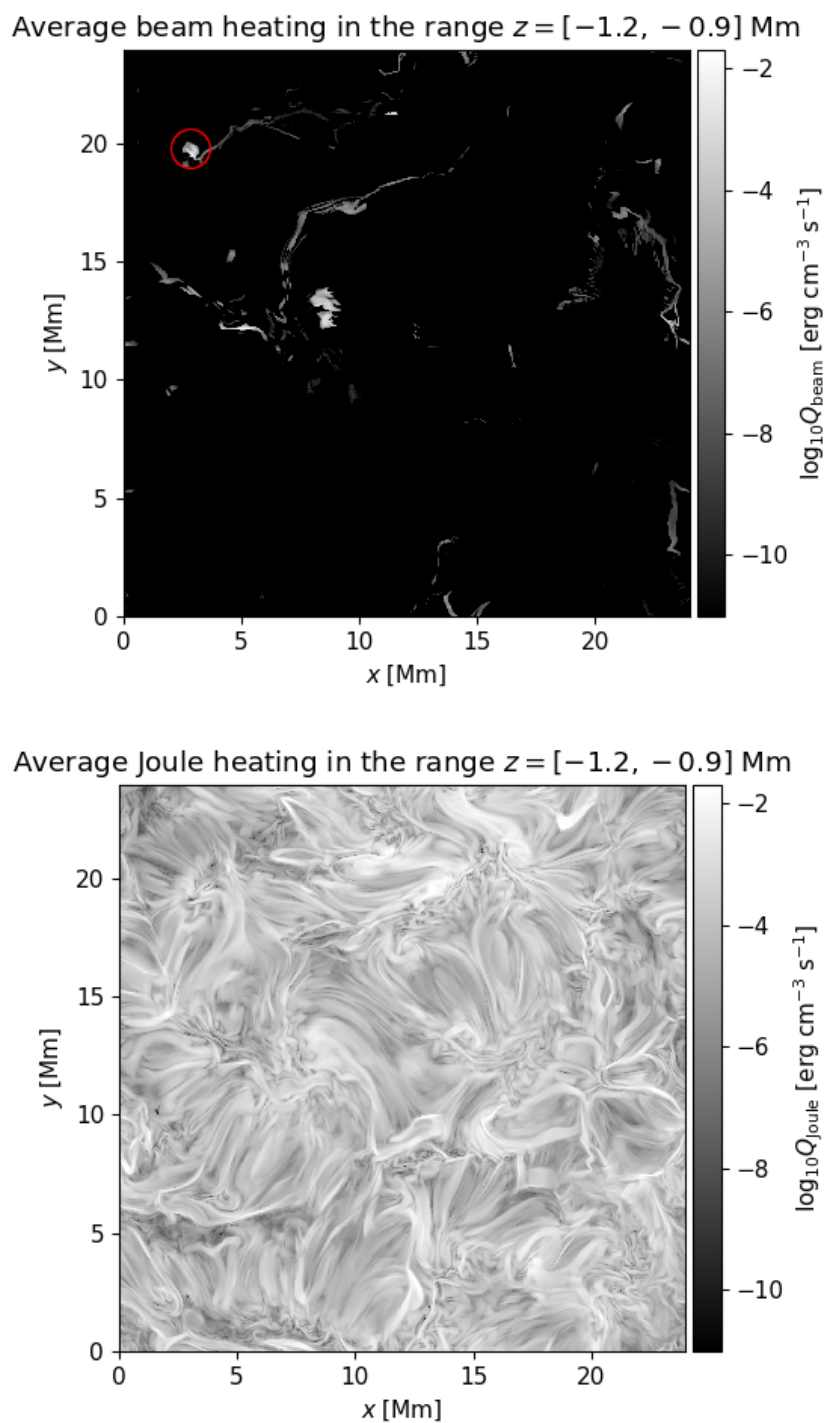
Figure 4.5: Top panel: Beam heating averaged vertically in a 0.3 Mm range centred on the positive heating peak in Figure 4.4. The red circle indicates a heating site discussed in Section 4.4.2. Bottom panel: Average Joule heating in the same height range.

4.3.3   Horizontal distribution of beam heating

While it can indeed be informative to look at averages over horizontal layers to study the effect of beam heating, it hides the fact that the sites of beam heating are not evenly distributed horizontally, but rather tend to cluster in small regions where the magnetic field lines converge. This is evident from the top panel in Figure 4.5, where the horizontal distribution of beam heating around the heating peak in Figure 4.4 is shown. Most of the locations in this layer have no beam heating at all, but there are a few isolated regions with relatively strong heating. These regions correspond to coronal loop foot points where a large number of field lines converge in a tight bundle.

The bottom panel shows the corresponding horizontal distribution of Joule heating for comparison. Overall, the total amount of beam heating is relatively minor compared to the total amount of Joule heating. However, the sites where there is significant beam heating have values comparable to those of Joule heating, indicating that the electron beams may have a notable heating impact in certain regions. Notice that the pattern of beam heating has many similarities to the pattern of Joule heating. This is reasonable considering that the electron beams converge along with the magnetic field lines to locations of strong magnetic flux emergence, and at these locations there also tends to be strong induced currents that lead to Joule heating.

4.4   Effect of beam heating on the atmosphere

4.4.1   Differences at beam heating sites

In order to get a general idea of how various atmospheric properties are affected by the presence of electron beams, I compared the values at certain locations in the beam simulation to the corresponding values in the original simulation (the one without beams). There are three separate effects that can lead to a discrepancy between the two simulations at some position. Firstly, energy could be deposited at the position by passing beams. Secondly, the position could correspond to an acceleration site, meaning that local Joule heating will have been removed and put into the beam that originates there. Finally, the two simulations could simply have evolved away from each other. A tiny perturbation to the state of the atmosphere will eventually lead to completely different results. This happens because of the non-linearity of the MHD equations and is sometimes referred to as the *butterfly effect*[2]. Even if we did not have the discrepancies caused by the extra beam physics, we would still be left with tiny perturbations throughout the atmosphere caused by variations in how floating-point values are rounded[3]. Since the aim here is to look at the heating of the plasma caused by electron beams, it is only the differences due to the first effect mentioned above that are of interest. Care

---

[2]Coined by American meteorologist Edward Lorenz in 1969, the term refers to a metaphor where the flapping of the wings of a butterfly affects the path of a tornado by introducing a tiny initial perturbation to the evolution of the non-linear weather system.

[3]Even though IEEE floating-point operations are in themselves deterministic, running the code on different machines will typically introduce variations that lead to different rounding, like expressions not being evaluated in the same order.
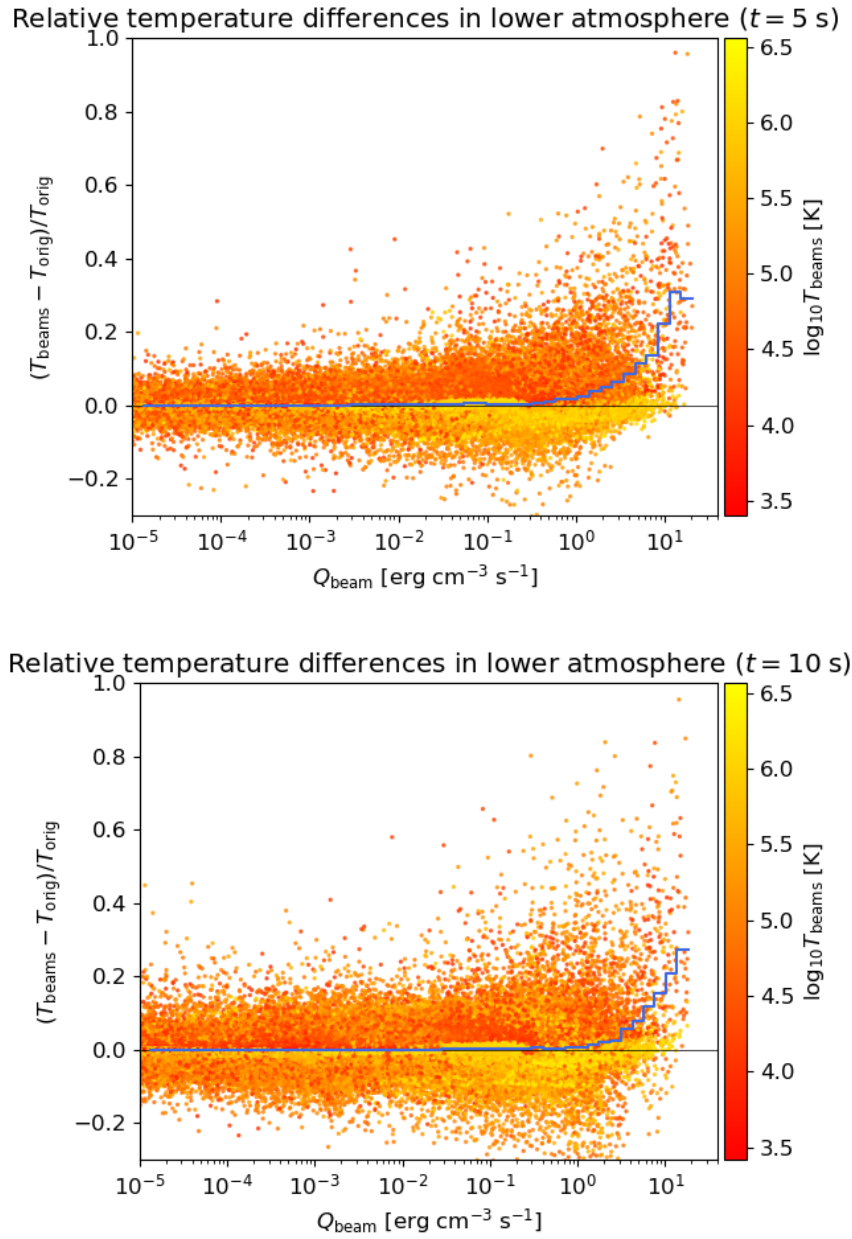
Figure 4.6: Relative differences in temperature between the simulations with and without electron beams, for varying amounts of beam heating, when the simulations have been run for 5 (upper panel) and 10 (lower panel) seconds of solar time. T_{beams} is the temperature at a point in the beam simulation, and T_{orig} is the temperature at the corresponding point in the original simulation. Q_{beam} was computed as the average beam heating since $t = 0$ s. The blue curve is the running average of the temperature difference, with each horizontal segment indicating the width of the bin in which the differences were averaged.

thus has to be taken when comparing the two simulations so that differences due to the butterfly effect are not attributed to the presence of beams.

Temperature

Figure 4.6 shows the relative difference in gas temperature between the two simulations after 5 and 10 seconds of solar time. Only the points below $z = -2$ Mm and where the beam heating is larger than 10^{-5} erg cm^{-3} s^{-1} are included in the plots. This automatically excludes points like acceleration sites where the total heating is reduced due to Joule heating being converted into beam energy. The temperature differences are plotted against the value of beam heating, and the colours represent the actual temperatures taken from the simulation with beams.

The plots contain less than 0.5% of all the points below $z = -2$ Mm. This immediately tells us that just a very small fraction of the points in the lower atmosphere will be directly influenced by beam heating. For points with $Q_\mathrm{beam} < 0.1$ erg cm^{-3} s^{-1}, the average temperature difference (as indicated by the blue curve) remains close to zero after both 5 and 10 seconds. However, their spread in temperature differences increases over time. Because these points have never been subjected to any significant amount of beam heating, this spread must have been caused by the butterfly effect. The fact that the butterfly effect has not changed the average temperature difference indicates that the average is a reasonable measure of the direct influence of the electron beams on the plasma. For larger values of beam heating, the average increase in temperature grows steadily, peaking at around 35% at the sites with the strongest beam heating. For these sites, the individual temperature differences range from no change at all to a doubling of temperature.

Density

Figure 4.7 shows the relative differences in density after 5 seconds, plotted in the same way as for the temperature. Unlike an increase in internal energy (and hence in temperature and pressure), a change in density is not a primary reaction to an enhanced rate of heating. The density is affected by the divergence of the velocity field: if more plasma flows out of a region than into it, the density decreases. Such a situation can be caused by a localised increase in pressure, but it takes time to accelerate the plasma. This resistance to change is reflected in the figure. On average, the density remains unchanged for most values of beam heating. There is a very slight increase in density for intermediate beam heating values, and a small decrease in density for the highest beam heating values, but these differences are only in the order of 1% on average. I will come back to the reason for these slight differences in Section 4.4.2.

Pressure

The corresponding plot of the relative pressure differences after 5 seconds looks almost identical to the upper panel in Figure 4.6, so I do not include it here. Considering the very minor average differences in density in Figure 4.7, this similarity between pressure
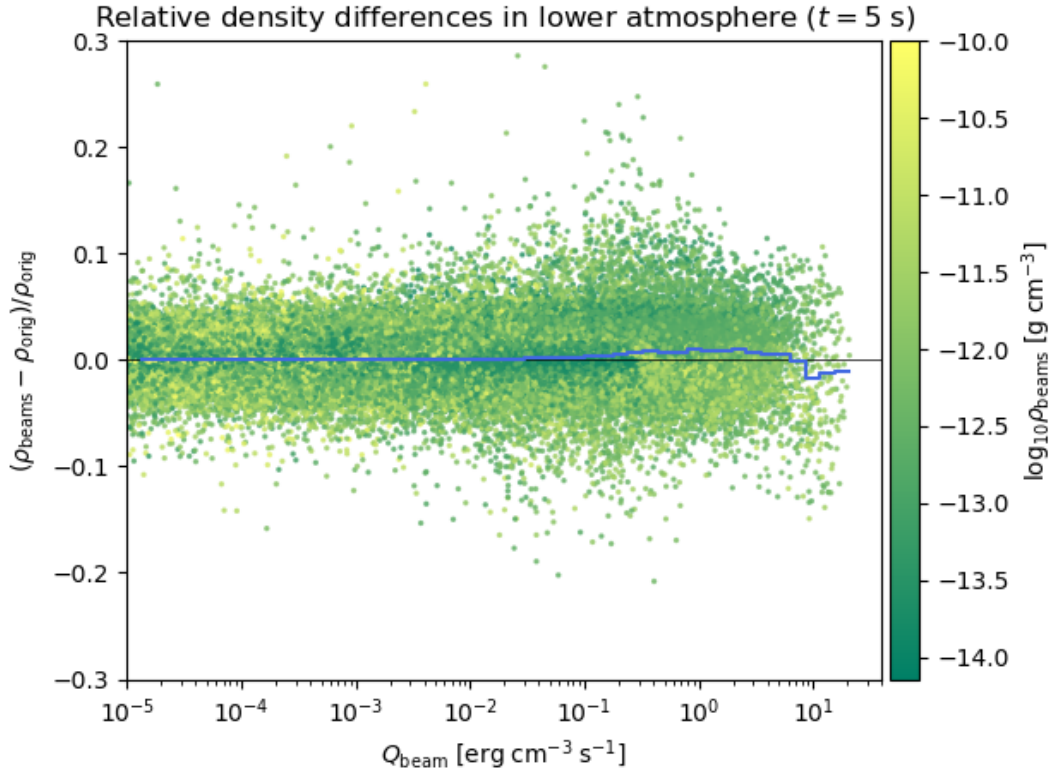
Figure 4.7: Relative differences in mass density after 5 seconds of solar time, in the same format as the plots in Figure 4.6.

and temperature is indeed what we would expect if we assume that the plasma follows the ideal gas law[4]:

$$P = \rho k_{\mathrm{B}} T. \tag{4.1}$$

It is clear from this equation that the pressure will behave in exactly the same way as the temperature if the density is kept constant.

Vertical velocity

Let us finally consider the differences in vertical plasma velocity between the two simulations. These are shown after 5 seconds in Figure 4.8. This time, absolute differences are used rather than relative differences. The points with significant beam heating appear to exhibit both downward and upward motions, but downward motion appears to be the most prevalent. In the beam simulation, the plasma has been accelerated upwards relative to the original simulation by up to 2 km/s on average, with the largest changes

---

[4]The equation of state used for the simulations is in fact not the ideal gas law, but a more detailed relation that has been pre-computed and tabulated. Still, the ideal gas law works as a decent approximation for understanding the rough behaviour of the plasma.
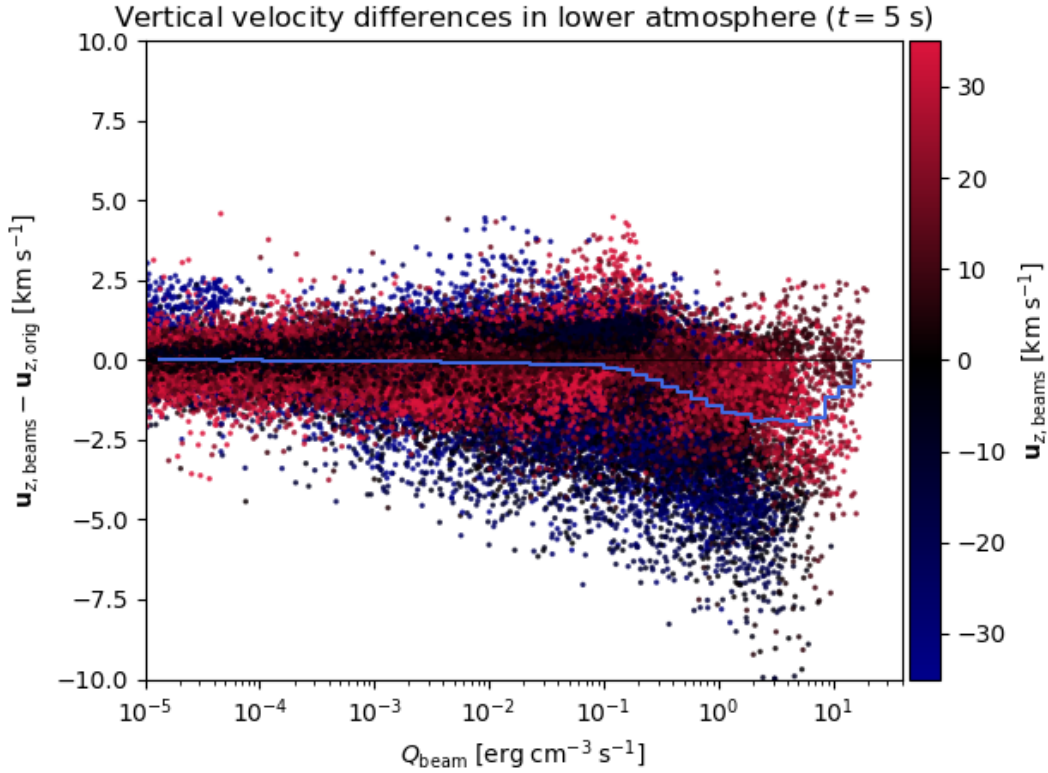
Figure 4.8: Absolute differences in vertical plasma velocity, plotted in the same way as in Figures 4.6 and 4.7. Note that positive vertical velocities result in motion downwards in the atmosphere. The colour coding corresponds to the Doppler-shifts of the velocities, with red indicating downward motion and blue indicating upward motion.

occurring at sites of intermediate to strong beam heating. Differences of up to 10 km/s in the upward direction can be seen for some of the sites.

Preliminary discussion

The results that we have looked at so far make it clear that heating from electron beams has a direct, appreciable influence on the plasma at a relatively small number of points in the lower atmosphere where the heating is most intense. We know from Section 4.3 that these points predominately are located near the bottom of the transition region (around $z = -1$ Mm), and that they form several extended heating sites at the foot points of coronal loops. In the corona, there are scattered sites of strong beam heating close to where electrons are accelerated, as we saw from Figure 4.3. However, the high coronal conductivity will quickly make sure that this heating just goes into compensating for the removal of energy at the acceleration sites. This justifies the exclusion of points above $z = -2$ Mm in the previous figures. We saw that after 5 seconds, the temperature and pressure at the transition region beam heating sites have

increased by up to 35% on average, and the heated plasma has been accelerated upwards by several km/s. Meanwhile, the density has remained relatively unaffected.

Although useful for providing an overview of how the plasma is affected by electron beams, the average differences considered in this section do not tell the whole story of what is happening at the beam heating sites. In the following section, we will gain a better understanding of what the electron beams do by studying the state of the plasma near the foot point of a coronal loop.

4.4.2 Effect on plasma in a coronal loop

The most intense beam heating regions occur where a large number of beams converge on the same spot in the lower atmosphere. This can happen near the foot points of coronal loops, where the associated magnetic field lines are tightly bundled together. In this section we will study the effect of beam heating on the transition region plasma inside
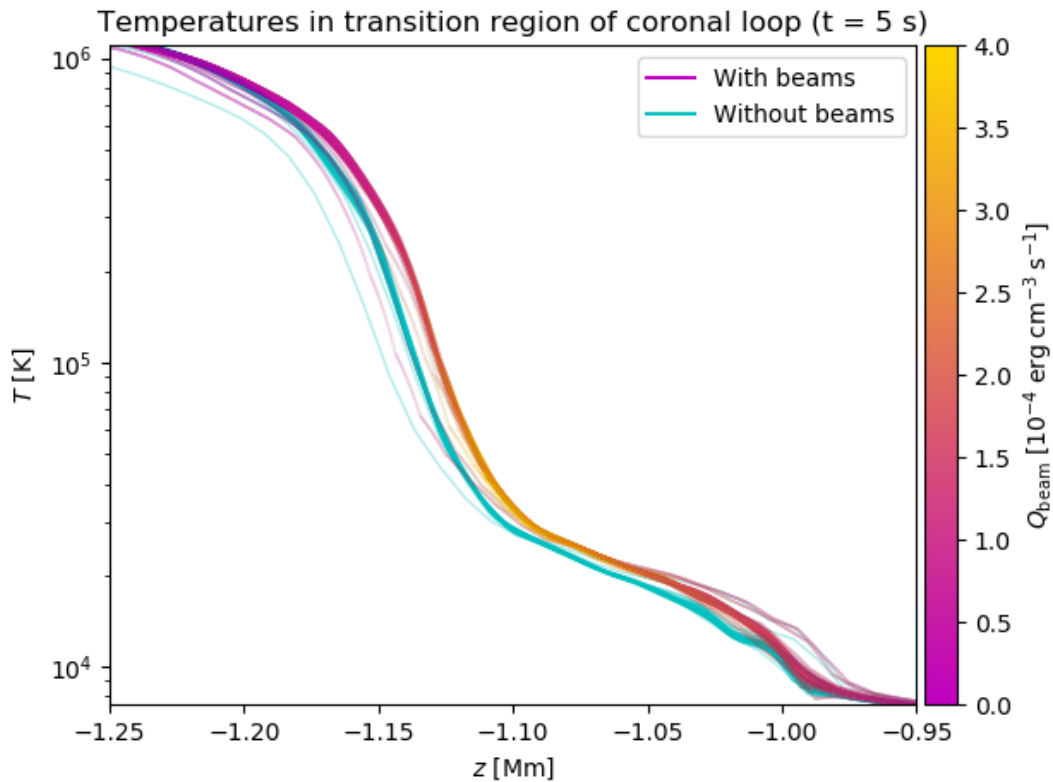
Figure 4.9: Temperature as a function of depth along coronal loop field lines near an intense beam heating region. There are two sets of field lines, taken from both the original and the beam simulation, in both cases when the simulations had been evolved for 5 seconds of solar time. The latter set of field lines are colour coded with the amount of heating that was produced by the beam traversing each field line.

such a bundle of field lines. The red circle in the upper panel of Figure 4.5 indicates the location of the particular heating region in question. To extract the relevant field lines, I started by transporting a full set of electron beams in both the simulation that was run with and without beams. In both cases I recorded which beams passed through a small box centred on the chosen beam heating region and extending about 100 km in each direction. Among these beams I only considered the ones originating near $z = -5$ to $z = -6$ Mm, and with a relatively high energy deposition in the transition region. This resulted in two sets of field lines corresponding to the same coronal loop, one for the run that included beam heating and one for the original run.

Temperature

The temperatures along the coronal loop field lines for both runs (after 5 seconds) are plotted as a function of depth in Figure 4.9. The depth range in the figure spans 300 km, centred on the depth where the beam heating is strongest ($z = -1.1$ Mm). For the field lines from the beam simulation, each curve is colour coded according to the amount of heating generated by the beam that traversed the field line. Both at the
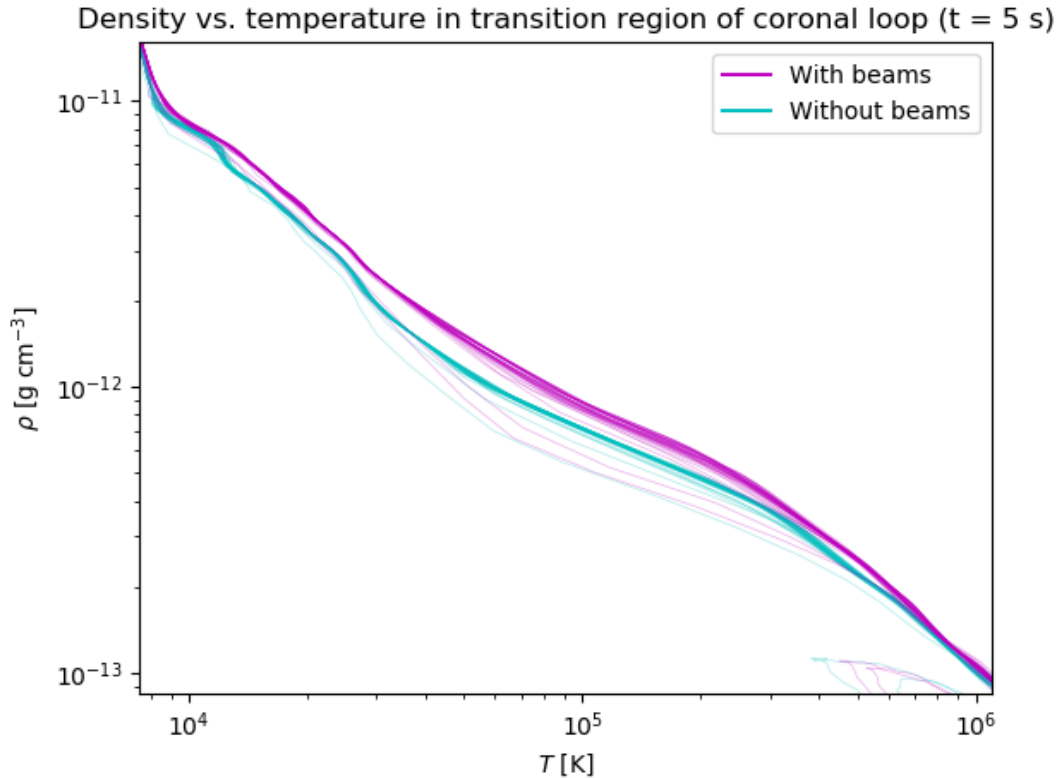
Figure 4.10: Density as a function of temperature along the same field lines as in Figure 4.9.

top and bottom of the depth range, the two bundles have the same temperature. Near the middle, where there is beam heating, the coronal loop from the beam simulation is hotter. The heating is evident for a large range of temperatures, from about 10^4 to 10^6 K. In fact, the beam heating effectively moves the entire transition region about 10 km deeper. This happens because the plasma just below the transition region is made hot enough to halt the radiative cooling and undergo a massive temperature increase, as described in Section 2.1.3.

Recall from Figure 4.7 that the density is rather unaffected by beam heating. The density increases very steeply with depth inside the transition region, so moving the transition region deeper should lead to a higher overall density inside the transition region. Figure 4.10 verifies this suspicion. It shows the densities inside the field line bundles plotted as a function of temperature. The increase in density is evident for nearly the full temperature range from 10^4 to 10^6 K. The largest relative increases occur between 30 000 and 80 000 K, and typically appear to be in the order of a few tens of percent. A higher density for the plasma at a given temperature will tend to enhance the emission of radiation. These results can therefore provide predictions for the degree of brightening that should be observed in the various spectral lines produced at transition region temperatures.

Pressure

Let us now look at the pressure inside the coronal loops, which is plotted as a function of depth in Figure 4.11. For both runs there is a significant pressure drop from $z = -1.15$ to $z = -1.1$ Mm, meaning that the transition region is not in hydrostatic equilibrium. If present at other beam heating sites as well, such a pressure drop would explain the trend of downflows evident in Figure 4.8, as plasma will flow from the higher-pressure region just above the beam heating site to the lower-pressure region below. The exact reason for the irregular pressure profile is not clear, but it is likely related to the high activity in the initial atmosphere. For the simulation with beam heating, the pressure is generally higher inside the beam heating region, in accordance with the elevated temperature.

Vertical velocity

Figure 4.12 shows the vertical plasma velocity along the field lines in the two coronal loops. The previously mentioned downflow is very apparent between $z = -1.2$ and $z = -1.1$ Mm in this figure (recall that a positive vertical velocity corresponds to downward motion). Higher up, the plasma changes direction and starts to move upwards. This can be understood from the moderate drop in pressure when moving upwards from the pressure peak near $z = -1.17$ Mm in Figure 4.11.

The velocity profile in the beam simulation is shifted slightly towards larger depths and has a lower peak compared to original velocity profile. Both of these features can be explained from the changes in pressure. The general increase of pressure by beam heating causes the pressure drop in Figure 4.11 to occur at a slightly larger depth, and
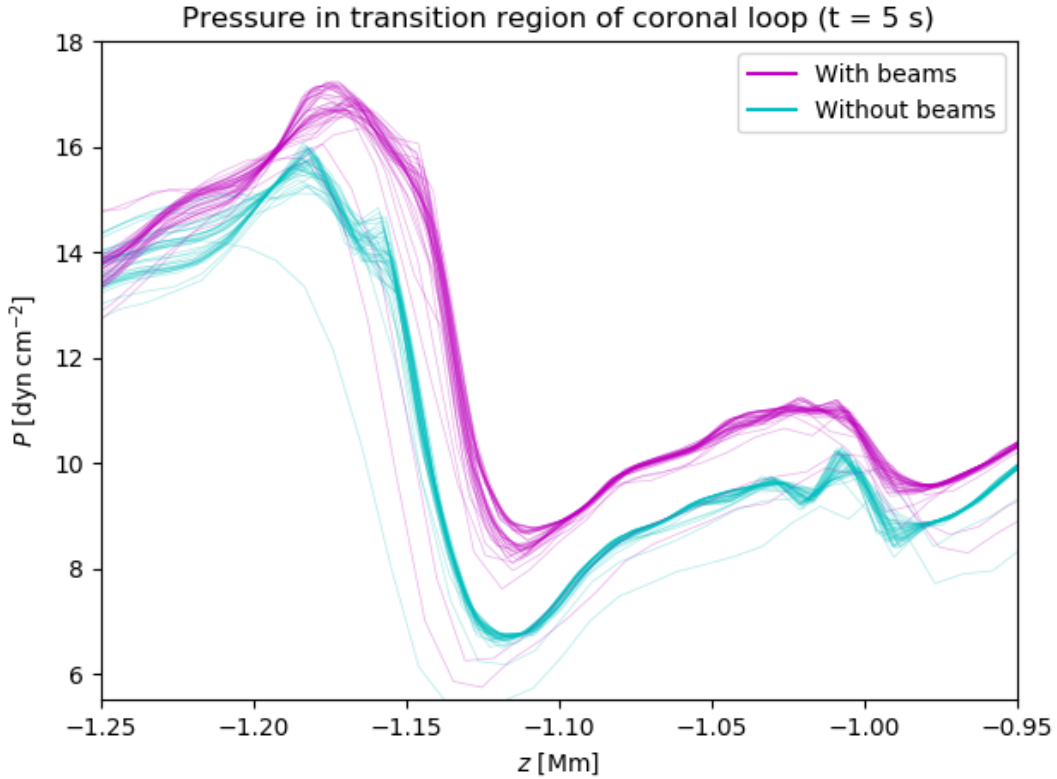
Figure 4.11: Pressure as a function of depth along the same field lines as in Figure 4.9.

hence the largest acceleration of the plasma will take place somewhat deeper down in the transition region. Because the maximum pressure enhancement by beam heating occurs near the bottom of the pressure drop, the pressure gradient is not quite as large in the simulation with beams, causing the maximum downward velocity to be a bit smaller.

Near $z = -1.12$ Mm, the vertical velocity profiles for the two cases intersect. Higher up, the plasma in the beam simulation has a slower downward (or faster upward) motion compared to in the original simulation, while the opposite is true farther down. However, the differences are significantly larger on the upper side of the intersection, and this explains the general trend of the beams increasing the upward velocity at the beam heating sites, as we saw in Figure 4.8. The switch between downward to upward velocity occurs at about $z = -1.2$ Mm in the beam simulation, while in the original simulation it occurs a few tens of kilometres higher up. At this height, where the plasma in the original simulation is stationary, the plasma in the beam simulation has an upward velocity of about 7 km/s.

Figure 4.12 also enables us to understand the average differences in density from Figure 4.7, which shows a very slight increase in density at locations of intermediate beam heating and decrease in density at locations of strong beam heating. The net
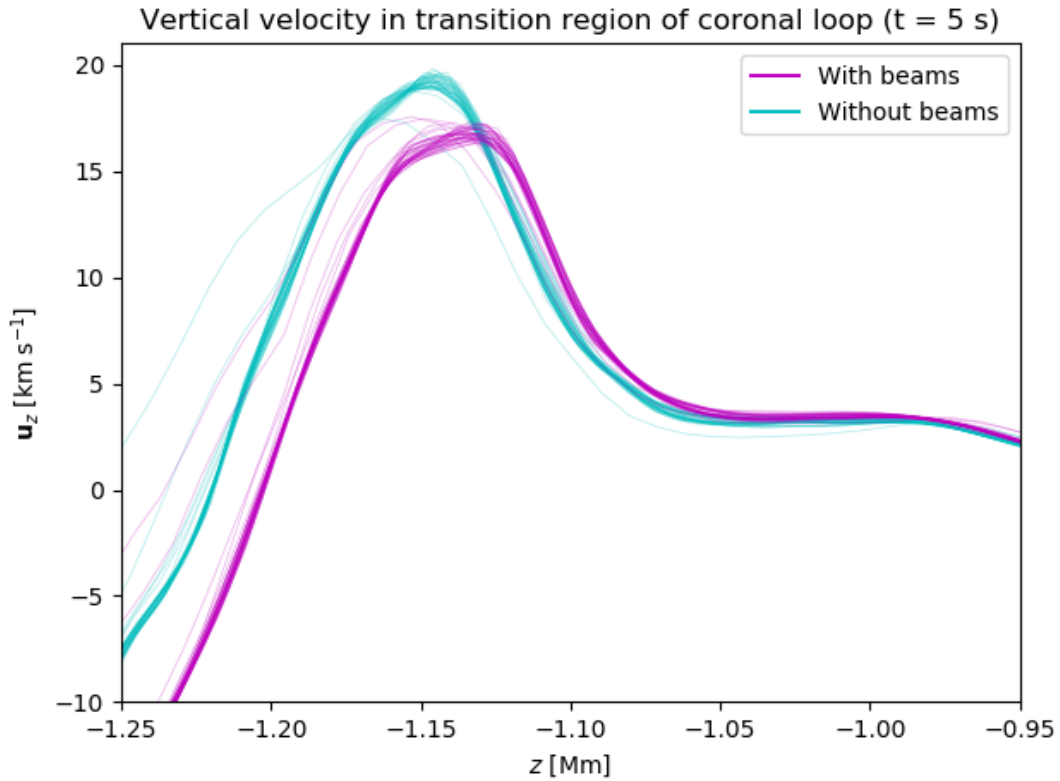
Figure 4.12: Vertical velocity as a function of depth along the same field lines as in Figure 4.9.

effect of the beams is to accelerate plasma in the direction away from the location of the strongest beam heating. When plasma flows away from these locations, the density there will decrease. As this plasma flows to the surrounding regions of intermediate beam heating, the density at these locations will in turn increase.

Time evolution

I conclude this examination of the beam heating's influence on the coronal loop with a quick look at how the temperature changes over time. The temperature at a single position (near the depth of $z = -1.12$ Mm), corresponding to where the beam heating in the coronal loop is strongest, is plotted as a function of time in Figure 4.13. The temperature at the same position in the original simulation is included for comparison, as well as the amount of beam heating. Initially, the beam heating is zero (since no beams have been simulated), and the temperature is about 35 000 K. Even without beam heating, there is a steady increase in temperature over time at a rate of about 500 K/s. This is caused by local Joule heating and heat conduction, as well as by the flow of hotter coronal plasma downwards through the transition region. When the beam heating is switched on, the temperature increases rapidly, reaching about 42 000 K after
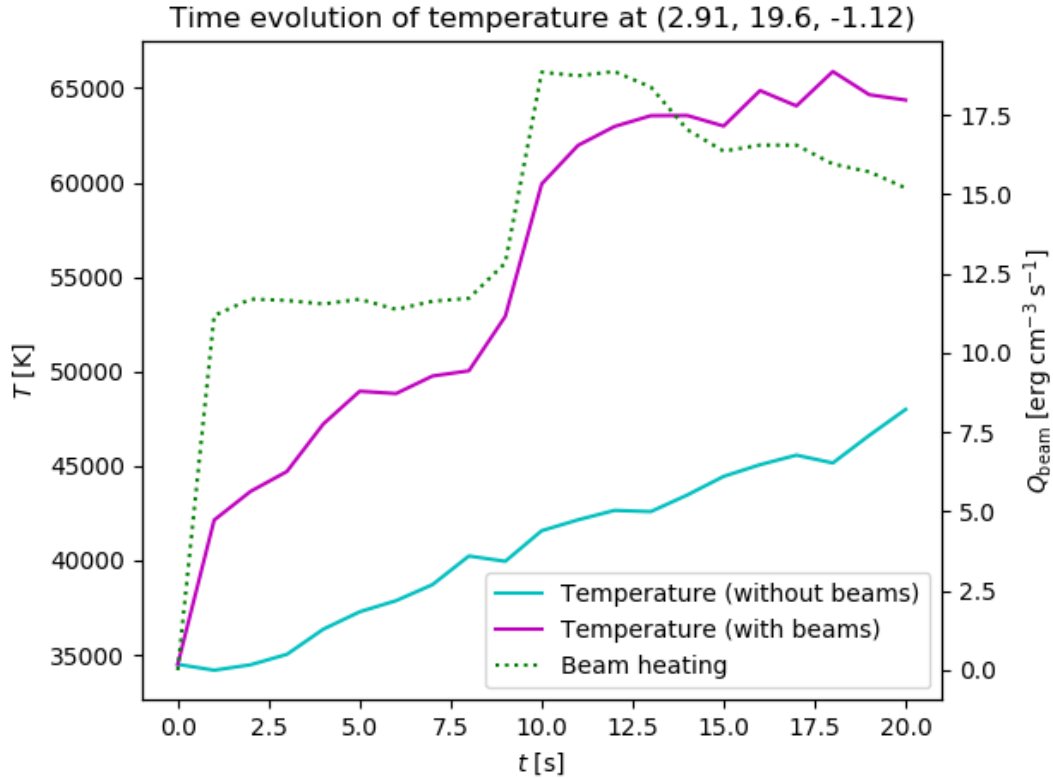
Figure 4.13: Time evolution of the temperature at the most intense beam heating site in the coronal loop, for both the simulation with and without electron beams. The amount of beam heating is also plotted.

the first second. For the next 7 seconds, the beam heating remains constant, while the temperature increases at about the same rate as the case without beams. This indicates that most of the extra heat input is balanced by an increased rate of radiative cooling. Between $t = 9$ and 10 s, the beam heating increases significantly, and this is accompanied by a roughly proportional increase in temperature. The temperature then returns to its original rate of increase, as the beam heating levels out and starts to go slightly down. After 18 seconds, the temperature is about 65 000 K in the beam simulation, compared to 45 000 K in the original simulation.

Discussion

The particular coronal loop explored in this section showed a response to the beam heating that was consistent with the differences studied in Section 4.4.1, and even allowed us to better understand some of the features in the plots of differences. This suggests that the situation is quite similar also in the other beam heating regions in the atmosphere, and that the detailed behaviour studied in this section to some extent is representative of typical heating response in other coronal loops.

The stronger emission at transition region temperatures caused by the shifting of the transition region to higher densities at the beam heating sites is an effect that potentially could be observed. Coronal loop foot point brightenings were indeed observed by Testa et al. (2014) in several transition region spectral lines, including the CII (1335.78 Å) and SI IV (1402.77 Å) lines, which are formed at temperatures of roughly 30 000 and 80 000 K respectively. These are temperatures where we saw some of the largest relative increases in density in Figure 4.10. However, the observed brightenings showed intensity variations by up to a couple of orders of magnitude, while Figure 4.10 suggests that any brightenings produced in the beam simulation at a given temperature would only be by a few tens of percent with respect to the original simulation. This matter could be made clearer by simulating the radiative transfer in detail to produce synthetic observed spectra, rather than just looking at the increase in density.

As mentioned in Section 2.2, Testa et al. (2014) found the observed brightenings to be consistent with 1D numerical simulations of an electron beam. Their best fit with the observations involved a value of $E_{\min}$ of about 10 keV, which enabled a large number of electrons to deposit their energy in the chromosphere, leading to a pressure pulse accelerating the transition region plasma upwards. This produced a blue-shift of the SI IV line, which is consistent with what they observed. The electron beam in their simulation had a total energy of about 10^{25} erg. It is not straightforward to compare this to the energy release in our model, due to the important qualitative difference that each beam heating site in our model is the result of a large number of different, relatively low-energy beams converging on the same location, rather than of a single high-energy beam. However, it is clear that the electron beams simulated in this thesis did not acquire nearly as hard energy distributions, because the method of determining $E_{\min}$ from the intersection between the thermal and non-thermal distributions (Section 3.2.2) naturally leads to small values of $E_{\min}$ in the relatively cool corona of the initial atmosphere. Because of the low fraction of very high-energy electrons, the beams did not penetrate significantly into the chromosphere. Still, we did see an increase in pressure centred at the beam heating sites, which caused the plasma to be accelerated in either direction away from the site. An upward velocity of 7 km/s was found at the point where the plasma would have been stationary in the absence of beams. While this is smaller than the several tens of km/s obtained by Testa et al. (2014), it represents a noteworthy qualitative agreement between the models.

Testa et al. (2014) also reported about 20–30 second variability in the observed brightenings. The time interval covered by Figure 4.13 is not sufficiently long to identify any such variability. However, the figure does show that the amount of beam heating at a given location can change rapidly and by a significant amount, indicating that variations with time scales of the order of seconds are indeed plausible.

The difference between a simulation with and without electron beams is essentially the means by which energy is transported along field lines from the corona to the lower atmosphere. Even in the absence of beams, energy will flow along the field lines due to heat conduction caused by collisions between electrons and ions (modelled in Bifrost by the Spitzer conductivity mentioned in Section 3.1.1). Our results show that electron beams represent a more effective process than Spitzer conductivity for transport of

energy from the corona. In addition, there is an important qualitative difference in that heat conduction will not proceed far past the transition region, as the plasma becomes cool and the resistivity increases. Electron beams, on the other hand, can in principle penetrate deep into the lower atmosphere before they deposit their energy, as long as the electrons are sufficiently energetic. This suggests that the simulation with electron beams would become increasingly different in character from the simulation without beams if we had obtained larger beam energies or harder energy distributions.

As a final note, it is worth stressing that the choice of the initial atmosphere will have had a great influence on the final results, both by determining the amount of energy available for the beams and by providing the setting in which the beams deposit this energy. As discussed in Section 4.1, the particular atmosphere used for this thesis had a relatively dense and cool corona. A hotter corona would have resulted in higher values of $E_{\min}$, and hence a larger fraction of higher-energy beam electrons. We would then have seen deeper penetration of the electron beams into the chromosphere, and possibly have obtained a different atmospheric response to the beam heating. Correspondingly, a more tenuous corona would have let more of the lower-energy beam electrons reach the transition region, increasing the total amount of transition region heating. It could therefore be informative to include the beam physics in different kinds of atmospheres and study the effect of beams under various conditions.

Chapter 5

Summary and concluding remarks

Summary

The aim of this thesis was to develop a model for computing the heating due to beams of non-thermal electrons accelerated by magnetic reconnection in the solar atmosphere, and use it to study the impact of such heating in the atmosphere. In order to perform a realistic 3D simulation of a limited region of the solar atmosphere, we used the radiative MHD code Bifrost. The implementation of the beam simulation code could then be broken down into four tasks.

Firstly, points in the atmosphere where the magnetic field is reconnecting had to be identified. For this we computed a quantity described in Biskamp (2005), which we denoted by K, and made use of the fact that reconnection in theory should happen when $K > 0$. For practical purposes we used a normalised variant $\tilde{K}$, and considered reconnection sites to occur where $\tilde{K}$ exceeded some limit $\tilde{K}_{\mathrm{lim}}$.

Next, the energy distribution of accelerated electrons at the reconnection sites had to be determined. We assumed a power law for the distribution, with a total energy E_{tot} corresponding to half of the energy released by reconnection. The power law index δ was simply set to a number in the range of typical values inferred from observations. Finally, we determined the minimum beam electron energy E_{min} by finding the intersection between the non-thermal power law distribution and the thermal Maxwell–Boltzmann distribution of electrons at the acceleration site.

Both of these tasks are the focus of Helle Bakke's master's thesis and have thus received less attention in this thesis. The primary focus of this thesis has been the latter two tasks: determining the trajectories that the electron beams would follow by tracing magnetic field lines, and computing the energy loss of the beam electrons along the way in order to determine the heating of the ambient plasma.

The tracing of magnetic field lines was performed by solving the ordinary differential equations for the direction of the field line using a Runge–Kutta scheme with an adaptive step length. We applied ideas from control theory in order to stabilise the sequence of step lengths. To avoid excessive computational effort in the lower part of the atmosphere, the field lines were gradually decoupled from the highly irregular magnetic field direction and instead forced to point straight down when the plasma β became

sufficiently high.

For simulating the evolution of the non-thermal electron distribution with distance along the field line, we started with a Fokker–Planck equation describing the effect of Coulomb collisions and magnetic mirroring on the electron beam. To simplify the problem, we assumed the beam electrons to have a trivial distribution of pitch angles. The beam heating at each point along the field line was then calculated as the collective energy loss from the beam electrons. We made the simplifying assumption that each individual electron would not lose any energy until it reached its thermalisation depth, where all of its initial energy would be deposited. This yielded a very simple expression for the beam heating, but resulted in a moderately overestimated beam heating in the transition region. The number of beams to simulate was chosen as a compromise between obtaining enough beam heating and avoiding a too high computational expense. A good balance was struck at around $160\,000$ beams ($\tilde{K}_{\mathrm{lim}} = 2$), as this was when the amount of beam heating started to stagnate.

We included the electron beam physics in an already well-evolved Bifrost simulation of the solar atmosphere, with significant magnetic flux emergence and an active, although relatively cool and dense corona. The atmosphere could then be simulated forward in time both with and without electron beams, so that the effect of beam heating on the atmosphere could be studied. We obtained a distribution of acceleration sites favouring points at relatively low heights, and consisting of clusters corresponding to the intersections of different magnetic domains. Acceleration sites lying too low in the atmosphere were automatically excluded, as their generated beams would deposit all of their energy immediately due to the high density and low temperature, and hence not contribute to any energy transport.

When looking at the heating caused by a single beam, we found that a large fraction of the beam energy was deposited in the immediate vicinity of the acceleration site. This was a result of the rapid thermalisation of the large fraction of relatively low-energy electrons in the beam. The remaining beam electrons were typically thermalised in the transition region, where the density increases rapidly with depth. The latter effect was also seen when we looked at the horizontally averaged heating generated by all the electron beams, where the majority of the beam heating was found to occur at a height of about 1 Mm above the top of the photosphere. The most intense beam heating sites were located at the positions where the bundles of magnetic field lines associated with coronal loops entered the transition region. In the corona, the injection of energy into the beams led to an average reduction of heating. This reduction would have been more significant if it were not for the rapid thermalisation of the lower-energy beam electrons.

We also studied the average differences in various quantities between the simulation with and without beams included, after the simulations had been evolved for a few seconds of solar time. There was a clear correlation between the increase in temperature in the beam simulation and the amount of beam heating. For the strongest beam heating sites, the relative temperature increase compared to the simulation without beams was around 35% on average. There was also a corresponding increase in pressure, while the density remained almost the same in both simulations. The transition region plasma tended to exhibit a downward motion at the positions that were compared, but the

presence of beam heating typically lead to upwardly directed changes in velocity of a few km/s.

For the purpose of better understanding the details of how the beam heating influenced the atmosphere, we went on to look at the state of the plasma in the vicinity of a beam heating region inside a particular coronal loop. The heating had the effect of shifting the transition region downwards by roughly 10 km, to a depth with a slightly higher density. As a result, we predicted an increase in intensity by a few tens of percent for the spectral lines formed at temperatures in the range of 30 000 to 80 000 K. We also saw an increase in pressure centred at the position with the strongest beam heating. This led to an acceleration of plasma in the directions away from this position, predominantly in the upward direction. Finally, we looked briefly at the time evolution of temperature at the position of maximum beam heating in the coronal loop, for both the beam simulation and the original simulation. In both cases there was a steady increase in temperature over time, but the addition of beam heating resulted in an approximately 50% higher final temperature. The rate of temperature increase did not always increase in proportion to the amount of beam heating, suggesting that the rate of radiative cooling increased to compensate for the extra heat input.

Conclusion

The strength of the flare model developed for this and Helle Bakke's thesis is that it is fully integrated into a realistic 3D model of the solar atmosphere. This allows for a self-consistent treatment where the presence of flares can affect the atmosphere and hence have an influence on later flare events[1]. In addition, we can avoid a number of free parameters that would otherwise be required. For instance, in 1D flare models like those of Abbett and Hawley (1999) and Allred et al. (2005, 2015), the total beam energy E_{tot} and minimum electron energy E_{min}, as well as the state of the plasma along the beam trajectory, have to be set based on the particular kind of situation that is the aim to study. In our model, these quantities are all derived from the state of a realistic 3D atmosphere. The drawback with this is a great increase in computational cost, which necessitates a relatively simple treatment of the underlying physical processes compared to what is done in the 1D models (the omission of pitch angle scattering being one example).

For the particular initial atmosphere used in this thesis, we found that the beam heating produced by our simple flare model had a notable effect on the atmospheric plasma in a limited number of regions where the heating was sufficiently strong. These regions correspond to coronal loop foot points, where beams converge along the increasingly concentrated magnetic field lines. Most of the energy was deposited at the bottom of the transition region, and the resulting shift of the transition region to larger depths could lead to slightly enhanced emission in transition region spectral lines like C II and SI IV. Together with the upwards acceleration of plasma found at the heating sites, this is in correspondence with the observations and modelling of Testa et al. (2014).

---

[1]Although, in the particular simulations performed for this thesis, the plasma response was too small to significantly influence the overall state of the atmosphere.

However, the response of the plasma in our model was relatively small compared to what their results would suggest. This can in part be attributed to the low fraction of high-energy electrons in the generated beams, which was a result of the relatively low temperatures at many of the coronal electron acceleration sites. In addition, the somewhat high coronal density of the initial atmosphere caused a large fraction of the total beam energy to be deposited before the electrons reached the transition region. This state of affairs would probably be quite different for a different choice of initial atmosphere.

Future work

Seeing as many aspects of the physical processes governing the interaction between beam particles and the plasma have been neglected or simplified (as discussed throughout Section 3.4.1), there are several improvements that could be made to the flare model to make it more realistic. For instance, including the increased rate of collisional ionisation due to the presence of non-thermal electrons could be important for obtaining the correct plasma response in higher-energy events, as this is the process that triggers the explosive phase of flares (Section 2.2.4). In addition, the ad hoc assumption of a peaked pitch angle distribution, which allowed us to neglect pitch angle scattering, could certainly be improved upon. This could be done by assuming a more general form of the pitch angle distribution, with parameters that are either set freely (like δ in the energy distribution) or derived from the underlying plasma properties (like $E_{\min}$). However, the latter approach would likely require a more detailed model for the acceleration process, which could quickly become computationally expensive. A large amount of extra computational effort would also be required for transporting the electron distribution while including the effect of pitch angle scattering, as the Fokker–Planck equation would have to be solved numerically for every depth. This would also be true if we were to drop the assumption that each electron deposits all of its energy at the thermalisation depth (rather than loosing it continuously), since we would need to numerically evaluate the integral in equation (3.74) at each depth. Rather than just speculating on which aspects of the beam transport simulation that should be given a more accurate treatment, it would probably be better to set up a comparison between the present transport model and that of a detailed 1D code like RADYN (Abbett and Hawley, 1999; Allred et al., 2005, 2015; Carlsson and Stein, 1992, 1995, 1997, 2002).

Due to the limited time at hand, I was not able to simulate the atmosphere for more than about 20 seconds of solar time, which is too short to gain much information about the temporal aspect of the plasma response to the electron beams. Simulations lasting several minutes would be able to reveal the presence of any periodicity similar to the 30-second variations observed by Testa et al. (2014). In addition, the generation of synthetic spectra would provide a better understanding of how the results of the model can be verified observationally. Finally, one of the limitations of these results is that relatively little beam energy reaches the transition region, so the response is very faint. The amount of transition region beam heating is determined by the coolness and denseness of the corona compared to the strength of magnetic reconnection. However,

in order to release a large amount of magnetic energy through reconnection, a strong magnetic flux has to be pushed up though the bottom of the simulation box, and this tends to lift cool, dense photospheric plasma into the corona. The solution to this is to use a larger simulation box, which would naturally allow for stronger magnetic fields higher up in the atmosphere while retaining a hot and tenuous corona. Simulations using larger grids would therefore make it easier to study the effect of beam heating, but this is currently too computationally expensive.

Some of the computational challenges discussed above could be made more manageable by the planned integration of Bifrost into the DISPATCH framework (Nordlund et al., 2018), which allows for the division of the simulation into individual tasks that can be performed in parallel without imposing constraints on the time steps used by other tasks. By letting certain tasks run with a longer time step than is permissible in the current Bifrost code, DISPATCH can free up computational capacity which can be allocated to the simulation of a larger domain. In addition, various solutions to tasks like simulating the electron acceleration process or transporting an electron beam can be integrated into the framework and take full advantage of the task-based approach to parallelisation.

Appendices

Appendix A

Neville's interpolation algorithm

Neville's algorithm is a recursive way of evaluating the interpolating polynomial $P(x)$ described in Section 3.3.1, by constructing it from lower order interpolating polynomials. We start with the zeroth order interpolating polynomials for the points $x_k, x_{k+1}, \ldots, x_{k+N}$. These will simply be the constants corresponding to the function values $y_k, y_{k+1}, \ldots, y_{k+N}$. Let us denote these polynomials by $P_{(i)}(x)$, where $i = 0, \ldots, N$, so that $P_{(i)}(x)$ is the interpolating polynomial for point x_{k+i}. We thus have $P_{(0)}(x) = y_k$, $P_{(1)}(x) = y_{k+1}$, and so forth. There are $N + 1$ such polynomials in total, one for each point. We will soon see how we can use these zeroth order polynomials to construct the N first order polynomials $P_{(i)(i+1)}(x)$ that interpolate between the N point pairs $(x_k, x_{k+1}), (x_{k+1}, x_{k+2}), \ldots, (x_{k+N-1}, x_{k+N})$. These will in turn be used to construct the $N - 1$ second order polynomials $P_{(i)(i+1)(i+2)}(x)$ that interpolate between the $N-1$ point triples $(x_k, x_{k+1}, x_{k+2}), (x_{k+1}, x_{k+2}, x_{k+3}), \ldots, (x_{k+N-2}, x_{k+N-1}, x_{k+N})$. The procedure is then repeated until we end up with the single N'th order polynomial $P_{(i)(i+1)\cdots(N)}(x) = P(x)$ that interpolates between all the points $(x_k, x_{k+1}, \ldots, x_{k+N})$.

So how do we determine the relationship between the $N - n + 1$ polynomials of order n and the $N - n$ polynomials of order $n + 1$? Take for instance the second order polynomial $P_{(0)(1)(2)}(x)$. It is the parabola that passes through the points (x_0, y_0), (x_1, y_1) and (x_2, y_2). We already know the two lines $P_{(0)(1)}(x)$ and $P_{(1)(2)}(x)$ that extend from (x_0, y_0) to (x_1, y_1) and from (x_1, y_1) to (x_2, y_2) respectively. We can use these to construct $P_{(0)(1)(2)}(x)$ by increasing their order by one (through multiplication with x), whilst making sure that the combined expression still produces the values y_0 at x_0, y_1 at x_1 and y_2 at x_2. Here is an expression that fulfils these requirements:

$$P_{(0)(1)(2)}(x) = \frac{(x_2 - x)P_{(0)(1)}(x) + (x - x_0)P_{(1)(2)}(x)}{x_2 - x_0} \tag{A.1}$$

We know that this must indeed be the polynomial we want since it passes through the three points and is of second order. The same idea can be used to derive a general expression for the relation between the two n'th order polynomials $P_{(i)\cdots(i+n)}(x)$ and $P_{(i+1)\cdots(i+n+1)}(x)$, and the $n + 1$'st order polynomial $P_{(i)\cdots(i+n+1)}(x)$:

$$P_{(i)\cdots(i+n)} = \frac{(x_{i+n} - x)P_{(i)\cdots(i+n-1)}(x) + (x - x_i)P_{(i+1)\cdots(i+n)}(x)}{x_{i+n} - x_i} \tag{A.2}$$

Equation (A.2) reduces to equation (A.1) if we set $i = 0$ and $n = 2$. A more practical alternative to equation (A.2) when it comes to implementation (Press et al., 2007) is to define the n'th order *correction polynomials*

$$C_{n,i}(x) \equiv P_{(i)\cdots(i+n)}(x) - P_{(i)\cdots(i+n-1)}(x) \tag{A.3}$$

$$D_{n,i}(x) \equiv P_{(i)\cdots(i+n)}(x) - P_{(i+1)\cdots(i+n)}(x), \tag{A.4}$$

valid for $n = 1, 2, \ldots, N$. In the $n = 0$ case, we simply have

$$C_{0,i}(x) \equiv P_{(i)}(x) \tag{A.5}$$

$$D_{0,i}(x) \equiv P_{(i)}(x). \tag{A.6}$$

So to obtain $P_{(0)(1)(2)}(x)$ from $P_{(0)(1)}(x)$ we could use

$$P_{(0)(1)(2)}(x) = P_{(0)(1)}(x) + C_{2,0}(x). \tag{A.7}$$

If we instead wanted to obtain $P_{(0)(1)(2)}(x)$ from $P_{(1)(2)}(x)$, we could use

$$P_{(0)(1)(2)}(x) = P_{(1)(2)}(x) + D_{2,0}(x). \tag{A.8}$$

Recurrence relations for computing $C_{n+1,i}$ and $D_{n+1,i}$ can be derived from equation (A.2) using the definitions in equations (A.3) and (A.4). The results are

$$C_{n+1,i}(x) = (x - x_i)\left[\frac{C_{n,i+1}(x) - D_{n,i}(x)}{x_{i+n+1} - x_i}\right] \tag{A.9}$$

$$D_{n+1,i}(x) = -(x_{i+n+1} - x)\left[\frac{C_{n,i+1}(x) - D_{n,i}(x)}{x_{i+n+1} - x_i}\right]. \tag{A.10}$$

To obtain $P(x)$, we start by setting $C_{0,i}(x)$ and $D_{0,i}(x)$ equal to y_i, then compute the corrections $C_{1,i}$ and $D_{1,i}$ using equations (A.9) and (A.10). These are in turn inserted into equations (A.9) and (A.10) to obtain $C_{1,i}$ and $D_{1,i}$, and this continues until we have computed $C_{N,i}$ and $D_{N,i}$. In each step n of the process we also maintain a cumulative sum $P_{n,i}(x)$ of correction polynomials that we correct with either $C_{n+1,i}$ or $D_{n+1,i-1}$ (yielding either $P_{n+1,i}(x)$ or $P_{n+1,i-1}(x)$ respectively)[1], landing us at $P(x)$ after the final iteration. The final correction yields an estimate of the error, in the sense that a large final correction indicates that the sequence of corrections has not converged, so that there will be a significant difference between solutions obtained using interpolations of different orders.

---

[1]It makes no difference for the result whether C or D is used, but sometimes only one of them will be defined.

Appendix B

Performing the simulations in parallel

The Bifrost code is typically executed in parallel on a large number of CPUs in order to reduce the execution time as much as possible. Individual instances of the executing program, or *processes*, are distributed among the CPUs. The full simulation box is divided into equally sized sub-domains, one for each process, and every process is responsible for performing the simulation in its assigned sub-domain. In order to fulfil the local boundary conditions, each processes also has to communicate its results at the boundaries of the sub-domain to the neighbouring processes, and in turn receive the corresponding results of the other processes. This is achieved with the help of a communication protocol known as the *Message Passing Interface* (MPI).

As with the rest of Bifrost, the simulations of the electron beams are also parallelised. However, the task of tracing a magnetic field line is inherently serial; each new position has to be computed from the previous position. As discussed at the end of Section 3.3.6, the beam heating is computed in tandem with the tracing of the associated field line, so we cannot split these tasks between processes either. But there is still a potential for concurrency due to the fact that we simulate a large number of beams in each time step.

Since a process only has access to the part of the atmosphere belonging to its sub-domain, each process is responsible for simulating a beam as long as it resides in its sub-domain. This includes the simulation of all beams that originate in the sub-domain until they pass into another domain, as well as beams originating in other domains but passing through the sub-domain belonging to the process.

Each process has a message buffer for storing the current state of all incoming beams. As soon as it has simulated the current beam, it checks the buffer and begins simulating the next incoming beam. If there are no incoming beams at the moment, it instead begins the simulation of the next beam originating in its sub-domain. When a process stops simulating a beam, it is either because the beam is completed (because there is no energy left in the beam or it has reached a non-periodic boundary), or because the beam has entered a neighbouring sub-domain. In the former case, the process simply

increments a count of how many beams it has completed. In the latter case, the process determines which process should take over the beam and stores the current state of the beam in the message buffer of that process.

There is one process, hereafter referred to as the communication master, which is additionally responsible for determining when all the beams have been completed. One of the processes lying at the top of the atmosphere is chosen for this task, because they typically have fewer beams to simulate than the processes farther down. After a process has no remaining beams originating in its sub-domain, it will check its count of completed beams every time it does not detect an incoming beam. If the count is non-zero (because it has completed new incoming beams), it sends the count to the communication master and resets it to zero. The communication master sums up the number of completed beams from all processes, and when this number reaches the total number of beams it puts a special message in the message buffers of the other processes, instructing them to exit the simulation loop.
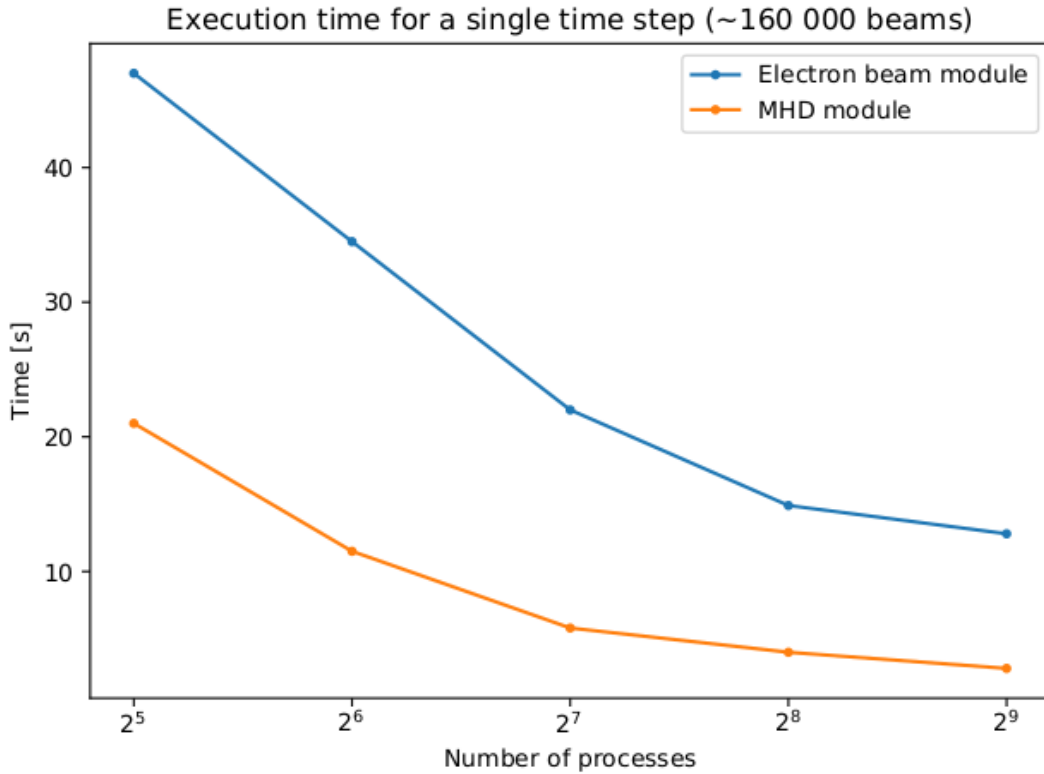
Figure B.1: Measured execution times of the electron beam module for a single time step with various numbers of MPI processes. The corresponding execution times for the module responsible for the MHD simulation are also included for reference. The processes were evenly distributed over 29 individual nodes (connected computers) during each run. Each run was performed with a threshold of $\tilde{K}_{\text{lim}} = 2$, resulting in roughly 160 000 beams.

The amount of work will not be evenly distributed between the processes, because some parts of the atmosphere will have a higher traffic of beams than other. In particular, processes with sub-domains very high in the corona or deep in the lower atmosphere where the beams cannot reach, are likely to be mostly idle. The communication of beam states and counts will also steal processing resources. We therefore cannot expect the execution time to decrease linearly with the number of processes. Figure B.1 shows how the execution time for the beam simulations varied with the number of MPI processes when run on a particular computer cluster.

The performance scales decently well going from 32 to 128 processes, with the execution time roughly being cut to half by the fourfold increase in number of processes. It then starts to level out, with relatively little speedup from 256 to 512 processes. This behaviour is likely due to the aforementioned uneven load balance between the processes. We actually see a similar behaviour for the MHD module. However, in this
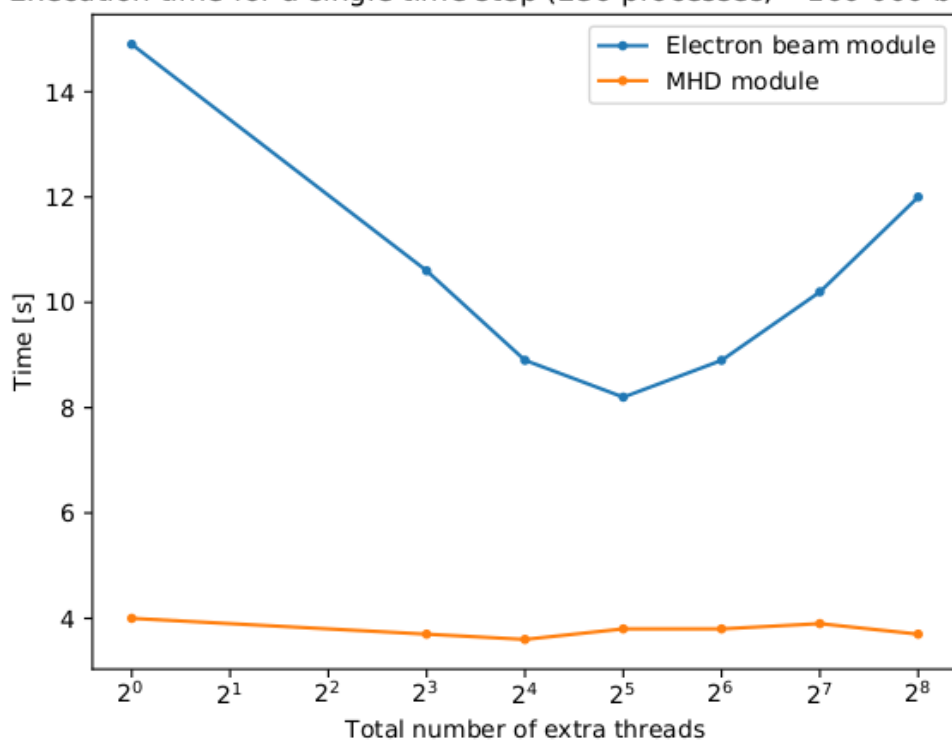
Figure B.2: Measured execution times of the electron beam module for a single time step with various numbers of extra threads. The number of processes is 256 in each run, and the threshold $\tilde{K}_{\text{lim}}$ is the same as in Figure B.1. Like in Figure B.1, the execution times for the MHD simulation are also shown for reference. The first data point (2^0 extra threads) in practice yielded no extra threads due to the numbers being rounded to zero when each process computes the number of additional threads to use.

case the work should be very evenly distributed, so the stagnation is more likely a result of the simulations reaching a point where the decrease in computation time is offset by the increase in communication time.

The uneven load balance can be improved by having multiple CPUs simulating beams within the same sub-domain. This can be achieved by letting each process spawn a number of *threads* before entering the simulation loop, and distribute among the threads the beams to be simulated in the sub-domain. A thread is a lightweight variant of a process: the threads spawned by a process share nearly all the data associated with the process, apart from a few quantities that each thread needs to have independent copies of, like the state of the current beam to simulate. By deciding the number of threads to use in each sub-domain based on the number of beams that originate there, the available processing resources can be applied where they are needed the most.

I implemented this functionality using a multithreading API[1] called *OpenMP* (shorthand for Open Multi-Processing). I then ran a simulation while varying the total number of extra threads that were made available for the processes. The corresponding execution times for a single time step are shown in Figure B.2.

With just a small number of additional threads, the performance improves quite substantially. The "sweet spot" appears to be 32 extra threads, which reduces the execution time by 45% while only requiring a 12.5% increase in the number of active CPU cores. This cheap performance increase is possible because the majority of all the work is done in just a hand-full of sub-domains, and this is where the additional threads are provided. However, too many threads in a single sub-domain will become counter-productive, because the CPUs running the threads will eventually have to be physically far apart, meaning that some of them only have relatively slow access to their shared memory[2]. These threads will then mostly get in the way of the faster threads while adding little in terms of computational resources.

---

[1] An API, or *application programming interface*, is essentially an interface to a set of software components, enabling you to use them within your application source code.

[2] This is a result of the NUMA (non-uniform memory access) design used for the memory system of the nodes in the cluster.

List of figures

Bibliography

Abbett, W. P. and Hawley, S. L. (1999), 'Dynamic Models of Optical Emission in Impulsive Solar Flares', *The Astrophysical Journal* **521**(2), 906–919. doi:10.1086/307576.

Alexander, D. and Metcalf, T. R. (1997), 'A Spectral Analysis of the Masuda Flare Using Yohkoh Hard X-Ray Telescope Pixon Reconstruction', *The Astrophysical Journal* **489**(1), 442–455. doi:10.1086/304762.

Allred, J. C., Hawley, S. L., Abbett, W. P. and Carlsson, M. (2005), 'Radiative Hydrodynamic Models of the Optical and Ultraviolet Emission from Solar Flares', *The Astrophysical Journal* **630**(1), 573–586. doi:10.1086/431751.

Allred, J. C., Kowalski, A. F. and Carlsson, M. (2015), 'A Unified Computational Model for Solar and Stellar Flares', *The Astrophysical Journal* **809**(1), 104. doi:10.1088/0004-637X/809/1/104.

Antonucci, E., Dennis, B. R., Gabriel, A. H. and Simnett, G. M. (1985), 'Initial phase of chromospheric evaporation in a solar flare', *Solar Physics* **96**(1), 129–142. doi:10.1007/BF00239797.

Bakke, H. M. (2018), Acceleration of High-Energy Electrons by Magnetic Reconnection in Coronal Nanoflares, Master's thesis, Institute of Theoretical Astrophysics, University of Oslo.

Biskamp, D. (2005), *Magnetic Reconnection in Plasmas*, Cambridge Monographs on Plasma Physics, Cambridge University Press.

Bogacki, P. and Shampine, L. F. (1989), 'A 3(2) pair of Runge - Kutta formulas', *Applied Mathematics Letters* **2**(4), 321 – 325. doi:https://doi.org/10.1016/0893-9659(89)90079-7.

Borissov, A., Kontar, E. P., Threlfall, J. and Neukirch, T. (2017), 'Particle acceleration with anomalous pitch angle scattering in 2D magnetohydrodynamic reconnection simulations', *Astronomy & Astrophysics* **605**, A73. doi:10.1051/0004-6361/201731183.

Brown, J. C. (1971), 'The Deduction of Energy Spectra of Non-Thermal Electrons in Flares from the Observed Dynamic Spectra of Hard X-Ray Bursts', *Solar Physics* **18**(3), 489–502. doi:10.1007/BF00149070.

Brown, J. C. (1972), 'The Directivity and Polarisation of Thick Target X-Ray Bremsstrahlung from Solar Flares', *Solar Physics* **26**(2), 441–459. doi:10.1007/BF00165286.

Cargill, P. (2013), 'From flares to nanoflares: magnetic reconnection on the Sun', *Astronomy and Geophysics* **54**(3), 3.16–3.20. doi:10.1093/astrogeo/att078.

Carlsson, M. and Hansteen, V. (2005), Chromospheric Waves, *in* D. E. Innes, A. Lagg and S. A. Solanki, eds, 'Chromospheric and Coronal Magnetic Fields', Vol. 596 of *ESA Special Publication*, p. 39.1.

Carlsson, M. and Stein, R. F. (1992), 'Non-LTE radiating acoustic shocks and CA II K2V bright points', *The Astrophysical Journal Letters* **397**, L59–L62. doi:10.1086/186544.

Carlsson, M. and Stein, R. F. (1995), 'Does a nonmagnetic solar chromosphere exist?', *The Astrophysical Journal Letters* **440**, L29–L32. doi:10.1086/187753.

Carlsson, M. and Stein, R. F. (1997), 'Formation of Solar Calcium H and K Bright Grains', *The Astrophysical Journal* **481**(1), 500–514. doi:10.1086/304043.

Carlsson, M. and Stein, R. F. (2002), 'Dynamic Hydrogen Ionization', *The Astrophysical Journal* **572**, 626–635. doi:10.1086/340293.

Christe, S., Hannah, I. G., Krucker, S., McTiernan, J. and Lin, R. P. (2008), 'RHESSI Microflare Statistics. I. Flare-Finding and Frequency Distributions', *The Astrophysical Journal* **677**(2), 1385–1394. doi:10.1086/529011.

Crosby, N. B., Aschwanden, M. J. and Dennis, B. R. (1993), 'Frequency distributions and correlations of solar X-ray flare parameters', *Solar Physics* **143**(2), 275–299. doi:10.1007/BF00646488.

De Pontieu, B., Erdélyi, R. and James, S. P. (2004), 'Solar chromospheric spicules from the leakage of photospheric oscillations and flows', *Nature* **430**(6999), 536–539. doi:10.1038/nature02749.

Dormand, J. R. and Prince, P. J. (1980), 'A family of embedded Runge-Kutta formulae', *Journal of Computational and Applied Mathematics* **6**(1), 19 – 26. doi:https://doi.org/10.1016/0771-050X(80)90013-3.

Emslie, A. G. (1980), 'The effect of reverse currents on the dynamics of nonthermal electron beams in solar flares and on their emitted X-ray bremsstrahlung', *The Astrophysical Journal* **235**, 1055–1065. doi:10.1086/157709.

Emslie, A. G., Dennis, B. R., Holman, G. D. and Hudson, H. S. (2005), 'Refinements to flare energy estimates: A followup to "Energy partition in two solar flare/CME events" by A. G. Emslie et al.', *Journal of Geophysical Research (Space Physics)* **110**(A9), A11103. doi:10.1029/2005JA011305.

Emslie, A. G., Dennis, B. R., Shih, A. Y., Chamberlin, P. C., Mewaldt, R. A., Moore, C. S., Share, G. H., Vourlidas, A. and Welsch, B. T. (2012), 'Global Energetics of Thirty-eight Large Solar Eruptive Events', *The Astrophysical Journal* **759**, 71. doi:10.1088/0004-637X/759/1/71.

Emslie, A. G., Kucharek, H., Dennis, B. R., Gopalswamy, N., Holman, G. D., Share, G. H., Vourlidas, A., Forbes, T. G., Gallagher, P. T., Mason, G. M., Metcalf, T. R., Mewaldt, R. A., Murphy, R. J., Schwartz, R. A. and Zurbuchen, T. H. (2004), 'Energy partition in two solar flare/CME events', *Journal of Geophysical Research (Space Physics)* **109**(A18), A10104. doi:10.1029/2004JA010571.

Feldman, U., Doschek, G. A., Behring, W. E. and Phillips, K. J. H. (1996), 'Electron Temperature, Emission Measure, and X-Ray Flux in A2 to X2 X-Ray Class Solar Flares', *The Astrophysical Journal* **460**, 1034. doi:10.1086/177030.

Ginzburg, V. L. and Syrovatskii, S. I. (1964), *The Origin of Cosmic Rays*, Macmillan, New York.

Gudiksen, B. V., Carlsson, M., Hansteen, V. H., Hayek, W., Leenaarts, J. and Martínez-Sykora, J. (2011), 'The stellar atmosphere simulation code Bifrost. Code description and validation', *Astronomy and Astrophysics* **531**, A154. doi:10.1051/0004-6361/201116520.

Gustafsson, K. (1994), 'Control-theoretic Techniques for Stepsize Selection in Implicit Runge-Kutta Methods', *ACM Transactions on Mathematical Software* **20**(4), 496–517. doi:10.1145/198429.198437.

Hayek, W., Asplund, M., Carlsson, M., Trampedach, R., Collet, R., Gudiksen, B. V., Hansteen, V. H. and Leenaarts, J. (2010), 'Radiative transfer with scattering for domain-decomposed 3D MHD simulations of cool stellar atmospheres. Numerical methods and application to the quiet, non-magnetic, surface of a solar-type star', *Astronomy and Astrophysics* **517**, A49. doi:10.1051/0004-6361/201014210.

Hyman, J. M. (1979), A method of lines approach to the numerical solution of conservation laws, *in* 'Advances in Computer Methods for Partial Differential Equations - III', pp. 313–321.

Jeffrey, N. L. S. (2014), The spatial, spectral and polarization properties of solar flare X-ray sources, PhD thesis, University of Glasgow.

Jess, D. B., Morton, R. J., Verth, G., Fedun, V., Grant, S. D. T. and Giagkiozis, I. (2015), 'Multiwavelength Studies of MHD Waves in the Solar Chromosphere. An Overview of Recent Results', *Space Science Reviews* **190**, 103–161. doi:10.1007/s11214-015-0141-3.

Kane, S. R. (1973), 'Characteristics of Nonthermal Electrons Accelerated During the Flash Phase of Small Solar Flares', *NASA Special Publication* **342**, 55.

Kontar, E. P., Hannah, I. G. and Bian, N. H. (2011), 'Acceleration, Magnetic Fluctuations, and Cross-field Transport of Energetic Electrons in a Solar Flare Loop', *The Astrophysical Journal Letters* **730**, L22. doi:10.1088/2041-8205/730/2/L22.

Kontar, E. P., Hannah, I. G., Jeffrey, N. L. S. and Battaglia, M. (2010), 'The Subarcsecond Hard X-ray Structure of Loop Footpoints in a Solar Flare', *The Astrophysical Journal* **717**(1), 250–256. doi:10.1088/0004-637X/717/1/250.

Leach, J. (1984), The Impulsive Hard X-Rays from Solar Flares., PhD thesis, Stanford University.

Leach, J. and Petrosian, V. (1981), 'Impulsive phase of solar flares. I - Characteristics of high energy electrons', *The Astrophysical Journal* **251**, 781–791. doi:10.1086/159521.

Masuda, S., Kosugi, T., Hara, H., Tsuneta, S. and Ogawara, Y. (1994), 'A loop-top hard X-ray source in a compact solar flare as evidence for magnetic reconnection', *Nature* **371**(6497), 495–497. doi:10.1038/371495a0.

Miller, J. A., Cargill, P. J., Emslie, A., Holamn, G. D., Dennis, B. R., Larosa, T. N., Winglee, R. M., Benka, S. G. and Tsuneta, S. (1997), 'Critical Issues For Understanding Particle Acceleration in Impulsive Solar Flares', *Journal of Geophysical Research* **102**.

Nordlund, Å., Ramsey, J. P., Popovas, A. and Küffmeier, M. (2018), 'DISPATCH: a numerical simulation framework for the exa-scale era - I. Fundamentals', *Monthly Notices of the Royal Astronomical Society* **477**(1), 624–638. doi:10.1093/mnras/sty599.

Parker, E. N. (1988), 'Nanoflares and the solar X-ray corona', *Astrophysical Journal* **330**, 474–479. doi:10.1086/166485.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (2007), *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3 edn, Cambridge University Press, New York, NY, USA.

Rosenbluth, M. N., MacDonald, W. M. and Judd, D. L. (1957), 'Fokker-Planck Equation for an Inverse-Square Force', *Physical Review* **107**(1), 1–6. doi:10.1103/PhysRev.107.1.

Shampine, L. F. (1986), 'Some Practical Runge-Kutta Formulas', *Mathematics of Computation* **46**(173), 135–150. doi:10.2307/2008219.

Shibata, K. (1999), 'Evidence of Magnetic Reconnection in Solar Flares and a Unified Model of Flares', *Astrophysics and Space Science* **264**(1/4), 129–144. doi:10.1023/A:1002413214356.

Shibata, K. and Magara, T. (2011), 'Solar Flares: Magnetohydrodynamic Processes', *Living Reviews in Solar Physics* **8**, 6. doi:10.12942/lrsp-2011-6.

Svestka, Z. (1989), 'Solar flares - The gradual phase', *Solar Physics* **121**(1-2), 399–417. doi:10.1007/BF00161709.

Syrovatskii, S. I. (1972), 'Solar Flare Time Development: Three Phases', *Comments on Astrophysics and Space Physics* **4**, 65.

Testa, P., De Pontieu, B., Allred, J., Carlsson, M., Reale, F., Daw, A., Hansteen, V., Martinez-Sykora, J., Liu, W., DeLuca, E. E., Golub, L., McKillop, S., Reeves, K., Saar, S., Tian, H., Lemen, J., Title, A., Boerner, P., Hurlburt, N., Tarbell, T. D., Wuelser, J. P., Kleint, L., Kankelborg, C. and Jaeggli, S. (2014), 'Evidence of nonthermal particles in coronal loops heated impulsively by nanoflares', *Science* **346**(6207), 1255724. doi:10.1126/science.1255724.

Vernazza, J. E., Avrett, E. H. and Loeser, R. (1981), 'Structure of the solar chromosphere. III - Models of the EUV brightness components of the quiet-sun', *The Astrophysical Journal Supplement* **45**, 635–725. doi:10.1086/190731.

Zarro, D. M., Canfield, R. C., Metcalf, T. R. and Strong, K. T. (1988), 'Explosive plasma flows in a solar flare', *The Astrophysical Journal* **324**, 582–589. doi:10.1086/165919.

Zharkova, V. V., Arzner, K., Benz, A. O., Browning, P., Dauphin, C., Emslie, A. G., Fletcher, L., Kontar, E. P., Mann, G., Onofri, M., Petrosian, V., Turkmani, R., Vilmer, N. and Vlahos, L. (2011), 'Recent Advances in Understanding Particle Acceleration Processes in Solar Flares', *Space Science Reviews* **159**, 357–420. doi:10.1007/s11214-011-9803-y.